Forecasting US Equity Returns in the 21st Century

John Y. Campbell, Harvard University
July 2001

What returns should investors expect the US stock market to deliver on average during the next century? Does the experience of the last century provide a reliable guide to the future? In this short note I first discuss alternative methodologies for forecasting average future equity returns, then discuss current market conditions, and finally draw conclusions for long-term return forecasts. Throughout I work in real, that is inflation-adjusted, terms.

I. Methods for forecasting returns

1. Average past returns

Perhaps the simplest way to forecast future returns is to use some average of past returns. Very naturally, this method has been favored by many investors and analysts. However there are several difficulties with it.

a) *Geometric average or arithmetic average?* The geometric average return is the cumulative past return on US equities, annualized. Siegel (1998) studies long-term historical data on value-weighted US share indexes. He reports a geometric average of 7.0% over two different sample periods, 1802–1997 and 1871–1997. The arithmetic average return is the average of one-year past returns on US equities. It is considerably higher than the geometric average return, 8.5% over 1802–1997 and 8.7% over 1871–1997.[1]

When returns are serially uncorrelated, the arithmetic average represents the best forecast of future return in any randomly selected future year. For long holding periods, the best forecast is the arithmetic average compounded up appropriately. If one is making a 75-year forecast, for example, one should forecast a cumulative return of 1.085^{75} based on 1802–1997 data.

When returns are negatively serially correlated, however, the arithmetic average is not necessarily superior as a forecast of long-term future returns. To understand this, consider an extreme example in which prices alternate deterministically between 100 and 150. The return is 50% when prices rise, and -33% when prices fall. Over any even number of periods, the geometric average return is zero, but the arithmetic average return is 8.5%. In this case the arithmetic average return is misleading because it fails to take account of the fact that high returns always multiply a low initial price of 100, while low returns always multiply a high initial price of 150. The geometric average is a better indication of long-term future

[1] When returns are lognormally distributed, the difference between the two averages is approximately one-half the variance of returns. Since stock returns have an annual standard deviation of about 18% over these long periods, the predicted difference is $0.18^2/2 = 0.016$ or 1.6%. This closely matches the difference in the data.

prospects in this example.[2]

This point is not just a theoretical curiosity, because in the historical data summarized by Siegel, there is strong evidence that the stock market is mean-reverting. That is, periods of high returns tend to be followed by periods of lower returns. This suggests that the arithmetic average return probably overstates expected future returns over long periods.

b) *Returns are very noisy.* The randomness in stock returns is extreme. With an annual standard deviation of real return of 18%, and 100 years of past data, a single year's stock return that is only one standard deviation above average increases the average return by 18 basis points. A lucky year that is two standard deviations above average increases the average return by 36 basis points. Even when a century or more of past data is used, forecasts based on historical average returns are likely to change substantially from one year to the next.

c) *Realized returns rise when expected returns fall.* To the extent that expected future equity returns are not constant, but change over time, they can have perverse effects on realized returns. Suppose for example that investors become more risk-tolerant and reduce the future return that they demand from equities. If expected future cash flows are unchanged, this drives up prices and realized returns. Thus an estimate of future returns based on average past realized returns will tend to increase just as expected future returns are declining.

Something like this probably occurred in the late 1990's. A single good year can have a major effect on historical average returns, and several successive good years have an even larger effect. But it would be a mistake to react to the spectacular returns of 1995–99 by increasing estimates of 21st Century returns.

d) *Unpalatable implications.* Fama and French (2000) point out that average past US stock returns are so high that they exceed estimates of the return to equity (ROE) calculated for US corporations from accounting data. Thus if one uses average past stock returns to estimate the cost of capital, the implication is that US corporate investments have destroyed value; corporations should instead have been paying all their earnings out to stockholders. This conclusion is so hard to believe that it further undermines confidence in the average-return methodology.

One variation of the average-past-returns approach is worth discussing. One might take the view that average past equity returns in other countries provide relevant evidence about US equity returns. Standard international data from Morgan Stanley Capital International, available since the early 1970's, show that equity returns in most other industrialized countries have been about as high as those in the US. The exceptions are the heavily commodity-dependent markets of Australia and Canada, and the very small Italian market (Campbell 1999). Jorion and Goetzmann (1999) argue that other countries' returns were

[2] One crude way to handle this problem is to measure the annualized variance of returns over a period such as 20 years that is long enough for returns to be approximately serially uncorrelated, and then to adjust the geometric average up by one-half the annualized 20-year variance as would be appropriate if returns are lognormally distributed. Campbell and Viceira (2001, Figure 4.2) report an annualized 20-year standard deviation of about 14% in long-term annual US data, which would imply an adjustment of $0.14^2/2 = 0.010$ or 1.0%.

lower than US returns in the early 20th Century, but this conclusion appears to be sensitive to their omission of the dividend component of return (Dimson, Marsh, and Staunton 2000). Thus the use of international data does not change the basic message that the equity market has delivered high average returns in the past.

2. Valuation ratios

An alternative approach is to use valuation ratios—ratios of stock prices to accounting measures of value such as dividends or earnings—to forecast future returns. In a model with constant valuation ratios and growth rates, the famous Gordon growth model says that the dividend-price ratio

$$\frac{D}{P} = R - G, \tag{1}$$

where R is the discount rate or expected equity return, and G is the growth rate of dividends (equal to the growth rate of prices when the valuation ratio is constant). This formula can be applied either to price per share and conventional dividends per share, or to the total value of the firm and total cash paid out by the firm (including share repurchases). A less well-known but just as useful formula says that in steady state, where earnings growth comes from reinvestment of retained earnings which earn an accounting ROE equal to the discount rate R,

$$\frac{E}{P} = R. \tag{2}$$

Over long periods of time summarized by Siegel (1998), these formulas give results consistent with average realized returns. Over the period 1802–1997, for example, the average dividend-price ratio was 5.4% while the geometric average growth rate of prices was 1.6%. These numbers add to the geometric average return of 7.0%. Over the period 1871–1997 the average dividend-price ratio was 4.9% while the geometric average growth rate of prices was 2.1%, again adding to 7.0%. Similarly, Campbell and Shiller (2001) report that the average P/E ratio for S&P500 shares over the period 1872-2000 was 14.5. The reciprocal of this is 6.9%, consistent with average realized returns.

When valuation ratios and growth rates change over time, these formulas are no longer exactly correct. Campbell and Shiller (1988) and Vuolteenaho (2000) derive dynamic versions of the formulas that can be used in this context. Campbell and Shiller show, for example, that the log dividend-price ratio is a discounted sum of expected future discount rates, less a discounted sum of expected future dividend growth rates. In this note I will work with the simpler deterministic formulas.

II. Current market conditions

Current valuation ratios are wildly different from historical averages, reflecting the unprecedented bull market of the last 20 years, and particularly the late 1990's. The attached figure, taken from Campbell and Shiller (2001), illustrates this point. The bottom left panel shows the dividend-price ratio D/P in January of each year from 1872–2000. The long-term historical average is 4.7%, but D/P has fallen dramatically since 1982 to about 1.2% in January 2000 (and 1.4% today).

The dividend-price ratio may have fallen in part because of shifts in corporate financial policy. An increased tendency for firms to repurchase shares rather than pay dividends increases the growth rate of dividends per share, by shrinking the number of shares. Thus it increases G in the Gordon growth formula and reduces conventionally measured D/P. One way to correct for this is to add repurchases to conventional dividends. Recent estimates of this effect by Liang and Sharpe (1999) suggest that it may be an upward adjustment of 75 to 100 basis points, and more in some years. Of course, this is not nearly sufficient to explain the recent decline in D/P.

Alternatively, one can look at the price-earnings ratio. The top left panel of the figure shows P/E over the same period. This has been high in recent years, but there are a number of earlier peaks that are comparable. Close inspection of these peaks shows that they often occur in years such as 1992, 1934, and 1922 when recessions caused temporary drops in (previous-year) earnings. To smooth out this effect, Campbell and Shiller (2001), following Graham and Dodd (1934), advocate averaging earnings over 10 years. The price-averaged earnings ratio is illustrated in the top right panel of the figure. This peaked at 45 in January 2000; the previous peak was 28 in 1929. The decline in the S&P500 since January 2000 has only brought the ratio down to the mid-30's, still higher than any level seen before the late 1990's.

The final panel in the figure, on the bottom right, shows the ratio of current to 10-year average earnings. This ratio has been high in recent years, reflecting robust earnings growth during the 1990's, but it is not unprecedentedly high. The really unusual feature of the recent stock market is the level of prices, not the growth of earnings.

III. Implications for future returns

The implications of current valuations for future returns depend on whether the market has reached a new steady state, in which current valuations will persist, or whether these valuations are the result of some transitory phenomenon.

If current valuations represent a new steady state, then they imply a substantial decline in the equity returns that can be expected in the future. Using Campbell and Shiller's (2001) data, the unadjusted dividend-price ratio has declined by 3.3 percentage points from the historical average. Even adjusting for share repurchases, the decline is at least 2.3 percentage points. Assuming constant long-term growth of the economy, this would imply that the geometric average return on equity is no longer 7%, but 3.7% or at most 4.7%. Looking at the price-averaged earnings ratio, adjusting for the typical ratio of current to averaged earnings, gives an even lower estimate. Current earnings are normally 1.12 times averaged earnings; $1.12/35 = 0.032$, implying a 3.2% return forecast. These forecasts allow for only a very modest equity premium relative to the yield on long-term inflation-indexed bonds, currently about 3.5%, or the 3% safe real return assumed recently by the Trustees.

If current valuations are transitory, then it matters critically what happens to restore traditional valuation ratios. One possibility is that earnings and dividends are below their long-run trend levels; rapid earnings and dividend growth will restore traditional valuations without any declines in equity returns below historical levels. While this is always a possibility, Campbell and Shiller (2001) show that it would be historically unprecedented. The US stock market has an extremely poor record of predicting future earnings and dividend growth. Historically stock prices have increased relative to earnings during decades of rapid earnings growth, such as the 1920's, 1960's, or 1990's, as if the stock market anticipates that rapid earnings growth will continue in the next decade. However there is no systematic tendency for a profitable decade to be followed by a second profitable decade; the 1920's, for example, were followed by the 1930's and the 1960's by the 1970's. Thus stock market optimism often fails to be justified by subsequent earnings growth.[3]

A second possibility is that stock prices will decline or stagnate until traditional valuations are restored. This has occurred at various times in the past after periods of unusually high stock prices, notably the 1900's and 1910's, the 1930's, and the 1970's. This would imply extremely low and perhaps even negative returns during the adjustment period, and then higher returns afterwards.

The unprecedented nature of recent stock market behavior makes it impossible to base forecasts on historical patterns alone. One must also form a view about what happened to drive stock prices up during the 1980's and particularly the 1990's. One view is that there has been a structural decline in the equity premium, driven either by the correction of mistaken perceptions of risk (aided perhaps by the work of economists on the equity premium puzzle), or by the reduction of barriers to participation and diversification by small investors.[4] Economists such as McGrattan and Prescott (2001) and Jagannathan, McGrattan, and Scherbina (2001) argue that the structural equity premium is now close to zero, consistent with theoretical models in which investors effectively share risks and have modest risk aversion, and consistent with the view that the US market has reached a new steady state.

An alternative view is that the equity premium has declined only temporarily, either because investors irrationally overreacted to positive fundamental news in the 1990's (Shiller 2000), or because the strong economy made investors more tolerant of risk.[5] On this view the equity premium will return to historical levels, implying extremely poor near-term returns and higher returns in the more distant future after traditional valuations have been restored.

It is too soon to tell which of these views is correct, and I believe it is sensible to put some weight on each of them. That is, I expect valuation ratios to return part way but not

[3]Vuolteenaho (2000) notes, however, that US corporations were unusually profitable in the late 1990's and that profitability has some predictive power for future earnings growth.

[4]Heaton and Lucas (1999) model barriers of this sort. It is hard to get large effects of increased participation on stock prices unless initial participation levels are extremely low. Furthermore, one must keep in mind that what matters for pricing is the wealth-weighted participation rate, that is, the probability that a randomly selected dollar of wealth is held by an individual who can participate in the market. This is higher than the equal-weighted participation rate, the probability that a randomly selected individual can participate.

[5]Campbell and Cochrane (1999) present a model in which investors judge their well-being by their consumption relative to a recent average of past aggregate consumption. In this model investors are more risk-tolerant when consumption grows rapidly and they have a "cushion of comfort" relative to their minimum expectations. The Campbell-Cochrane model fits past cyclical variations in the stock market, which will likely continue in the future, but it is hard to explain the extreme recent movements using this model.

fully to traditional levels.[6] A rough guess for the long term, after the adjustment process is complete, might be a geometric average equity return of 5% to 5.5% or an arithmetic average return of 6.5% to 7%.

If equity returns are indeed lower on average in the future, it is likely that short-term and long-term real interest rates will be somewhat higher. That is, the total return to the corporate capital stock is determined primarily by the production side of the economy and by national saving and international capital flows; the division of total return between riskier and safer assets is determined primarily by investor attitudes towards risk. Reduced risk aversion then reduces the equity premium both by driving down the equity return and by driving up the riskless interest rate. The yield on long-term inflation-indexed Treasury securities (TIPS) is about 3.5%, while short-term real interest rates have recently averaged about 3%. Thus 3% to 3.5% would be a reasonable guess for safe real interest rates in the future, implying a long-run average equity premium of 1.5% to 2.5% in geometric terms or about 3% to 4% in arithmetic terms.

Finally, I note that it is tricky to use these numbers appropriately in policy evaluation. Average equity returns should never be used in base-case calculations without showing alternative calculations to reflect the possibilities that realized returns will be higher or lower than average. These calculations should include an alternative in which equities underperform Treasury bills. Even if the probability of underperformance is small over a long holding period, it cannot be zero or the stock market would be offering an arbitrage opportunity or "free lunch". Equally important, the bad states of the world in which underperformance occurs are heavily weighted by risk-averse investors. Thus policy evaluation should use a broad range of returns to reflect the uncertainty about long-run stock market performance.

[6]This compromise view also implies that negative serial correlation, or mean-reversion, is likely to remain a characteristic of stock returns in the 21st Century.

Bibliography

Campbell, John Y. and John H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy* 107, 205–251.

Campbell, John Y. and Robert J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", *Review of Financial Studies* 1, 195–228.

Campbell, John Y. and Robert J. Shiller, 2001, "Valuation Ratios and the Long-Run Stock Market Outlook: An Update", NBER Working Paper No. 8221.

Campbell, John Y. and Luis M. Viceira, 2001, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, forthcoming Oxford University Press, New York, NY.

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2000, *The Millennium Book: A Century of Investment Returns*, ABN Amro and London Business School, London.

Fama, Eugene and Kenneth R. French, 2000, "The Equity Premium", unpublished paper, University of Chicago and MIT.

Graham, Benjamin and David Dodd, 1934, *Security Analysis*, McGraw-Hill, New York, NY.

Heaton, John and Deborah Lucas, 1999, "Stock Prices and Fundamentals", *NBER Macroeconomics Annual* 213–242, MIT Press, Cambridge, MA.

Jagannathan, Ravi, Ellen R. McGrattan, and Anna Scherbina, 2001, "The Declining US Equity Premium", NBER Working Paper No. 8172.

Jorion, Philippe and William N. Goetzmann, 1999, "Global Stock Markets in the Twentieth Century", *Journal of Finance* 54, 953–980.

Liang, Nellie and Steven A. Sharpe, 1999, "Share Repurchases and Employee Stock Options and their Implications for S&P500 Share Retirements and Expected Returns", unpublished paper, Board of Governors of the Federal Reserve System.

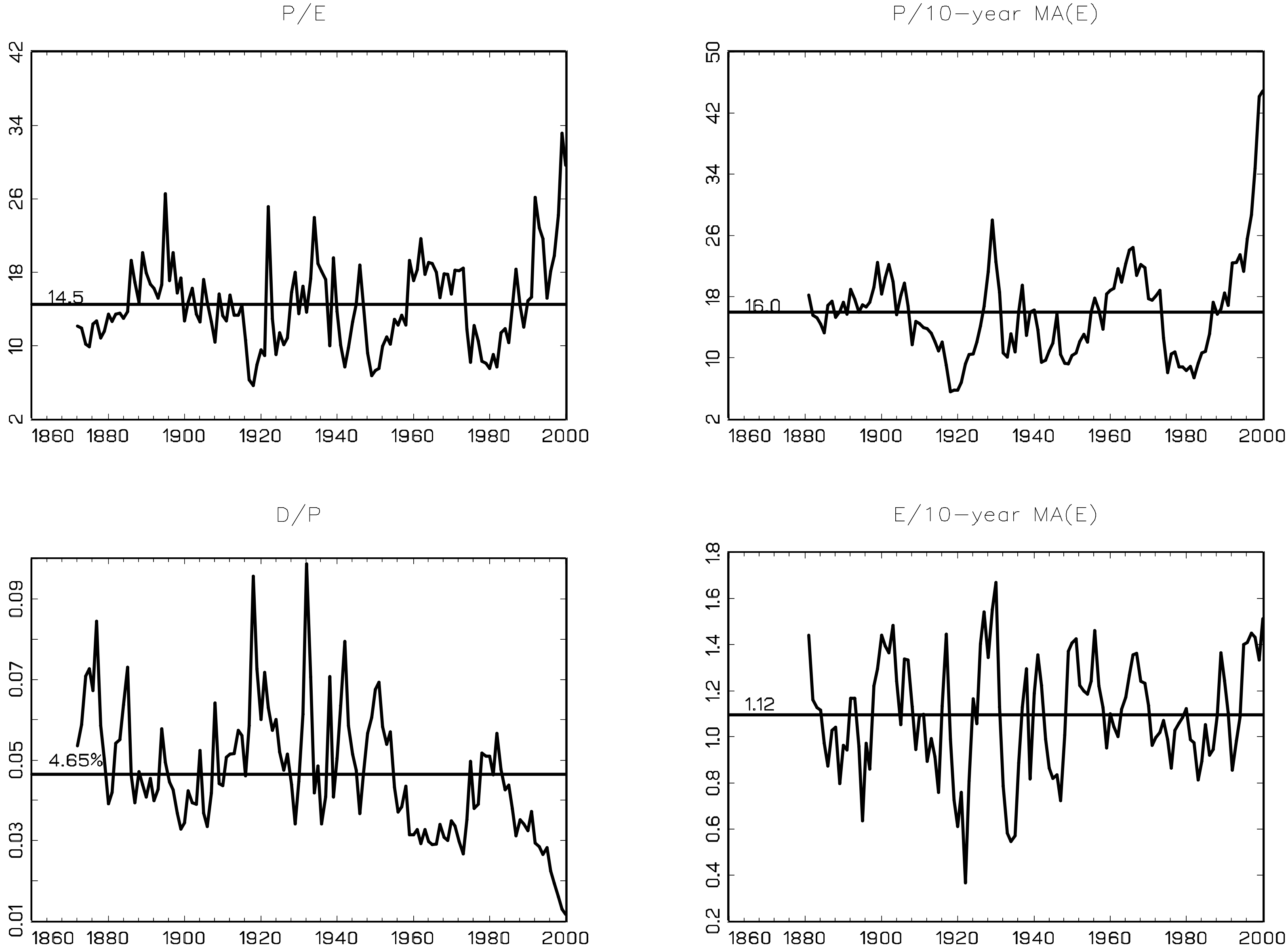
McGrattan, Ellen R., and Edward C. Prescott, 2001, "Is the Stock Market Overvalued?", NBER Working Paper No. 8077.

Shiller, Robert J., 2000, *Irrational Exuberance*, Princeton University Press, Princeton, NJ.

Siegel, Jeremy, 1998, *Stocks for the Long Run*, 2nd ed., McGraw-Hill, New York, NY.

Vuolteenaho, Tuomo, 2000, "Understanding the Aggregate Book-to-Market Ratio and its Implications to Current Equity-Premium Expectations", unpublished paper, Harvard University.

Figure 4. S&P Composite Stock Data, January Values 1872-1997



Viewpoint: Estimating the equity premium

John Y. Campbell *Department of Economics, Harvard University*

Abstract. Finance theory restricts the time-series behaviour of valuation ratios and links the cross-section of stock prices to the level of the equity premium. This can be used to strengthen the evidence for predictability in stock returns. Steady-state valuation models are useful predictors of stock returns, given the persistence in valuation ratios. A steady-state approach suggests that the world geometric average equity premium fell considerably in the late twentieth century, rose modestly in the early years of the twenty-first century, and was almost 4% at the end of March 2007. JEL classification: G12

Evaluer la prime des actions par rapport aux obligations. La théorie financière contraint le comportement diachronique des ratios de valorisation et relie transversalement les prix des actions au niveau de prime des actions sur les obligations. Voilà qui peut être utilisé pour renforcer la prédictibilité des rendements sur les actions. Les modèles de valorisation en régime permanent sont des prédicteurs utiles des rendements sur les actions, compte tenu du caractère stable des ratios de valorisation. Une approche en termes de régime permanent suggère que la moyenne géométrique mondiale de la prime des actions sur les obligations a chuté considérablement à la fin du 20e siècle, qu'elle a été modestement en hausse dans les premières années du 21e siècle, et qu'elle était à presque 4% à la fin de mars 2007.

The author is also affiliated with Arrowstreet Capital, LP, and NBER. This paper was presented in June 2007 as a State of the Art lecture at the Canadian Economics Association annual meeting at Dalhousie University in Halifax, Nova Scotia. A precursor was presented in January 2007 to the D-CAF Conference on Return Predictability at Copenhagen Business School. I am grateful to participants at both conferences, to John Cochrane, Jon Lewellen, Lubos Pastor, Ivo Welch, and Jeff Wurgler, and particularly to Angelo Melino for their thoughtful comments; to Bob Shiller, Moto Yogo, and my colleagues at Arrowstreet Capital, Sam Thompson and Tuomo Vuolteenaho, for joint research and many conversations on this subject; and to Alex Ogan, also of Arrowstreet Capital, for his able assistance with the data illustrated in figures 1 through 5.
Email: john_campbell@harvard.edu.

Canadian Journal of Economics / Revue canadienne d'Economique, Vol. 41, No. 1
February / février 2008. Printed in Canada / Imprimé au Canada

0008-4085 / 08 / 1–21 / © Canadian Economics Association

1. Introduction

What return should investors expect the stock market to deliver, above the interest rate on a safe short-term investment? In other words, what is a reasonable estimate of the equity premium?

This question is a basic one for investors who must decide how to allocate their portfolios to safe and risky assets. In the academic world, it has for over three decades played a central role in the development of asset pricing theory and financial econometrics. In the 1960s and 1970s, the efficient market hypothesis was interpreted to mean that the true equity premium was a constant. Investors might update their estimates of the equity premium as more data became available, but eventually these estimates should converge to the truth. This viewpoint was associated with the use of historical average excess stock returns to forecast future returns.

In the early 1980s, a number of researchers reported evidence that excess stock returns could be predicted by regressing them on lagged financial variables. In particular, valuation ratios that divide accounting measures of cash flow by market valuations, such as the dividend-price ratio, earnings-price ratio, or smoothed earnings-price ratio, appeared to predict returns. Value-oriented investors in the tradition of Graham and Dodd (1934) had always asserted that high valuation ratios are an indication of an undervalued stock market and should predict high subsequent returns, but these ideas did not carry much weight in the academic literature until authors such as Rozeff (1984), Fama and French (1988), and Campbell and Shiller (1988a,b) found that valuation ratios are positively correlated with subsequent returns and that the implied predictability of returns is substantial at longer horizons. Around the same time, several papers pointed out that yields on short- and long-term Treasury and corporate bonds are correlated with subsequent stock returns (Fama and Schwert 1977; Keim and Stambaugh 1986; Campbell 1987; Fama and French 1989).

These results suggested that the equity premium is not a constant number that can be estimated ever more precisely, but an unknown state variable whose value must be inferred at each point in time on the basis of observable data. Meanwhile, research in asset pricing theory made financial economists more comfortable with the idea that the equity premium can change over time even in an efficient market with rational investors, so that a time-varying equity premium does not necessarily require abandonment of the traditional paradigm of financial economics for a behavioural or inefficient-markets alternative. Campbell and Cochrane (1999), for example, showed that rational investors with habit formation preferences might become more averse to volatility in consumption and wealth, driving up the equilibrium equity premium, when the economy is weak.

During the 1990s, research continued on regressions predicting stock returns from valuation ratios (Kothari and Shanken 1997; Lamont 1998; Pontiff and Schall 1998) and interest rates (Hodrick 1992). However the 1990s also saw challenges to the new view that valuation ratios predict stock returns.

A first challenge came from financial econometricians, who began to express concern that the apparent predictability of stock returns might be spurious. Many of the predictor variables in the literature are highly persistent: Nelson and Kim (1993) and Stambaugh (1999) pointed out that persistence leads to biased coefficients in predictive regressions if innovations in the predictor variable are correlated with returns (as is strongly the case for valuation ratios, although not for interest rates). Under the same conditions the standard t-test for predictability has incorrect size (Cavanagh, Elliott, and Stock 1995). These problems are exacerbated if researchers are data mining, considering large numbers of variables and reporting only those results that are apparently statistically significant (Foster, Smith, and Whaley 1997; Ferson, Sarkissian, and Simin 2003). An active recent literature discusses alternative econometric methods for correcting the bias and conducting valid inference (Cavanagh, Elliott, and Stock 1995; Lewellen 2004; Torous, Valkanov, and Yan 2004; Campbell and Yogo 2006; Jansson and Moreira 2006; Polk, Thompson, and Vuolteenaho 2006; Ang and Bekaert 2007; Cochrane 2007).

A second challenge was posed by financial history. In the late 1990s valuation ratios were extraordinarily low, so regression forecasts of the equity premium became negative (Campbell and Shiller 1998). Yet stock returns continued to be high until after the turn of the millennium. Data from these years were sufficiently informative to weaken the statistical evidence for stock return predictability. Although low returns in the early 2000s have partially restored this evidence, Goyal and Welch (2003, 2007) and Butler, Grullon, and Weston (2005) have argued that overall, the out-of-sample forecasting power of valuation ratios is often worse than that of a traditional model predicting the equity premium using only the historical average of past stock returns.

The ultimate test of any predictive model is its out-of-sample performance. My personal experience using regression models to forecast stock returns in the late 1990s was humbling, although these models were partially vindicated by the stock market decline of the early 2000s. The lesson I draw from this experience is that one is more likely to predict stock returns successfully if one uses finance theory to reduce the number of parameters that must be freely estimated from the data and to restrict estimates of the equity premium to a reasonable range.

In the next section of this paper I show how finance theory can be used if one believes that valuation ratios, in particular the dividend-price ratio, are stationary around a constant mean. Even under stationarity, the persistence of valuation ratios has led researchers to concentrate on situations where valuation ratios have a root that is close to unity. In section 3 I discuss the limiting case where one believes that the dividend-price ratio follows a geometric random walk. I show that this case allows an even larger role for theory: it implies that one should forecast returns by adding a growth estimate to the dividend-price ratio, in the manner of the classic Gordon growth model. I argue that this approach has historically generated successful out-of-sample forecasts and is likely to do so in the future as well. In section 4 I apply this methodology to estimate the current equity premium for Canada, for the U.S., and for the world stock market as a whole. In section 5 I briefly discuss how finance theory can be used to predict the equity premium from the cross-section of stock prices. Section 6 concludes.

2. Regression-based return prediction with a stationary dividend-price ratio

When the dividend-price ratio is stationary, a basic tool for analysing stock returns is the loglinear approximate relation derived by Campbell and Shiller (1988a). This relation says that the log stock return r_{t+1}, the log stock price p_t, and the log dividend d_t approximately satisfy

$$\begin{aligned} r_{t+1} &= k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t \\ &= k + (d_t - p_t) + \Delta d_{t+1} - \rho(d_{t+1} - p_{t+1}), \end{aligned} \tag{1}$$

where ρ is a coefficient of loglinearization equal to the reciprocal of one plus the steady-state level of the dividend-price ratio. Thus ρ is slightly smaller than one; for annual U.S. data, $\rho = 0.96$ is a reasonable value, given an average dividend-price ratio in the late twentieth century of about 4% or 0.04 in levels. This equation says that proportional changes in stock prices have a larger effect on returns than equal proportional changes in dividends, because the level of dividends is small relative to the level of prices.

Equation (1) is a difference equation for the log dividend-price ratio. Solving it forward, imposing a condition that there are no explosive bubbles in stock prices, and taking expectations at time t allows us to interpret the dividend-price ratio as

$$d_t - p_t = \frac{k}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j [r_{t+1+j} - \Delta d_{t+1+j}]. \tag{2}$$

This formula delivers a number of insights. First, it helps to motivate regressions of stock returns on the log dividend-price ratio. The ratio is a linear combination of discounted expectations of future stock returns and dividend growth. If dividend growth is not too predictable (and there is little direct evidence for long-term dividend predictability in U.S. data), and if the dynamics of discount rates are such that short- and long-term expected stock returns are highly correlated, then the log dividend-price ratio should be a good proxy for the expected stock return over the next period.

Second, equation (2) shows that in the absence of price bubbles, the log dividend-price ratio will be stationary if stock returns and dividend growth are stationary, conditions that seem quite plausible. In particular, if returns and dividend growth rates do not have time trends, then the log dividend-price ratio will not have a time trend either. (This model cannot be used to say what would happen if there were time trends in returns or dividend growth rates, because such trends would invalidate the linear approximation (1).) Third, however, persistent variation in returns or dividend growth rates can lead to persistent variation in the log dividend-price ratio even if that ratio is stationary.

The effect of persistence on predictive regressions has been highlighted by Stambaugh (1999). Stambaugh discusses the two-equation system,

$$r_{t+1} = \alpha + \beta x_t + u_{t+1} \tag{3}$$

$$x_{t+1} = \mu + \phi x_t + \eta_{t+1}, \tag{4}$$

where x_t can be any persistent predictor variable but attention focuses on the level or log of the dividend-price ratio.

OLS estimates of equation (3) in twentieth-century U.S. data, with the log dividend-price ratio $x_t = d_t - p_t$ as the explanatory variable and the annualized stock return as the dependent variable, tend to deliver estimates in the range 0.1 to 0.2. An estimate of 0.04, the historical average level of the dividend-price ratio, would imply that around the average, a percentage point increase in the level of the dividend-price ratio increases the expected stock return by one percentage point. The OLS estimates imply a sensitivity of the return to the dividend-price ratio that is several times greater than this. They imply that when the dividend-price ratio is unusually high, it tends to return to normal through increases in prices that magnify the effect on stock returns. Campbell and Shiller (1998) emphasize this pattern in the historical data.

To understand Stambaugh's concern about persistence, define

$$\gamma = \frac{\sigma_{u\eta}}{\sigma_\eta^2}. \tag{5}$$

The coefficient γ is the regression coefficient of return innovations on innovations to the predictor variable. In the case where the explanatory variable is the log dividend-price ratio, γ is negative because rising stock prices tend to be associated with a falling dividend-price ratio. More precisely, dividend growth is only weakly correlated with and much less volatile than stock returns, so from equation (1) γ is about $-\rho$, that is, slightly greater than -1.

Stambaugh points out that the bias in estimating the coefficient β is γ times the bias in estimating the persistence of the predictor variable, ϕ:

$$\mathrm{E}[\hat{\beta} - \beta] = \gamma \mathrm{E}[\hat{\phi} - \phi]. \tag{6}$$

This is significant because it has been understood since the work of Kendall (1954) that there is downward bias in estimates of ϕ of about $-(1 + 3\phi)/T$, where T is the sample size, primarily resulting from the fact that x_t has an unknown mean that must be estimated. With a highly persistent predictor variable and γ slightly greater than -1, the Stambaugh bias in $\hat{\beta}$ is almost $4/T$. With 50 years of data the bias is almost 0.08, substantial relative to the OLS estimates discussed above.

Recent responses to Stambaugh's critique have all used theory in one way or another. Lewellen (2004) first writes an expression for the bias conditional on the estimated persistence $\hat{\phi}$ and the true persistence ϕ:

$$\mathrm{E}[\hat{\beta} - \beta \mid \hat{\phi}, \phi] = \gamma[\hat{\phi} - \phi]. \tag{7}$$

At first sight this expression does not seem particularly useful because we do not know the true persistence coefficient. However, Lewellen argues on the basis of theory that ϕ cannot be larger than one – the dividend-price ratio is not explosive – so the largest bias occurs when $\phi = 1$. He proposes the conservative approach of adjusting the estimated coefficient using this worst-case bias:

$$\hat{\beta}_{adj} = \hat{\beta} - \gamma(\hat{\phi} - 1). \tag{8}$$

In the data, the log dividend-price ratio appears highly persistent. That is, $\hat{\phi}$ is close to one; Lewellen reports a monthly estimate of 0.997 for the period 1946–2000, or about 0.965 on an annual basis. Lewellen's bias adjustment is therefore about 0.035, much smaller than Stambaugh's bias adjustment for a 50-year sample and somewhat smaller whenever the sample size is less than 114 years. Lewellen argues that stock returns are indeed predictable from the log dividend-price ratio, almost as much so as a naive researcher, unaware of Stambaugh's critique, might believe. Another way to express Lewellen's point is that data samples with spurious return predictability are typically samples in which the log dividend-price ratio appears to mean-revert more strongly than it truly does. In the historical data, the log dividend-price ratio has a root very close to unity – it barely seems to mean-revert at all – and thus we should not expect important spurious predictability in the historical data.

Cochrane (2007) responds to Stambaugh by directing attention to the inability of the log-dividend price ratio to forecast dividend growth. At first sight this response does not seem connected to Lewellen's, but in fact it is closely related. The Campbell-Shiller loglinearization (1) implies that r_{t+1}, Δd_{t+1}, $d_{t+1} - p_{t+1}$, and $d_t - p_t$ are deterministically linked. It follows that if we regress r_{t+1}, Δd_{t+1}, and $d_{t+1} - p_{t+1}$ onto $d_t - p_t$, the coefficients β, β_d, and ϕ are related by

$$\beta = 1 - \rho\phi + \beta_d, \tag{9}$$

where ρ is the coefficient of loglinearization from equation (1).

If we have prior knowledge about ϕ, then β and β_d are linked. For example, if $\rho = 0.96$ and we know that $\phi \leq 1$, then $\beta_d \leq \beta - 0.04$. If $\beta = 0$, then β_d must be negative and less than -0.04. The fact that regression estimates of β_d are close to zero is therefore indirect evidence that $\beta > 0$, in other words that stock returns are predictable – given our prior knowledge, based on theory, that the log dividend-price ratio is not explosive.

Another way to express Cochrane's point is that if the dividend-price ratio fails to predict stock returns, it will be explosive unless it predicts dividend growth. Since the dividend-price ratio cannot be explosive, the absence of predictable dividend growth strengthens the evidence for predictable returns.

Campbell and Yogo (2006) offer a third response to Stambaugh. They point out that if we knew persistence, we could reduce noise by adding the innovation to the predictor variable to the predictive regression, estimating

$$r_{t+1} = \alpha' + \beta x_t + \gamma(x_{t+1} - \phi x_t) + v_{t+1}. \tag{10}$$

The additional regressor, $(x_{t+1} - \phi x_t) = \eta_{t+1}$, is uncorrelated with the original regressor x_t but correlated with the dependent variable r_{t+1}. Thus, the regression (10) still delivers a consistent estimate of the original predictive coefficient β, but it does so with increased precision because it controls for some of the noise in unexpected stock returns.

Of course, in practice we do not know the persistence coefficient ϕ, but Campbell and Yogo argue that we can construct a confidence interval for it by inverting a unit root test. By doing this we 'de-noise' the return and get a more powerful test. The test delivers particularly strong evidence for predictability if we rule out a persistence coefficient $\phi > 1$ on prior grounds.

A way to understand Campbell and Yogo's results is to recall the challenge posed by the late 1990s. In that period, the dividend-price ratio was low, which led Campbell and Shiller (1998) to predict low stock returns based on a regression like (3). In fact, stock returns remained high until the early 2000s. These high returns were accompanied by falling dividend yields, despite the fact that the dividend yield was already below its historical mean. If we believe that the dividend yield was below its true mean and that it should be forecast to return to that mean rather than exploding away from it, then the late 1990s declines in the dividend-price ratio must have been unexpected. Unexpected declines in the dividend-price ratio are associated with unexpected high stock returns, accounting for the poor performance of the basic predictability regression in the late 1990s. The regression (10) corrects for this effect, limiting the negative influence of the late 1990s on the estimated predictive coefficient β.

The econometric issues discussed in this section have little effect on regressions that use nominal interest rates or yield spreads to predict excess stock returns. Although nominal interest rates are highly persistent, their innovations are not strongly correlated with innovations in stock returns, and thus the coefficient γ is close to zero for these variables, implying only a trivial bias in OLS regression estimates. Even papers that are sceptical of stock return predictability from the dividend-price ratio, such as Ang and Bekaert (2007), emphasize the strength of the statistical evidence that interest rates predict stock returns. The challenge in this case is primarily a theoretical one: to understand the economic forces that cause common variation in nominal interest rates and the equity premium.

All the papers discussed above combine prior knowledge with classical statistical methods. It is possible, of course, to use finance theory in an explicit Bayesian manner. Several recent papers have done this, notably Pastor and Stambaugh (2007) and Wachter and Warusawitharana (2007). Consistent with the results reported here, these papers find that tight priors on the persistence of the predictor variable tend to deliver stronger evidence for predictability of stock returns.

3. Steady-state return prediction

The papers discussed in the previous section address the question of whether the equity premium varies with market valuations, or whether it is constant. Even if one believes that the equity premium is time varying, however, there remains the important question of how best to estimate it at each point in time. Given the noise in stock returns, equity premium models with multiple free coefficients are hard to estimate and may fail out of sample because of errors in estimating the coefficients. Indeed, Goyal and Welch (2007) argue that almost all the regression models proposed in the recent literature fail to beat the historical sample mean when predicting excess stock returns out of sample.

In response to Goyal and Welch, Campbell and Thompson (2007) propose to use steady-state valuation models to estimate the equity premium. Such models tightly restrict the way in which historical data are used to predict future returns, and Campbell and Thompson find that they work well out of sample. Fama and French (2002) and Pastor, Sinha, and Swaminathan (2007) also use this approach to analyse the equity premium. The approach is analogous to the familiar procedure of forecasting the return on a bond, using its yield rather than its historical average return.

The classic steady-state model is the Gordon growth model, named after Canadian economist Myron Gordon. The model describes the level of the dividend-price ratio in a steady state with a constant discount rate and growth rate. Using upper-case letters to denote levels of variables, the Gordon growth model can be written as

$$\frac{D}{P} = R - G. \tag{11}$$

This formula can be used directly with historical dividend growth rates, but it can also be rewritten in several ways that suggest alternative empirical strategies for forecasting stock returns. First, one can substitute out growth by using the steady-state relation between growth and accounting return on equity,

$$G = \left(1 - \frac{D}{E}\right) ROE, \tag{12}$$

where D/E is the payout ratio, to obtain a growth-adjusted return forecast

$$\hat{R}_{DP} = \frac{D}{P} + \left(1 - \frac{D}{E}\right) ROE. \tag{13}$$

This return forecast is linear in D/P, with a slope coefficient of one and an intercept that is determined by the reinvestment rate and profitability. Importantly, neither the slope coefficient nor the intercept need to be estimated from noisy historical stock returns.

Second, one can restate the model in terms of the earnings-price ratio by using $D/P = (D/E)(E/P)$ to obtain

$$\hat{R}_{EP} = \left(\frac{D}{E}\right) \frac{E}{P} + \left(1 - \frac{D}{E}\right) ROE, \tag{14}$$

a payout-ratio-weighted average of the earnings-price ratio and the accounting return on equity. When return on equity equals the expected return, as might be the case in long-run equilibrium, then this implies that $\hat{R}_{EP} = E/P$.

Finally, one can rewrite the model in terms of the book-market ratio. Since $E/P = (B/M)ROE$,

$$\hat{R}_{BM} = ROE\left[1 + \frac{D}{E}\left(\frac{B}{M} - 1\right)\right]. \tag{15}$$

To use these formulas in practice, one must decide how to combine historical and contemporaneous data on the right-hand-side variables. Campbell and Thompson (2007) follow Fama and French (2002) by using historical average data on payouts and profitability, but differ from them by using current rather than historical average data on valuation ratios to obtain a return forecast conditional on the market's current valuation level. This procedure assumes that movements in valuation ratios, relative to historical cash flows, are explained by permanent changes in expected returns, so that each percentage point increase in the level of the dividend-price ratio generates a percentage point increase in the return forecast. It is a compromise between the view that valuation ratios are driven by changing forecasts of profitability, in which case the implied movements in returns would be smaller, and the view that valuation ratios are driven by *temporary* changes in discount rates, in which case the implied return movements would be larger, as discussed in the previous section.

Campbell and Thompson evaluate the out-of-sample performance of these models and several other variants over the period 1927–2005 and subsamples with breakpoints at 1956 and 1980. They find that steady-state valuation models typically perform better when more theoretical restrictions are imposed, and that they almost always outperform the historical mean return as a predictor of future returns. Dividend-based and earnings-based models, equations (13) and (14), generally appear to be more successful than the book-market model (15). In the next section I illustrate this approach using a model that averages both the dividend-price ratio and the recent history of earnings to generate a return forecast that is a blend of those from (13) and (14).

3.1. The Gordon model with a random walk dividend-price ratio

It may at first sight appear strange that steady-state valuation models based on the Gordon growth model perform well, given that they assume constant valuation ratios, while in the data valuation ratios vary in a highly persistent manner. It turns out, however, that a variant of the Gordon growth model can be derived using the assumption that the log dividend-price ratio follows a random walk. Under this assumption the Campbell-Shiller loglinear model, used in the previous section, breaks down because the dividend-price ratio has no fixed mean around which to take a loglinear approximation. However, in this case a suitable version of the original Gordon growth model is available to take the place of the Campbell-Shiller model.

To show this I assume, as in the Gordon growth model, that the dividend is known one period in advance. Then we can write

$$\frac{D_{t+1}}{P_t} = \exp(x_t), \tag{16}$$

where x_t now denotes the log dividend-price ratio using a forward or indicated dividend rather than a historical dividend. I assume that x_t follows a random walk:

$$x_t = x_{t-1} + \varepsilon_t. \tag{17}$$

Since the dividend growth rate is known one period in advance, I can write

$$\frac{D_{t+1}}{D_t} = 1 + G_t = \exp(g_t). \tag{18}$$

Finally, I assume that x_{t+1} and g_{t+1} are conditionally normal given time t information.

The definition of the stock return implies that

$$\begin{aligned} 1 + R_{t+1} &= \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{D_{t+1}}{P_t} + \frac{D_{t+2}}{D_{t+1}}\frac{D_{t+1}}{P_t}\left(\frac{D_{t+2}}{P_{t+1}}\right)^{-1} \\ &= \exp(x_t)[1 + \exp(g_{t+1} - x_{t+1})]. \end{aligned} \tag{19}$$

The conditionally expected stock return can be calculated using the formula for the conditional expectation of lognormally distributed random variables and the martingale property that $\mathrm{E}_t\, x_{t+1} = x_t$:

$$\begin{aligned} E_t(1+R_{t+1}) &= \exp(x_t)[1+\mathrm{E}_t \exp(g_{t+1}-x_{t+1})] \\ &= \exp(x_t)\left[1+\exp\left(\mathrm{E}_t g_{t+1}-x_t+\sigma_g^2/2+\sigma_x^2/2-\sigma_{gx}\right)\right] \\ &= \frac{D_{t+1}}{P_t}+\exp(\mathrm{E}_t g_{t+1})\exp(\mathrm{Var}_t(p_{t+1}-p_t)/2). \end{aligned} \tag{20}$$

Finally, the right-hand side of (20) can be approximated using the facts that for small y, $\exp(y) \approx 1+y$, and that unexpected log stock returns are approximately equal to unexpected changes in log stock prices:

$$E_t(1+R_{t+1}) \approx \frac{D_{t+1}}{P_t}+\exp(\mathrm{E}_t g_{t+1})+\frac{1}{2}\mathrm{Var}_t(r_{t+1}). \tag{21}$$

This equation expresses the expected stock return as the level of the dividend yield, plus geometric average dividend growth, plus one-half the variance of stock returns. In the original Gordon model, $\sigma_x^2=0$, so the variance of stock returns equals the variance of dividend growth. Since arithmetic average dividend growth equals geometric average dividend growth plus one-half the variance of dividend growth, in this case we get the original Gordon formula that the arithmetic average stock return equals dividend yield plus arithmetic average dividend growth.

If one subtracts half the variance of stock returns from each side of (20), one finds that the geometric average stock return equals the level of the dividend-price ratio plus the geometric average of dividend growth. Under the assumptions of the original Gordon model, the geometric implementation of the model is equivalent to an arithmetic implementation because stock returns and dividend growth have the same variance, so their geometric and arithmetic averages differ by the same amount. In the data, however, returns are much more volatile, so the geometric implementation and the arithmetic implementation are different. The analysis here shows that the geometric implementation is correct. Interestingly, this is exactly the way in which the model is used by Siegel (1994).

4. What is the equity premium today?

I now use a version of the above methodology, starting from equation (14), to estimate the equity premium. Following the previous discussion, I first estimate the conditional geometric average stock return, then subtract the real interest rate to get an equity premium number, and finally discuss the adjustment that is needed to convert from a geometric average to an arithmetic average equity premium. I look at data for the world as a whole (measured using the Morgan Stanley Capital International all-world index), and also for the U.S. and Canada, over the period from 1982 through the end of March 2007.

Figure 1 shows that for all three indices smoothed earnings-price ratios, with earnings smoothed over three years to eliminate cyclical noise, have fallen

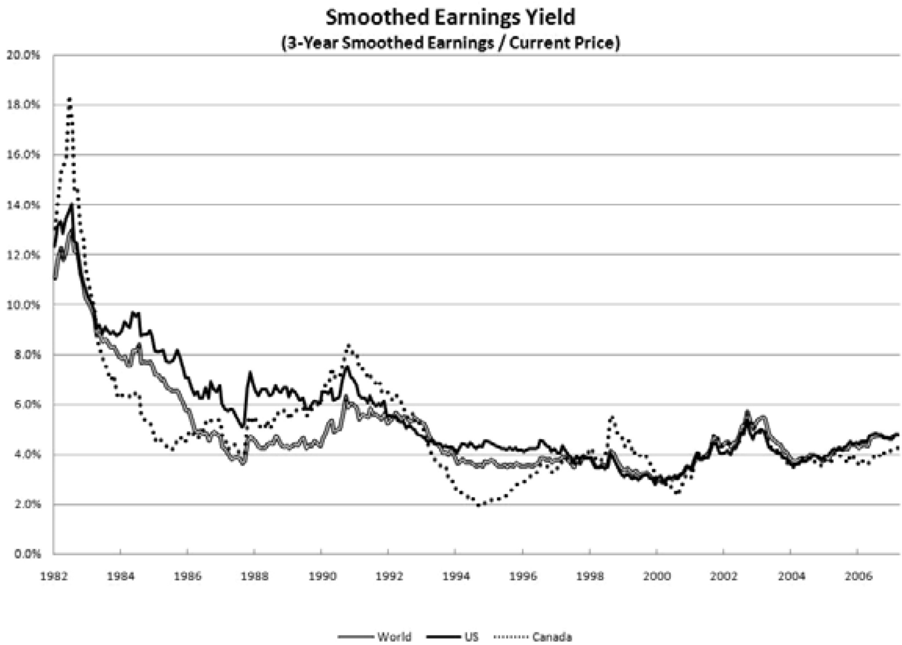


FIGURE 1 Three-year smoothed earnings-price ratios in the world, the U.S., and Canada

dramatically since the early 1980s and have been in the 3% to 5% range for the last ten years. During the same period, however, figure 2 shows that profitability has increased from a long-run historical average of around 6% to much higher values around 10%. Meanwhile, payout ratios have fluctuated widely around an average of about 50%.

In constructing a return forecast, it is desirable to combine historical earnings with some forward-looking measure of earnings. One possibility is to use analysts' earnings forecasts (Pastor, Sinha, and Swaminathan 2007); another is to use dividends. I average historical earnings, smoothed over three years, and the current dividend, divided by the payout rate, to construct a forward-looking measure of permanent earnings that can be used in equation (14).

When I put these numbers together, an earnings-based estimate of the real return on U.S. equities, assuming constant 6% real profitability and a 50% payout rate, was about 9% in the early 1980s and fell to just above 4% in the year 2000. Since then it has increased to slightly over 5%. This estimate assumes that profitability and payouts are best forecast to be constant; alternatively, if one uses the three-year moving average of profitability illustrated in figure 2, and a similar three-year moving average of the payout ratio, the current real return estimate increases by almost 4% to 9%, reflecting the high recent profitability and low payout ratios of U.S. corporations. At the world level, the current real return number is comparable to the U.S. number if a fixed profitability estimate

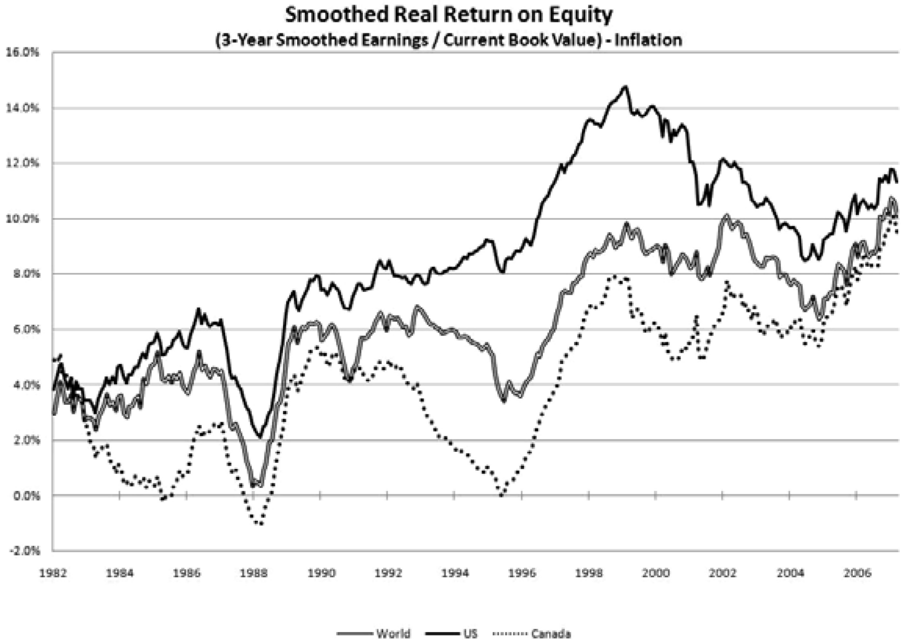


FIGURE 2 Three-year smoothed profitability in the world, the U.S., and Canada

is used, but the adjustment for recent profitability and payouts is much smaller, only slightly above 2%. The Canadian real return number is also very similar to that in the U.S. on the basis of fixed profitability, but lower Canadian profitability and higher payouts in the last few years imply that the use of recent data increases the estimated real return by less than 2%.

To convert these numbers into estimates of the equity premium, one needs to subtract a safe real interest rate. Figure 3 plots real yields on inflation-indexed bonds in three large markets, the U.K., the U.S., and Canada. The figure shows that the average real yield on inflation-indexed bonds across the three countries was about 3.5% in the 1990s but fell below 2% in the early 2000s. By the end of March 2007, it had recovered to just over 2%.

The implied current equity premium, assuming constant profitability and payouts, is just over 3%: 3.3% for the world as a whole, 3.2% for the U.S., and 3.1% for Canada. If instead one uses recent profitability and payouts, the current equity premium is 5.7% for the world as a whole, a startling 6.9% for the U.S., and 5.0% for Canada. Figures 4, 5, and 6 illustrate the history of the equity premium in the world, the U.S., and Canada under these two alternative assumptions.

Obviously a key question is whether the high profitability of global, and particularly U.S., corporations can be expected to continue. On the one hand, globalization has increased the supply of labour relative to capital, reducing wage pressure and increasing profitability; on the other hand, profitability has been increased

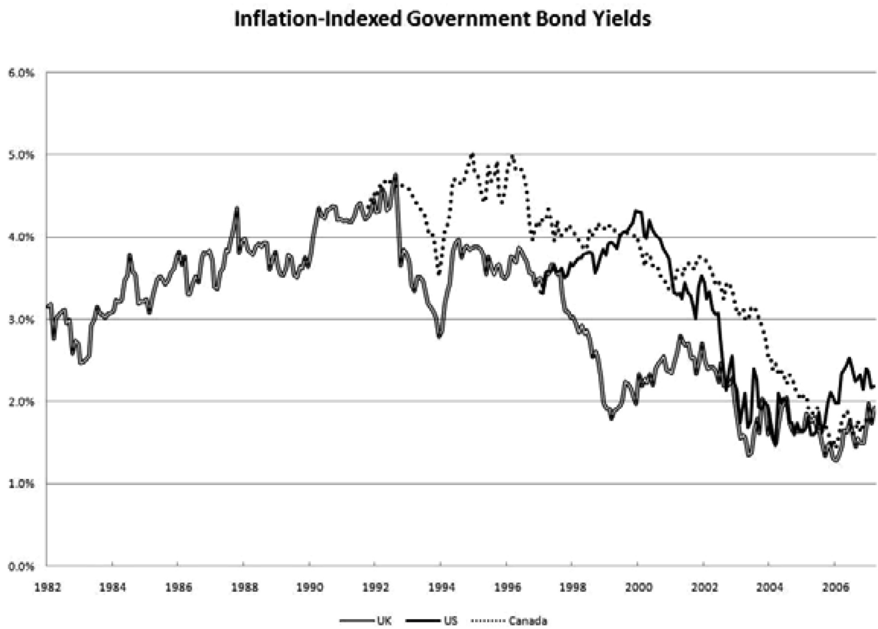


FIGURE 3 Long-term real interest rates in the U.K., the U.S., and Canada

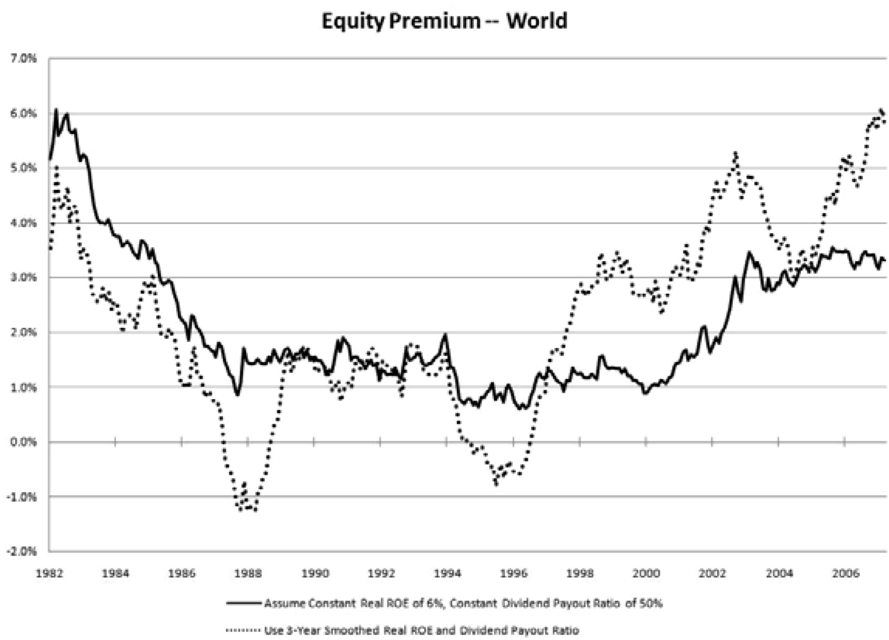


FIGURE 4 The world equity premium since 1982

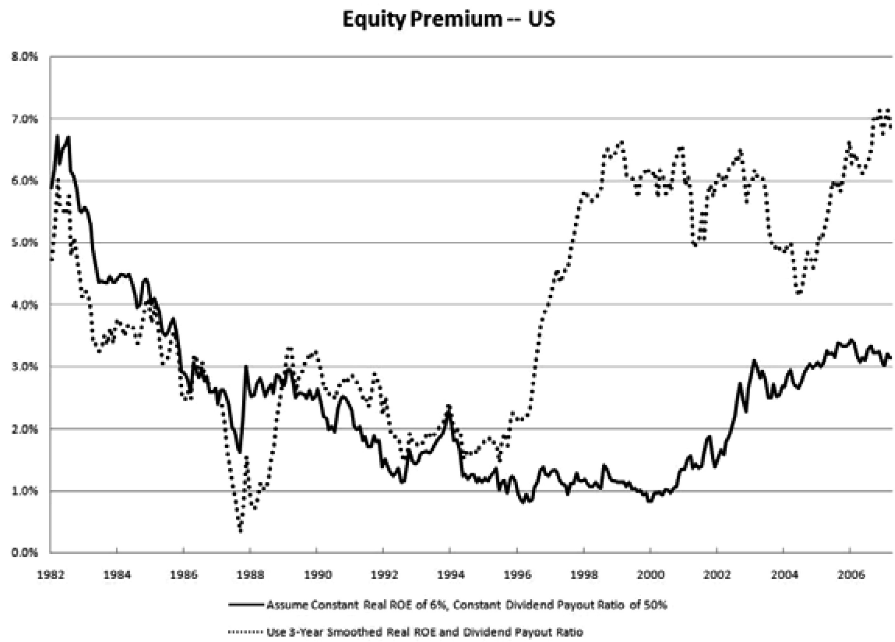


FIGURE 5 The U.S. equity premium since 1982

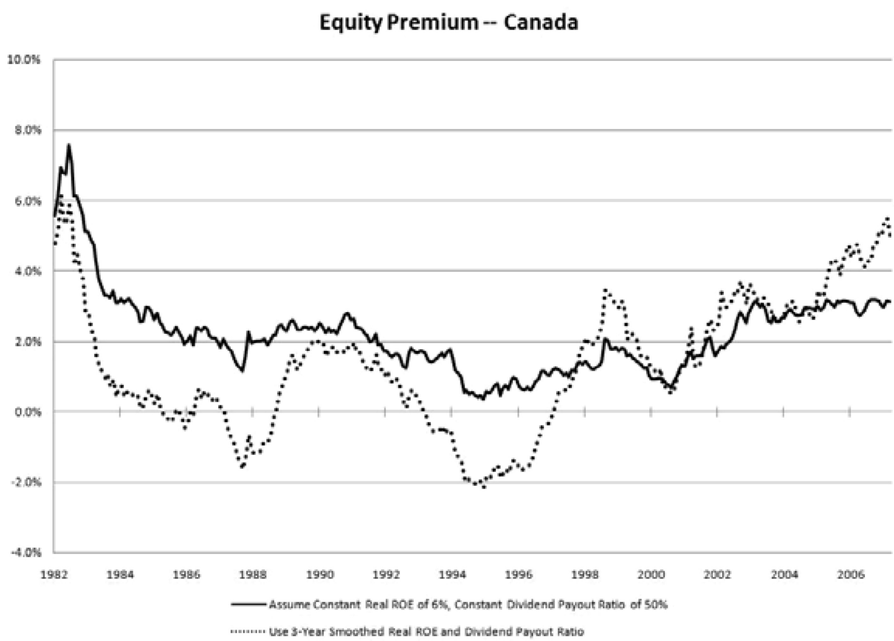


FIGURE 6 The Canadian equity premium since 1982

by favourable business cycle and political conditions that may not persist. Historically, profitability has shown temporary fluctuations and low payout rates (high reinvestment rates) have predicted declining profitability. Also, equity premium estimates based on current profitability and payout rates have been highly volatile, even turning negative on occasion. For both these reasons it seems wise to place considerably more weight on long-term averages than on recent data. If one puts a weight of 0.75 on the long-term average, with 0.25 on the recent data, the implied equity premium at the end of March 2007 is in the range 3.6% to 4.1%: 3.9% in the world as a whole, 4.1% in the U.S., and 3.6% in Canada. This number is a geometric average equity premium; for an arithmetic average, one should add one-half the variance of stock returns, or almost 1.3% if stock returns have a conditional standard deviation of 16%. The resulting arithmetic equity premium numbers are in the range 4.9% to 5.4%. Note that the equity premium is this high in large part because the safe real interest rate has declined over the past decade, as illustrated in figure 3.

These numbers are lower than historical average excess stock returns reported by Dimson, Marsh, and Staunton (2006). Using data for the period 1900–2005, Dimson, Marsh, and Staunton report geometric average equity premia of 4.7% for the world as a whole, 5.5% for the U.S., and 4.5% for Canada. The difference reflects two facts. First, historical average returns have been driven up by declining valuation ratios; this effect cannot be expected to continue in the future because valuation ratios should not have trends, a point emphasized by Fama and French (2002). Second, historical average returns were obtained by investors who paid lower stock prices and thus benefited from higher dividend-price ratios.

It is interesting to note that chief financial officers of major corporations, surveyed by Graham and Harvey (2007), have modest expectations of the equity premium, which implies that they do not expect recent profitability to continue. Their median estimate of the geometric average U.S. equity premium at the end of November 2006 was 3.4%, much closer to the constant-profitability number reported here than to the recent-profitability number and far below the historical average equity premium.

5. Return prediction with cross-sectional variables

Finance theory can also be used to predict excess stock returns using information in the cross-section of stock prices. This is valuable both to corroborate the predictions from aggregate valuation ratios and possibly as a way to pick up higher-frequency components of the equity premium that may be missed by a steady-state approach.

Polk, Thompson, and Vuolteenaho (2006) argue that if the Capital Asset Pricing Model (CAPM) is true, then a high equity premium implies low prices for stocks that have high betas with the aggregate market index. That is, high-beta stocks should be value stocks with low ratios of market prices to accounting measures of fundamental value. Reversing the argument, value stocks should tend to have high betas. This was true in the mid-twentieth-century, roughly from the 1930s through the 1950s, but in recent decades growth stocks have had higher betas than value stocks (Franzoni 2006). Polk, Thompson, and Vuolteenaho argue that this change in cross-sectional stock pricing reflects a decline in the equity premium. They construct a predictor of the aggregate market return, based on the relative pricing of high- and low-beta stocks, and show that it correlates well with the smoothed earnings-price ratio except in the early 1980s when inflation may have distorted the relationship.

It is possible to push this idea even further, exploiting the fact that the CAPM may not fully describe the cross-section of stock returns when returns are predictable in the time series. Merton (1973) developed an intertemporal CAPM (ICAPM) that showed that in the presence of time-varying expected returns, long-lived investors care not only about shocks to their wealth but also about shocks to the expected return on wealth. Intuitively, they value wealth not for its own sake but for the consumption stream it can provide; thus, they want to hedge against declines in the rate of return just as much as against declines in market value. Campbell (1993) implemented this idea using a vector autoregression (VAR) to break market movements into permanent movements driven by news about cash flows and temporary movements driven by news about discount rates. Long-lived investors are more concerned about the former than about the latter. Thus, stocks that covary with cash-flow news should have higher average returns than stocks that covary with discount-rate news, when betas with the overall market return are controlled for.

One of the main deviations from the CAPM in recent decades has been the value effect, the high average returns that value stocks have delivered despite their low market betas. If the ICAPM is to explain the value effect, it must be that value stocks covary with cash-flow news while growth stocks covary with discount-rate news. This implies that a moving average of past excess returns on growth stocks should be a good predictor of aggregate stock returns.

The value spread, the relative valuation of value and growth stocks (normally measured as the difference between the log book-market ratios of these two types of stocks) is one possible summary of past excess returns on growth stocks. Eleswarapu and Reinganum (2004) find that the value spread for small stocks predicts the aggregate market return, and Campbell and Vuolteenaho (2004) use the same variable in a VAR model to estimate and test the ICAPM. They find that the ICAPM explains the average returns of value and growth stocks much better than does the standard CAPM. Cohen, Polk, and Vuolteenaho (2006) and Campbell, Polk, and Vuolteenaho (2007) explore the robustness of these results, using both VAR-based and direct measures of cash-flow and discount-rate news. Empirically, the effect of including the small-stock value spread in a model of the equity premium is to lower the estimated equity premium at the turn of the millennium, when growth stocks were abnormally expensive relative to value stocks, and to increase it in 2006 and early 2007, when growth stocks were abnormally cheap.

All this work relies on theoretically motivated, but not fully restricted, time-series models of the aggregate market return. A natural next step is to use the theoretical restrictions of the ICAPM to jointly estimate a time-series model of the aggregate market return and a cross-sectional model of average stock returns. Campbell (1996) was an early implementation of this approach, but that paper did not find systematic deviations from the CAPM because it did not use the information in the relative prices of growth and value stocks. Recent research suggests that with the proper information variables and test assets, cross-sectional information can play an important role in a jointly estimated model of the equity premium.

6. Conclusion

In this paper I have tried to illustrate the usefulness of finance theory for statistical analysis of stock returns, in particular for estimation of the equity premium. The literature on this topic is vast, and inevitably I have neglected some important aspects. Five omissions deserve special mention.

First, I have not reviewed the simple but important point that excess stock returns should be difficult to predict, because highly predictable excess returns would imply extremely large profits for market-timing investors. Campbell and Thompson (2007) explore the mapping from R^2 statistics in predictive regressions to profits and welfare gains for market timers. The basic lesson is that investors should be suspicious of predictive regressions with high R^2 statistics, asking the old question, 'If you're so smart, why aren't you rich?'

Second, I have confined attention to short-term predictive regressions and have not considered direct forecasts of long-horizon returns. It has been known since Fama and French (1988) that long-horizon regressions often have higher R^2 statistics than short-horizon regressions, but their statistical properties are controversial. Campbell (2001) and Cochrane (2007) argue that in certain circumstances, long-horizon regressions can have superior power to detect predictability when in fact it exists.

Third, I have not discussed recent work that uses finance theory to infer the equity premium from the actions of market participants. Lettau and Ludvigson (2001), for example, argue that the level of consumption in relation to aggregate financial wealth and labour income reveals consumers' expectations of future stock returns. In a similar spirit Baker and Wurgler (2000) use the financing decisions of corporations to infer corporate managers' beliefs about expected stock returns.

Fourth, I have presented estimates of the equity premium without discussing the uncertainty of these estimates. I have suggested that finance theory can reduce our uncertainty about the equity premium, but a more formal Bayesian analysis would be needed to quantify this effect.

Finally, I have not attempted to review the important body of empirical work on the estimation of stock market risk. Mechanically, the volatility of stock returns determines the wedge between geometric and arithmetic average stock returns. Economically, both risk and return matter to investors, and it is plausible that changing risk is one factor that drives the changing equity premium. Merton (1980), Campbell (1987), French, Schwert, and Stambaugh (1987), Harvey (1989), and Glosten, Jagannathan, and Runkle (1993) are a few of the earlier papers that explore this relation. Recent contributions by Ghysels, Santa-Clara, and Valkanov (2005) and Pastor, Sinha, and Swaminathan (2007) find that the equity premium does covary positively with estimated risk, but that this effect does not explain the predictability of stock returns from valuation ratios or interest rates.

Despite the size and complexity of the literature on the equity premium, it has a simple unifying theme. Campbell, Lo, and MacKinlay (1997) argue that 'what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation.' Theory tells us why stock returns are so hard to predict. But it also holds out the promise of better prediction than we can hope to achieve by purely statistical forecasting methods.

References

Ang, Andrew, and Geert Bekaert (2007) 'Stock return predictability: is it there?' *Review of Financial Studies* 20, 651–707

Baker, Malcolm, and Jeffrey Wurgler (2000) 'The equity share in new issues and aggregate stock returns,' *Journal of Finance* 55, 2219–57

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael Roberts (2007) 'On the importance of measuring payout yield: Implications for empirical asset pricing,' *Journal of Finance* 62, 877–916

Butler, Alexander W., Gustavo Grullon, and James P. Weston (2005) 'Can managers forecast aggregate market returns?' *Journal of Finance* 60, 963–86

Campbell, John Y. (1987) 'Stock returns and the term structure,' *Journal of Financial Economics* 18, 373–99

— (1993) 'Intertemporal asset pricing without consumption data,' *American Economic Review* 83, 487–512

— (1996) 'Understanding risk and return,' *Journal of Political Economy* 104, 298–345

— (2001) 'Why long horizons? A study of power against persistent alternatives,' *Journal of Empirical Finance* 8, 459–91

Campbell, John Y., and John H. Cochrane (1999) 'By force of habit: A consumption-based explanation of aggregate stock market behavior,' *Journal of Political Economy* 107, 205–51

Campbell, John Y., and Robert J. Shiller (1988a) 'The dividend-price ratio and expectations of future dividends and discount factors,' *Review of Financial Studies* 1, 195–228

— (1988b), 'Stock prices, earnings, and expected dividends,' *Journal of Finance* 43, 661–76

— (1998), 'Valuation ratios and the long-run stock market outlook,' *Journal of Portfolio Management* 24(2), 11–26

Campbell, John Y., and Samuel B. Thompson (2007) 'Predicting excess stock returns out of sample: Can anything beat the historical average?' *Review of Financial Studies*, forthcoming

Campbell, John Y., and Tuomo O. Vuolteenaho (2004) 'Bad beta, good beta,' *American Economic Review* 94, 1249–75
Campbell, John Y., and Motohiro Yogo (2006) 'Efficient tests of stock return predictability,' *Journal of Financial Economics* 81, 27–60
Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (1997) *The Econometrics of Financial Markets* (Princeton, NJ: Princeton University Press)
Campbell, John Y., Christopher Polk, and Tuomo O. Vuolteenaho (2007) 'Growth or glamour? Fundamentals and systematic risk in stock returns,' working paper, Harvard University and London School of Economics
Cavanagh, Christopher L., Graham Elliott, and James H. Stock (1995) 'Inference in models with nearly integrated regressors,' *Econometric Theory* 11, 1131–47
Cochrane, John H. (2007) 'The dog that did not bark: a defense of return predictability,' *Review of Financial Studies*, forthcoming
Cohen, Randolph B., Christopher Polk, and Tuomo O. Vuolteenaho (2006) 'The price is (almost) right,' working paper, Harvard University and Northwestern University
Dimson, Elroy, Paul Marsh, and Michael Staunton (2006) 'The worldwide equity premium: a smaller puzzle,' working paper, London Business School
Eleswarapu, Venkat R. and Marc R. Reinganum (2004) 'The predictability of aggregate stock market returns: evidence based on glamour stocks,' *Journal of Business* 77, 275–94
Fama, Eugene F., and Kenneth R. French (1988) 'Dividend yields and expected stock returns,' *Journal of Financial Economics* 22, 3–25
— (1989) 'Business conditions and expected returns on stocks and bonds,' *Journal of Financial Economics* 25, 23–49
— (2002) 'The equity premium,' *Journal of Finance* 57, 637–59
Fama, Eugene F., and G. William Schwert (1977) 'Asset returns and inflation,' *Journal of Financial Economics* 5, 115–46
Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin (2003) 'Spurious regressions in financial economics?' *Journal of Finance* 58, 1393–413
Foster, F. Douglas, Tom Smith, and Robert E. Whaley (1997) 'Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R^2,' *Journal of Finance* 52, 591–607
Franzoni, Francesco (2006) 'Where is beta going? The riskiness of value and small stocks,' working paper, HEC School of Management
French, Kenneth R., G. William Schwert, and Robert F. Stambaugh (1987) 'Expected stock returns and volatility,' *Journal of Financial Economics* 19, 3–29
Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov (2005) 'There is a risk-return tradeoff after all,' *Journal of Financial Economics* 76, 509–48
Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle (1993) 'On the relation between the expected value and the volatility of the nominal excess return on stocks,' *Journal of Finance* 48, 1779–801
Goyal, Amit, and Ivo Welch (2003) 'Predicting the equity premium with dividend ratios,' *Management Science* 49, 639–54
— (2007) 'A comprehensive look at the empirical performance of equity premium prediction,' *Review of Financial Studies*, forthcoming
Graham, Benjamin, and David L. Dodd (1934) *Security Analysis*, 1st ed. (New York: McGraw-Hill)
Graham, John, and Campbell Harvey (2007) 'The equity risk premium in January 2007: evidence from the global CFO outlook survey,' working paper, Duke University
Harvey, Campbell (1989) 'Time-varying conditional covariances in tests of asset pricing models,' *Journal of Financial Economics* 24, 289–317
Hodrick, Robert J. (1992) 'Dividend yields and expected stock returns: alternative procedures for inference and measurement,' *Review of Financial Studies* 5, 257–86

Jansson, Michael, and Marcelo J. Moreira (2006) 'Optimal inference in regression models with nearly integrated regressors,' *Econometrica* 74, 681–715
Keim, Donald B., and Robert F. Stambaugh (1986) 'Predicting returns in the stock and bond markets,' *Journal of Financial Economics* 17, 357–90
Kendall, Maurice G. (1954) 'Note on bias in the estimation of autocorrelation,' *Biometrika* 41, 403–4
Kothari, S.P., and Jay Shanken (1997) 'Book-to-market, dividend yield, and expected market returns: a time-series analysis,' *Journal of Financial Economics* 44, 169–203
Lamont, Owen (1998) 'Earnings and expected returns,' *Journal of Finance* 53, 1563–87
Lettau, Martin, and Sydney Ludvigson (2001) 'Consumption, aggregate wealth, and expected stock returns,' *Journal of Finance* 56, 815–49
Lewellen, Jonathan (2004) 'Predicting returns with financial ratios,' *Journal of Financial Economics* 74, 209–35
Merton, Robert C. (1973) 'An intertemporal capital asset pricing model,' *Econometrica* 41, 867–87
— (1980) 'On estimating the expected return on the market: an exploratory investigation,' *Journal of Financial Economics* 8, 323–61
Nelson, Charles, and M. Kim (1993) 'Predictable stock returns: the role of small sample bias,' *Journal of Finance* 48, 641–61
Pastor, Lubos, and Robert F. Stambaugh (2007) 'Predictive systems: living with imperfect predictors,' NBER Working Paper 12814
Pastor, Lubos, Meenakshi Sinha, and Bhaskaran Swaminathan (2007) 'Estimating the intertemporal risk-return tradeoff using the implied cost of capital,' *Journal of Finance*, forthcoming
Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho (2006) 'Cross-sectional forecasts of the equity premium,' *Journal of Financial Economics* 81, 101–41
Pontiff, Jeffrey, and Lawrence D. Schall (1998) 'Book-to-market ratios as predictors of market returns,' *Journal of Financial Economics* 49, 141–60
Rozeff, Michael S. (1984) 'Dividend yields are equity risk premiums,' *Journal of Portfolio Management* 11(1), 68–75
Siegel, Jeremy (1994) *Stocks for the Long Run* (New York: McGraw-Hill)
Stambaugh, Robert F. (1999) 'Predictive regressions,' *Journal of Financial Economics* 54, 375–421
Torous, Walter, Rossen Valkanov, and Shu Yan (2004) 'On predicting stock returns with nearly integrated explanatory variables,' *Journal of Business* 77, 937–66
Wachter, Jessica A., and Missaka Warusawitharana (2007) 'Predictable returns and asset allocation: should a skeptical investor time the market?' NBER Working Paper No.13165

Estimating the Real Rate of Return on Stocks Over the Long Term

Papers by

John Y. Campbell
Peter A. Diamond
John B. Shoven

Presented to the
Social Security Advisory Board

August 2001

Social Security Advisory Board

An independent, bipartisan Board created by Congress and appointed by the President and the Congress to advise the President, the Congress, and the Commissioner of Social Security on matters related to the Social Security and Supplemental Security Income programs.

TABLE OF CONTENTS

INTRODUCTION

In recent years there have been a variety of proposals that would change the current Social Security system to include some form of investment of funds in private equities. These proposals include allowing or requiring individuals to use a portion of the payroll tax to fund individual investment accounts, either as part of the Social Security system or as an addition to it. They also include proposals to require the government to invest a portion of the Social Security Trust Funds in equities.

A key element in evaluating these proposals is the rate of return that can be expected on such investments. The members of the 1994-1996 Advisory Council on Social Security agreed to use a real annual rate of 7 percent (the average for the period 1900-1995) to compare the three plans put forward by the Council. The Office of the Chief Actuary (OCACT) of the Social Security Administration has continued to use 7 percent to evaluate proposals for investment in stocks. However, there is a question as to whether the historical rate for the last century should be used to make long-term projections over the coming decades or whether an alternative rate or range of rates is more appropriate.

This document includes papers by three distinguished economists that examine this important question, including the issue of how to reflect the higher risk inherent in stock investment relative to investment in U.S. Treasury securities. The papers are by John Campbell, Otto Eckstein Professor of Applied Economics at Harvard University; Peter Diamond, Institute Professor at the Massachusetts Institute of Technology; and John Shoven, Charles Schwab Professor of Economics at Stanford University. The Board is publishing them in order to make them available to policy makers and members of the public who are interested in the issue of how to ensure the long-term solvency of the Social Security system.

The papers (which have been updated for purposes of this document) were the basis for a discussion sponsored by the Social Security Advisory Board on May 31, 2001. The purpose of the discussion was to enable individuals from OCACT who have the responsibility of estimating the effects of changes in the Social Security system to hear a range of views on the likely real yields on equities over the long term. Participants in the discussion from OCACT included Stephen Goss, Chief Actuary; Alice Wade, Deputy Chief Actuary; Patrick Skirvin, Lead Economist; and Anthony Cheng, Economist.

Participants also included three other distinguished economists who were on the 1999 Technical Panel on Assumptions and Methods: Eugene Steuerle, Senior Fellow, The Urban Institute; Deborah Lucas, Professor of Finance, Northwestern University and currently Chief Economist, Congressional Budget Office; and Andrew Samwick, Assistant Professor of Economics, Dartmouth College. The 1999 Technical Panel, which was sponsored by the Advisory Board, was charged with reviewing the assumptions and methods used in the long-term projections of the Social Security Trust Funds. The Panel also examined the question of how to evaluate the returns and risks involved in stock market investments. The Panel's report was published by the Board in November 1999 and is available on the Board's Web site (www.ssab.gov).

Forecasting U.S. Equity Returns in the 21st Century

John Y. Campbell, Professor of Economics
Harvard University
July 2001

What returns should investors expect the U.S. stock market to deliver on average during the next century? Does the experience of the last century provide a reliable guide to the future? In this short note I first discuss alternative methodologies for forecasting average future equity returns, then discuss current market conditions, and finally draw conclusions for long-term return forecasts. Throughout I work in real, that is inflation-adjusted, terms.

I. Methods for Forecasting Returns

1. Average past returns

Perhaps the simplest way to forecast future returns is to use some average of past returns. Very naturally, this method has been favored by many investors and analysts. However there are several difficulties with it.

a) ***Geometric average or arithmetic average?*** The geometric average return is the cumulative past return on U.S. equities, annualized. Siegel (1998) studies long-term historical data on value-weighted U.S. share indexes. He reports a geometric average of 7.0% over two different sample periods, 1802-1997 and 1871-1997. The arithmetic average return is the average of one-year past returns on U.S. equities. It is considerably higher than the geometric average return, 8.5% over 1802-1997 and 8.7% over 1871-1997.[1]

When returns are serially uncorrelated, the arithmetic average represents the best forecast of future return in any randomly selected future year. For long holding periods, the best forecast is the arithmetic average compounded up appropriately. If one is making a 75-year forecast, for example, one should forecast a cumulative return of 1.085^{75} based on 1802-1997 data.

When returns are negatively serially correlated, however, the arithmetic average is not necessarily superior as a forecast of long-term future returns. To understand this, consider an extreme example in which prices alternate deterministically between 100 and 150. The return is 50% when prices rise, and -33% when prices fall. Over any even number of periods, the geometric average return is zero, but the arithmetic average return is 8.5%. In this case the arithmetic average return is misleading because it fails to take account of the fact that high returns always multiply a low initial price of 100, while low returns always multiply a high initial price of

---

[1] When returns are lognormally distributed, the difference between the two averages is approximately one-half the variance of returns. Since stock returns have an annual standard deviation of about 18% over these long periods, the predicted difference is $0.18^2/2=0.016$ or 1.6%. This closely matches the difference in the data.

150. The geometric average is a better indication of long-term future prospects in this example.[2]

This point is not just a theoretical curiosity, because in the historical data summarized by Siegel, there is strong evidence that the stock market is mean-reverting. That is, periods of high returns tend to be followed by periods of lower returns. This suggests that the arithmetic average return probably overstates expected future returns over long periods.

b) *Returns are very noisy.* The randomness in stock returns is extreme. With an annual standard deviation of real return of 18%, and 100 years of past data, a single year's stock return that is only one standard deviation above average increases the average return by 18 basis points. A lucky year that is two standard deviations above average increases the average return by 36 basis points. Even when a century or more of past data is used, forecasts based on historical average returns are likely to change substantially from one year to the next.

c) *Realized returns rise when expected returns fall.* To the extent that expected future equity returns are not constant, but change over time, they can have perverse effects on realized returns. Suppose for example that investors become more risk-tolerant and reduce the future return that they demand from equities. If expected future cash flows are unchanged, this drives up prices and realized returns. Thus an estimate of future returns based on average past realized returns will tend to increase just as expected future returns are declining.

Something like this probably occurred in the late 1990's. A single good year can have a major effect on historical average returns, and several successive good years have an even larger effect. But it would be a mistake to react to the spectacular returns of 1995-99 by increasing estimates of 21st Century returns.

d) *Unpalatable implications.* Fama and French (2000) point out that average past U.S. stock returns are so high that they exceed estimates of the return to equity (ROE) calculated for U.S. corporations from accounting data. Thus if one uses average past stock returns to estimate the cost of capital, the implication is that U.S. corporate investments have destroyed value; corporations should instead have been paying all their earnings out to stockholders. This conclusion is so hard to believe that it further undermines confidence in the average-return methodology.

One variation of the average-past-returns approach is worth discussing. One might take the view that average past equity returns in other countries provide relevant evidence about U.S. equity returns. Standard international data from Morgan Stanley Capital International,

[2] One crude way to handle this problem is to measure the annualized variance of returns over a period such as 20 years that is long enough for returns to be approximately serially uncorrelated, and then to adjust the geometric average up by one-half the annualized 20-year variance as would be appropriate if returns are lognormally distributed. Campbell and Viceira (2001, Figure 4.2) report an annualized 20-year standard deviation of about 14% in long-term annual U.S. data, which would imply an adjustment of $0.14^2/2=0.010$ or 1.0%.

available since the early 1970's, show that equity returns in most other industrialized countries have been about as high as those in the U.S. The exceptions are the heavily commodity-dependent markets of Australia and Canada, and the very small Italian market (Campbell 1999). Jorion and Goetzmann (1999) argue that other countries' returns were lower than U.S. returns in the early 20th Century, but this conclusion appears to be sensitive to their omission of the dividend component of return (Dimson, Marsh, and Staunton 2000). Thus the use of international data does not change the basic message that the equity market has delivered high average returns in the past.

2. Valuation ratios

An alternative approach is to use valuation ratios—ratios of stock prices to accounting measures of value such as dividends or earnings—to forecast future returns. In a model with constant valuation ratios and growth rates, the famous Gordon growth model says that the dividend-price ratio

$$\frac{D}{P} = R - G, \tag{1}$$

where R is the discount rate or expected equity return, and G is the growth rate of dividends (equal to the growth rate of prices when the valuation ratio is constant). This formula can be applied either to price per share and conventional dividends per share, or to the total value of the firm and total cash paid out by the firm (including share repurchases). A less well-known but just as useful formula says that in steady state, where earnings growth comes from reinvestment of retained earnings which earn an accounting ROE equal to the discount rate R,

$$\frac{E}{P} = R. \tag{2}$$

Over long periods of time summarized by Siegel (1998), these formulas give results consistent with average realized returns. Over the period 1802-1997, for example, the average dividend-price ratio was 5.4% while the geometric average growth rate of prices was 1.6%. These numbers add to the geometric average return of 7.0%. Over the period 1871-1997 the average dividend-price ratio was 4.9% while the geometric average growth rate of prices was 2.1%, again adding to 7.0%. Similarly, Campbell and Shiller (2001) report that the average P/E ratio for S&P 500 shares over the period 1872-2000 was 14.5. The reciprocal of this is 6.9%, consistent with average realized returns.

When valuation ratios and growth rates change over time, these formulas are no longer exactly correct. Campbell and Shiller (1988) and Vuolteenaho (2000) derive dynamic versions of the formulas that can be used in this context. Campbell and Shiller show, for example, that the log dividend-price ratio is a discounted sum of expected future discount rates, less a discounted sum of expected future dividend growth rates. In this note I will work with the simpler deterministic formulas.

II. Current Market Conditions

Current valuation ratios are wildly different from historical averages, reflecting the unprecedented bull market of the last 20 years, and particularly the late 1990's. The attached figure, taken from Campbell and Shiller (2001), illustrates this point. (See p. 9) The bottom left panel shows the dividend-price ratio D/P in January of each year from 1872-2000. The long-term historical average is 4.7%, but D/P has fallen dramatically since 1982 to about 1.2% in January 2000 (and 1.4% today).

The dividend-price ratio may have fallen in part because of shifts in corporate financial policy. An increased tendency for firms to repurchase shares rather than pay dividends increases the growth rate of dividends per share, by shrinking the number of shares. Thus it increases G in the Gordon growth formula and reduces conventionally measured D/P. One way to correct for this is to add repurchases to conventional dividends. Recent estimates of this effect by Liang and Sharpe (1999) suggest that it may be an upward adjustment of 75 to 100 basis points, and more in some years. Of course, this is not nearly sufficient to explain the recent decline in D/P.

Alternatively, one can look at the price-earnings ratio. The top left panel of the figure shows P/E over the same period. This has been high in recent years, but there are a number of earlier peaks that are comparable. Close inspection of these peaks shows that they often occur in years such as 1992, 1934, and 1922 when recessions caused temporary drops in (previous-year) earnings. To smooth out this effect, Campbell and Shiller (2001), following Graham and Dodd (1934), advocate averaging earnings over 10 years. The price-averaged earnings ratio is illustrated in the top right panel of the figure. This peaked at 45 in January 2000; the previous peak was 28 in 1929. The decline in the S&P 500 since January 2000 has only brought the ratio down to the mid-30's, still higher than any level seen before the late 1990's.

The final panel in the figure, on the bottom right, shows the ratio of current to 10-year average earnings. This ratio has been high in recent years, reflecting robust earnings growth during the 1990's, but it is not unprecedentedly high. The really unusual feature of the recent stock market is the level of prices, not the growth of earnings.

III. Implications for Future Returns

The implications of current valuations for future returns depend on whether the market has reached a new steady state, in which current valuations will persist, or whether these valuations are the result of some transitory phenomenon.

If current valuations represent a new steady state, then they imply a substantial decline in the equity returns that can be expected in the future. Using Campbell and Shiller's (2001) data, the unadjusted dividend-price ratio has declined by 3.3 percentage points from the historical average. Even adjusting for share repurchases, the decline is at least 2.3 percentage points. Assuming constant long-term growth of the economy, this would imply that the geometric average return on equity is no longer 7%, but 3.7% or at most 4.7%. Looking at the price-averaged earnings ratio, adjusting for the typical ratio of current to averaged earnings, gives an even lower estimate. Current earnings are normally 1.12 times averaged earnings; 1.12/35=0.032, implying a 3.2% return forecast. These forecasts allow for only a very modest equity premium relative to the yield on long-term inflation-indexed bonds, currently about 3.5%, or the 3% safe real return assumed recently by the Trustees.

If current valuations are transitory, then it matters critically what happens to restore traditional valuation ratios. One possibility is that earnings and dividends are below their long-run trend levels; rapid earnings and dividend growth will restore traditional valuations without any declines in equity returns below historical levels. While this is always a possibility, Campbell and Shiller (2001) show that it would be historically unprecedented. The U.S. stock market has an extremely poor record of predicting future earnings and dividend growth. Historically stock prices have increased relative to earnings during decades of rapid earnings growth, such as the 1920's, 1960's, or 1990's, as if the stock market anticipates that rapid earnings growth will continue in the next decade. However there is no systematic tendency for a profitable decade to be followed by a second profitable decade; the 1920's, for example, were followed by the 1930's and the 1960's by the 1970's. Thus stock market optimism often fails to be justified by subsequent earning growth.[3]

A second possibility is that stock prices will decline or stagnate until traditional valuations are restored. This has occurred at various times in the past after periods of unusually high stock prices, notably the 1900's and 1910's, the 1930's, and the 1970's. This would imply extremely low and perhaps even negative returns during the adjustment period, and then higher returns afterwards.

The unprecedented nature of recent stock market behavior makes it impossible to base forecasts on historical patterns alone. One must also form a view about what happened to drive stock prices up during the 1980's and particularly the 1990's. One view is that there has been a structural decline in the equity premium, driven either by the correction of mistaken perceptions of risk (aided perhaps by the work of economists on the equity premium puzzle), or by the reduction of barriers to participation and diversification by small investors.[4] Economists such as McGrattan and Prescott (2001) and Jagannathan, McGrattan, and Scherbina (2001) argue that the structural equity premium is now close to zero, consistent with theoretical models in which investors effectively share risks and have modest risk aversion, and consistent with the view that the U.S. market has reached a new steady state.

---

[3] Vuolteenaho (2000) notes, however, that U.S. corporations were unusually profitable in the late 1990's and that profitability has some predictive power for future earnings growth.

[4] Heaton and Lucas (1999) model barriers of this sort. It is hard to get large effects of increased participation on stock prices unless initial participation levels are extremely low. Furthermore, one must keep in mind that what matters for pricing is the wealth-weighted participation rate, that is, the probability that a randomly selected dollar of wealth is held by an individual who can participate in the market. This is higher than the equal-weighted participation rate, the probability that a randomly selected individual can participate.

An alternative view is that the equity premium has declined only temporarily, either because investors irrationally overreacted to positive fundamental news in the 1990's (Shiller 2000), or because the strong economy made investors more tolerant of risk.[5] On this view the equity premium will return to historical levels, implying extremely poor near-term returns and higher returns in the more distant future after traditional valuations have been restored.

It is too soon to tell which of these views is correct, and I believe it is sensible to put some weight on each of them. That is, I expect valuation ratios to return part way but not fully to traditional levels.[6] A rough guess for the long term, after the adjustment process is complete, might be a geometric average equity return of 5% to 5.5% or an arithmetic average return of 6.5% to 7%.

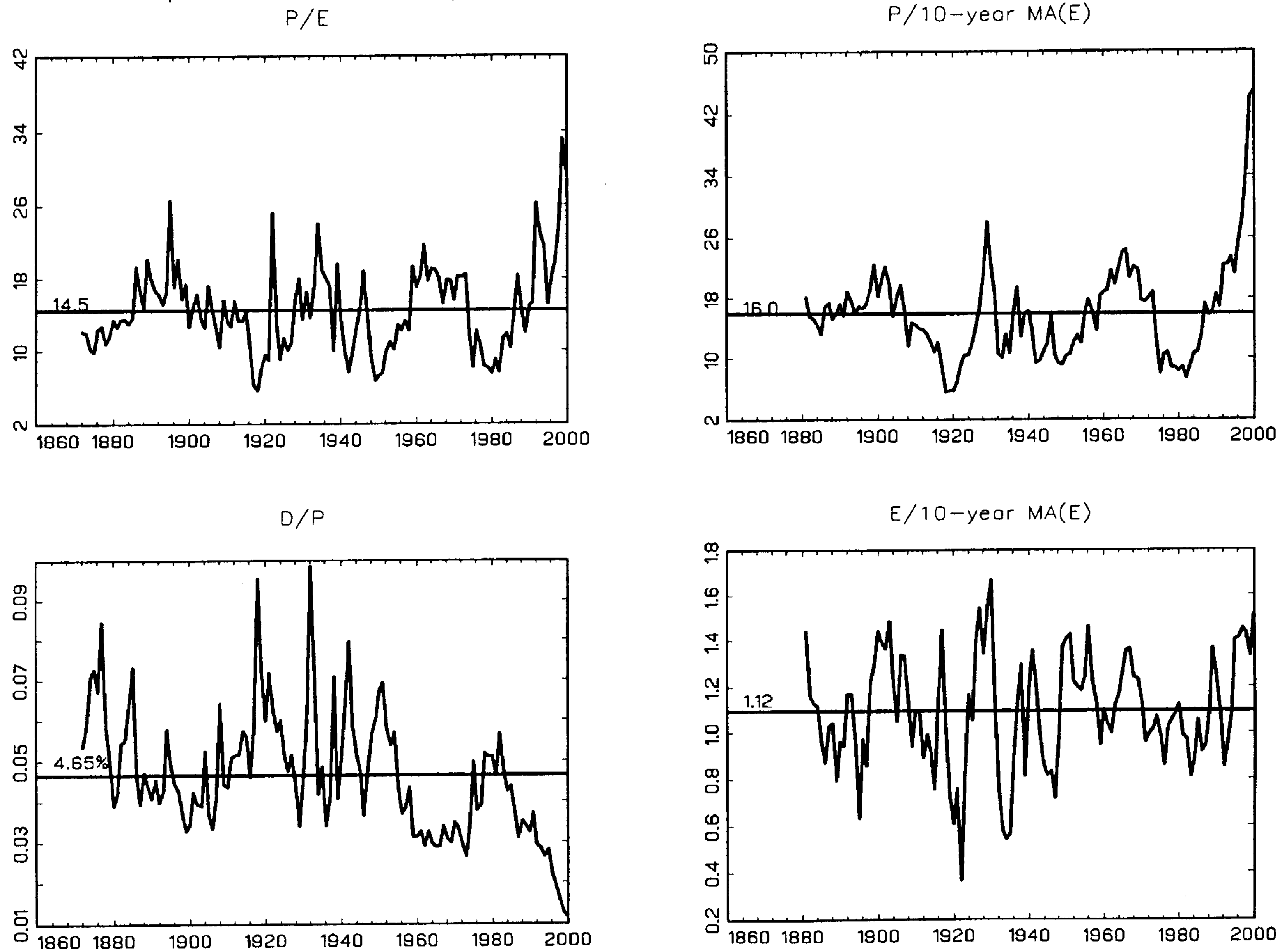
If equity returns are indeed lower on average in the future, it is likely that short-term and long-term real interest rates will be somewhat higher. That is, the total return to the corporate capital stock is determined primarily by the production side of the economy and by national saving and international capital flows; the division of total return between riskier and safer assets is determined primarily by investor attitudes towards risk. Reduced risk aversion then reduces the equity premium both by driving down the equity return and by driving up the riskless interest rate. The yield on long-term inflation-indexed Treasury securities (TIPS) is about 3.5%, while short-term real interest rates have recently averaged about 3%. Thus 3% to 3.5% would be a reasonable guess for safe real interest rates in the future, implying a long-run average equity premium of 1.5% to 2.5% in geometric terms or about 3% to 4% in arithmetic terms.

Finally, I note that it is tricky to use these numbers appropriately in policy evaluation. Average equity returns should never be used in base-case calculations without showing alternative calculations to reflect the possibilities that realized returns will be higher or lower than average. These calculations should include an alternative in which equities underperform Treasury bills. Even if the probability of underperformance is small over a long holding period, it cannot be zero or the stock market would be offering an arbitrage opportunity or "free lunch". Equally important, the bad states of the world in which underperformance occurs are heavily weighted by risk-averse investors. Thus policy evaluation should use a broad range of returns to reflect the uncertainty about long-run stock market performance.

[5] Campbell and Cochrane (1999) present a model in which investors judge their well-being by their consumption relative to a recent average of past aggregate consumption. In this model investors are more risk-tolerant when consumption grows rapidly and they have a "cushion of comfort" relative to their minimum expectations. The Campbell-Cochrane model fits past cyclical variations in the stock market, which will likely continue in the future, but it is hard to explain the extreme recent movements using this model.

[6] This compromise view also implies that negative serial correlation, or mean-reversion, is likely to remain a characteristic of stock returns in the 21st Century.

Figure 4. S&P Composite Stock Data, January Values 1872–1997



Bibliography

Campbell, John Y. and John H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy* 107, 205-251.

Campbell, John Y. and Robert J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", *Review of Financial Studies* 1, 195-228.

Campbell, John Y. and Robert J. Shiller, 2001, "Valuation Ratios and the Long-Run Stock Market Outlook: An Update", NBER Working Paper No. 8221.

Campbell, John Y. and Luis M. Viceira, 2001, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, forthcoming Oxford University Press, New York, NY.

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2000, *The Millennium Book: A Century of Investment Returns,* ABN Amro and London Business School, London.

Fama, Eugene and Kenneth R. French, 2000, "The Equity Premium", unpublished paper, University of Chicago and MIT.

Graham, Benjamin and David Dodd, 1934, *Security Analysis*, McGraw-Hill, New York, NY.

Heaton, John and Deborah Lucas, 1999, "Stock Prices and Fundamentals", *NBER Macroeconomics Annual* 213-242, MIT Press, Cambridge, MA.

Jagannathan, Ravi, Ellen R. McGrattan, and Anna Scherbina, 2001, "The Declining US Equity Premium", NBER Working Paper No. 8172.

Jorion, Philippe and William N. Goetzmann, 1999, "Global Stock Markets in the Twentieth Century", *Journal of Finance* 54, 953-980.

Liang, Nellie and Steven A. Sharpe, 1999, "Share Repurchases and Employee Stock Options and their Implications for S&P 500 Share Retirements and Expected Returns", unpublished paper, Board of Governors of the Federal Reserve System.

McGrattan, Ellen R., and Edward C. Prescott, 2001, "Is the Stock Market Overvalued?", NBER Working Paper No. 8077.

Shiller, Robert J., 2000, *Irrational Exuberance*, Princeton University Press, Princeton, NJ.

Siegel, Jeremy, 1998, *Stocks for the Long Run*, 2nd ed., McGraw-Hill, New York, NY.

Vuolteenaho, Tuomo, 2000, "Understanding the Aggregate Book-to-Market Ratio and its Implications to Current Equity-Premium Expectations", unpublished paper, Harvard University.

What Stock Market Returns to Expect for the Future: An Update

Peter A. Diamond, Professor of Economics
Massachusetts Institute of Technology
July 23, 2001

This note updates the calculations in my previous analysis of this issue (Social Security Bulletin, 2000, vol. 63, no. 2, pp. 38-52).* The calculations address two issues. First, what are the implications of assuming an annual 7% real return on equities throughout the next 75 years (along with the assumptions in the Trustees' Report), as has been the practice in OCACT projections of Social Security reform proposals that include equities. While the numbers are changed some from those based on the end of 1998, calculations done for the end of 2000 and the end of the first quarter of 2001 continue to show that a 7% return throughout the next 75 years from these starting points is implausible.

Second, what are the implications for stock market values in ten years if there is to be a lower rate of return for the next decade, followed by a return to the historical average return thereafter. As before, the returns over the next decade need to be very low, indeed an unchanged nominal value for stocks at the end of the decade is roughly consistent with close to a 7% return thereafter.

The calculations reported here are based on the Gordon formula, relating stock values to returns and the growth of returns. A first step in considering stock market returns is to project the future net cash flow to stockholders. This is normally done in three steps. First is to estimate the current net cash flow. Second is to adjust that for reasons to believe that the long-run relationship to GDP may be different from the current relationship. And third is to assume a constant relationship to GDP given the first two steps.

The cash flow to holders of publicly traded stocks as a whole contains many pieces. Easy to measure is the flow of dividends. Then there is the cash flow arising from share repurchase. This happens in two ways – direct repurchase of a corporation's own shares and acquisition of the shares of other corporations for cash or debt. Sometimes acquired shares are retired and sometimes they are not. This may be a complication in estimation given how data are presented – I have not reviewed measurement in data sources.

In order to maintain any given fraction of the value of shares outstanding, there are also pieces that are equivalent to negative cash flows. When employees exercise stock options and so acquire shares at less than market value, there is a dilution of the stock value of existing owners. This can be approached by thinking about the excess of market value over exercise price or by considering the value of options that are given to employees.

---

* See article beginning on p. 17.

I am grateful to Mauricio Soto for excellent research assistance, doing the calculations reported here. I am also grateful for financial support from the Retirement Research Center at Boston College.

Some existing firms go out of business while new firms are created. For considering the return on a given fraction of the entire outstanding traded stock, it is necessary to include the negative cash flow associated with additional traded companies. The direct cash flow of IPO's that are previously owned by individuals is such a negative cash flow. In addition, the value retained by the original owners also represents a dilution in the value of existing shareholders and also needs to be counted. Thus actual cash flow for new firms that were previously private needs to be increased by a multiplier – with 3 being a reasonable estimate. However, the analysis is different for new companies that are spin-offs from existing firms. The cash flow paid for them is a negative cash flow for shareholders as a whole. However, there is no need for a multiplier since the value of retained shares by corporations is retained by the aggregate of current shareholders. Thus there is a need to separate out these two types of IPO's. I have not seen an estimate separating these two parts.

In the methodology used in my previous paper, these various steps, along with any divergence of the current position from a steady state, were combined to produce a range of values referred to as adjusted dividend flow. In Table 1 are the implied ratios of stock market value to GDP at the end of the 75-year projection period based on stock market and GDP values at the end of 1998 and the assumptions in the 1999 Trustees' Report as well as values at the end of 2000 and end of the first quarter of 2001 and the assumptions in the 2001 Trustees' Report. The Table suggests that the 7 percent assumption throughout the next 75 years is not plausible in that it requires a rise in stock values to GDP that is implausible. The level of implausibility is not quite as high as two years ago, but it is still implausible. A sensitivity analysis is presented in Table 2 that varies the growth rate of GDP. Moderate increases in GDP growth above the levels assumed in the Trustees' Report still leave a 7% return throughout the next 75 years implausible.

Table 3 presents the size of the real drop in stock market values over the next ten years that are sufficient for the Gordon formula to yield a steady return of 7 percent thereafter (along with calculations for 6.5 and 6.0). Poor returns over the next ten years are needed for consistency with a higher ultimate long-run number, almost as poor as two years ago, for a given adjusted dividend level. Table 4 presents sensitivity analysis.

An important issue is whether it is more plausible to have a poor short-run return followed by a return to historic yields or to believe that the long-run ultimate return has dropped. Given the rest of the assumptions used by OCACT (particularly the assumption of a 3% real yield on long-term Treasuries), that is tantamount to a drop in the equity premium. I think many investors are not expecting as low a return as would be called for by the assumption that we are now in a steady state. Therefore, I continue to think a poor return over the next decade is a more plausible assumption. It seems sensible to lower the long-run return a little from the 7% historic norm in recognition of the unusually long period of very high returns that we have experienced (although one can wonder what would have happened in the late 20's and early 30's if Alan Greenspan had headed the Fed). Moreover, since it is impossible to predict timing of market corrections and it is sensible to work with a single rate of return for projection purposes, a lower rate of return is appropriate to correct for a period of lower returns even if the correction scenario returning all the way to 7% is right. Thus projection values around 6.0% or 6.5% seem to me appropriate for projection purposes. Of course, a wider band is important for high and low cost projections in order to show the extreme uncertainty associated with such a projection.

Table 1

Projections of the Ratio of Stock Market Value To GDP Assuming 7 Percent Real Return

End of 1998 Projections

	Adjusted Dividends			
	2.0%	2.5%	3.0%	3.5%
2073 Market to GDP	68.49	58.32	48.16	38.00
Ratio 2073 to Current	37.76	32.15	26.55	20.95

End of 2000 Projections

	Adjusted Dividends			
	2.0%	2.5%	3.0%	3.5%
2075 Market to GDP	44.93	37.73	30.54	23.34
Ratio 2075 to Current	26.47	22.23	17.99	13.75

End of First Quarter 2001 Projections

	Adjusted Dividends			
	2.0%	2.5%	3.0%	3.5%
2075 Market to GDP	39.54	33.29	27.03	20.77
Ratio 2075 to Current	26.81	22.57	18.33	14.08

Table 2

Projections of the Ratio of Stock Market Value To GDP Assuming 7 Percent Real Return

End of First Quarter 2001 Projections

	Adjusted Dividends 2.0%	2.5%	3.0%	3.5%
Under Current Projections				
2075 Market to GDP	39.54	33.29	27.03	20.77
Ratio 2075 to Current	26.81	22.57	18.33	14.08
GDP Growth 0.1% Higher				
2075 Market to GDP	36.34	30.43	24.51	18.60
Ratio 2075 to Current	24.64	20.63	16.62	12.61
GDP Growth 0.3% Higher				
2075 Market to GDP	30.65	25.37	20.08	14.79
Ratio 2075 to Current	20.78	17.20	13.61	10.02
GDP Growth 0.5% Higher				
2075 Market to GDP	25.81	21.07	16.34	11.60
Ratio 2075 to Current	17.50	14.29	11.08	7.86

*Assuming 7% stock yield, and using 2001 trustees projections.
** Using Estimated Market Value for April 1, 2001.

Table 3

Required Percentage Decline in Real Stock Prices Over the Following Ten Years To Justify a 7.0, 6.5, and 6.0 Percent Return Thereafter (end 1998)

Adjusted Dividend Yield	Long-run Return 7.0	6.5	6.0
2.0	55	51	45
2.5	44	38	31
3.0	33	26	18
3.5	21	13	4

Required Percentage Decline in Real Stock Prices Over the Following Ten Years To Justify a 7.0, 6.5, and 6.0 Percent Return Thereafter (end 2000)

Adjusted Dividend Yield	Long-run Return 7.0	6.5	6.0
2.0	53	48	42
2.5	41	35	28
3.0	29	22	13
3.5	17	9	-1

Source: Author's Calculations

Note: Derived from the Gordon Formula. Dividends are assumed to grow in line with GDP, which the OCACT assumed in 1999 is 2.0 percent over the next 10 years and 1.5 percent for the long run; and in 2001, 2.3 percent and then 1.6 percent.

Table 4

Required Percentage Decline in Real Stock Prices Over the Next Ten Years To Justify a 7.0, 6.5, and 6.0 Percent Return Thereafter (end 2000)

Under Current Projections

Adjusted Dividend Yield	Long-run Return 7.0	6.5	6.0
2.0	53	48	42
2.5	41	35	28
3.0	29	22	13
3.5	17	9	-1

GDP Growth 0.3% Higher Each Year

Adjusted Dividend Yield	Long-run Return 7.0	6.5	6.0
2.0	48	43	36
2.5	35	28	20
3.0	23	14	4
3.5	10	0	-12

Source: Author's Calculations

Note: Derived from the Gordon Formula. Dividends are assumed to grow in line with GDP, which the OACT assumes is 2.3 percent over the next 10 years. For long-run GDP growth, the OACT assumes 1.6 percent.

What Stock Market Returns to Expect for the Future?

Peter A. Diamond

Social Security Bulletin •Vol. 63 • No. 2 • 2000

High stock prices, together with projected slow economic growth, are not consistent with the 7.0 percent return that the Office of the Chief Actuary has generally used when evaluating proposals with stock investments. Routes out of the inconsistency include assuming higher GDP growth, a lower long-run stock return, or a lower short-run stock return with a 7.0 percent return on a lower base thereafter. In short, either the stock market is overvalued and requires a correction to justify a 7.0 percent return thereafter, or it is correctly valued and the long-run return is substantially lower than 7.0 percent (or some combination of the two). This article argues that the former view is more convincing, since accepting the "correctly valued" hypothesis implies an implausibly small equity premium.

This article originally appeared as an Issue in Brief of the Center for Retirement Research at Boston College (No. 2, September 1999). The research reported herein was performed pursuant to a grant from the Social Security Administration (SSA) funded as part of the Retirement Research Consortium. The opinions and conclusions expressed are solely those of the author and should not be construed as representing the opinions or policy of SSA, any agency of the federal government, or the Center for Retirement Research at Boston College.

I. Summary

In evaluating proposals for reforming Social Security that involve stock investments, the Office of the Chief Actuary (OCACT) has generally used a 7.0 percent real return for stocks. The 1994-96 Advisory Council specified that OCACT should use that return in making its 75-year projections of investment-based reform proposals. The assumed ultimate real return on Treasury bonds of 3.0 percent implies a long-run equity premium of 4.0 percent. There are two equity-premium concepts: the *realized* equity premium, which is measured by the actual rates of return; and the *required* equity premium, which investors expect to receive for being willing to hold available stocks and bonds. Over the past two centuries, the realized premium was 3.5 percent on average, but 5.2 percent for 1926 to 1998.

Some critics argue that the 7.0 percent projected stock returns are too high. They base their arguments on recent developments in the capital market, the current high value of the stock market, and the expectation of slower economic growth.

Increased use of mutual funds and the decline in their costs suggest a lower required premium, as does the rising fraction of the American public investing in stocks. The size of the decrease is limited, however, because the largest cost savings do not apply to the very wealthy and to large institutional investors, who hold a much larger share of the stock market's total value than do new investors. These trends suggest a lower equity premium for projections than the 5.2 percent of the past 75 years. Also, a declining required premium is likely to imply a temporary increase in the realized premium because a rising willingness to hold stocks tends to increase their price. Therefore, it would be a mistake during a transition period to extrapolate what may be a temporarily high realized return. In the standard (Solow) economic growth model, an assumption of slower long-run growth lowers the marginal product of capital if the savings rate is constant. But lower savings as growth slows should partially or fully offset that effect.

The present high stock prices, together with projected slow economic growth, are not consistent with a 7.0 percent return. With a plausible level of adjusted dividends (dividends plus net share repurchases), the ratio of stock value to gross domestic product (GDP) would rise more than 20-fold over 75 years. Similarly, the steady-state Gordon formula—that stock returns equal the adjusted dividend yield plus the growth rate of stock prices (equal to that of GDP)—suggests a return of roughly 4.0 percent to 4.5 percent. Moreover, when relative stock values have been high, returns over the following decade have tended to be low.

To eliminate the inconsistency posed by the assumed 7.0 percent return, one could assume higher GDP growth, a lower long-run stock return, or a lower short-run stock return with a 7.0 percent return on a lower base thereafter. For example, with an adjusted dividend yield of 2.5 percent to 3.0 percent, the market would have to decline about 35 percent to 45 percent in real terms over the next decade to reach steady state.

In short, either the stock market is overvalued and requires a correction to justify a 7.0 percent return thereafter, or it is correctly valued and the long-run return is substantially lower than 7.0 percent (or some combination). This article argues that the "overvalued" view is more convincing, since the "correctly valued" hypothesis implies an implausibly small equity premium. Although OCACT could adopt a lower rate for the entire 75-year period, a better approach would be to assume lower returns over the next decade and a 7.0 percent return thereafter.

II. Introduction

All three proposals of the 1994-96 Advisory Council on Social Security (1997) included investment in equities. For assessing the financial effects of those proposals, the Council members agreed to specify a 7.0 percent long-run real (inflation-adjusted) yield from stocks.[1] They devoted little attention to different short-run returns from stocks.[2] The Social Security Administration's Office of the Chief Actuary (OCACT) used this 7.0 percent return, along with a 2.3 percent long-run real yield on Treasury bonds, to project the impact of the Advisory Council's proposals.

Since then, OCACT has generally used 7.0 percent when assessing other proposals that include equities.[3] In the 1999 Social Security Trustees Report, OCACT used a higher long-term real rate on Treasury bonds of 3.0 percent.[4] In the first 10 years of its projection period, OCACT makes separate assumptions about bond rates for each year and assumes slightly lower real rates in the short run.[5] Since the assumed bond rate has risen, the assumed equity premium, defined as the difference between yields on equities and Treasuries, has declined to 4.0 percent in the long run.[6] Some critics have argued that the assumed return on stocks and the resulting equity premium are still too high.[7]

This article examines the critics' arguments and, rather than settling on a single recommendation, considers a range of assumptions that seem reasonable.[8] The article:

• Reviews the historical record on rates of return,

• Assesses the critics' reasons why future returns may be different from those in the historical record and examines the theory about how those rates are determined, and

• Considers two additional issues: the difference between gross and net returns, and investment risk.

Readers should note that in this discussion, a decline in the equity premium need not be associated with a decline in the return on stocks, since the return on bonds could increase. Similarly, a decline in the return on stocks need not be associated with a decline in the equity premium, since the return on bonds could also decline. Both rates of return and the equity premium are relevant to choices about Social Security reform.

III. Historical Record

Realized rates of return on various financial instruments have been much studied and are presented in Table 1.[9] Over the past 200 years, stocks have produced a real return of 7.0 percent per year. Even though annual returns fluctuate enormously, and rates vary significantly over periods of a decade or two, the return on stocks over very long periods has been quite stable (Siegel 1999).[10] Despite that long-run stability, great uncertainty surrounds both a projection for any particular period and the relevance of returns in any short period of time for projecting returns over the long run.

The equity premium is the difference between the rate of return on stocks and on an alternative asset—Treasury bonds, for the purpose of this article. There are two concepts of equity premiums. One is the *realized* equity premium, which is measured by the actual rates of return. The other is the *required* equity premium, which equals the premium that investors expect to get in exchange for holding available quantities of assets. The two concepts are closely related but different—significantly different in some circumstances.

The realized equity premium for stocks relative to bonds has been 3.5 percent for the two centuries of available data, but it has increased over time (Table 2).[11, 12] That increase has resulted

Table 1.
Compound annual real returns, by type of investment, 1802-1998 (in percent)

Period	Stocks	Bonds	Bills	Gold	Inflation
1802-1998	7.0	3.5	2.9	-0.1	1.3
1802-1870	7.0	4.8	5.1	0.2	0.1
1871-1925	6.6	3.7	3.2	-0.8	0.6
1926-1998	7.4	2.2	0.7	0.2	3.1
1946-1998	7.8	1.3	0.6	-0.7	4.2

Source: Siegel (1999).

Table 2.
Equity premiums: Differences in annual rates of return between stocks and fixed-income assets, 1802-1998

	Equity premium (percent)	
Period	With bonds	With bills
1802-1998	3.5	5.1
1802-1870	2.2	1.9
1871-1925	2.9	3.4
1926-1998	5.2	6.7
1946-1998	6.5	7.2

Source: Siegel (1999).

from a significant decline in bond returns over the past 200 years. The decline is not surprising considering investors' changing perceptions of default risk as the United States went from being a less-developed country (and one with a major civil war) to its current economic and political position, where default risk is seen to be virtually zero.[13]

These historical trends can provide a starting point for thinking about what assumptions to use for the future. Given the relative stability of stock returns over time, one might initially choose a 7.0 percent assumption for the return on stocks—the average over the entire 200-year period. In contrast, since bond returns have tended to decline over time, the 200-year number does not seem to be an equally good basis for selecting a long-term bond yield. Instead, one might choose an assumption that approximates the experience of the past 75 years—2.2 percent, which suggests an equity premium of around 5.0 percent. However, other evidence, discussed below, argues for a somewhat lower value.[14]

IV. Why Future Returns May Differ From Past Returns

Equilibrium and Long-Run Projected Rates of Return

The historical data provide one way to think about rates of return. However, thinking about how the future may be different from the past requires an underlying theory about how those returns are determined. This section lists some of the actions by investors, firms, and government that combine to determine equilibrium; it can be skipped without loss of continuity.

In asset markets, the demand by individual and institutional investors reflects a choice among purchasing stocks, purchasing Treasury bonds, and making other investments.[15] On the supply side, corporations determine the supplies of stocks and corporate bonds through decisions on dividends, new issues, share repurchases, and borrowing. Firms also choose investment levels. The supplies of Treasury bills and bonds depend on the government's budget and debt management policies as well as monetary policy. Whatever the supplies of stocks and bonds, their prices will be determined so that the available amounts are purchased and held by investors in the aggregate.

The story becomes more complicated, however, when one recognizes that investors base decisions about portfolios on their projections of both future prices of assets and future dividends.[16] In addition, market participants need to pay transactions costs to invest in assets, including administrative charges, brokerage commissions, and the bid-ask spread. The risk premium relevant for investors' decisions should be calculated net of transactions costs. Thus, the greater cost of investing in equities than in Treasuries must be factored into any discussion of the equity premium.[17] Differences in tax treatments of different types of income are also relevant (Gordon 1985; Kaplow 1994).

In addition to determining the supplies of corporate stocks and bonds, corporations also choose a debt/equity mix that affects the risk characteristics of both bonds and stocks. Financing a given level of investment more by debt and less by equity leaves a larger interest cost to be paid from the income of corporations before determining dividends. That makes both the debt and the equity more risky. Thus, changes in the debt/equity mix (possibly in response to prevailing stock market prices) should affect risk and, therefore, the equilibrium equity premium.[18]

Since individuals and institutions are generally risk averse when investing, greater expected variation in possible future yields tends to make an asset less valuable. Thus, a sensible expectation about long-run equilibrium is that the expected yield on equities will exceed that on Treasury bonds. The question at hand is how much more stocks should be expected to yield.[19] That is, assuming that volatility in the future will be roughly similar to volatility in the past, how much more of a return from stocks would investors need to expect in order to be willing to hold the available supply of stocks. Unless one thought that stock market volatility would collapse, it seems plausible that the premium should be significant. For example, equilibrium with a premium of 70 basis points (as suggested by Baker 1999a) seems improbable, especially since transactions costs are higher for stock than for bond investments. In considering this issue, one needs to recognize that a greater willingness to bear the risk associated with stocks is likely to be accompanied by greater volatility of stock prices if bond rates are unchanged. That is, fluctuations in expected growth in corporate profits will have bigger impacts on expected discounted returns (which approximate prices) when the equity premium, and so the discount rate, is lower.[20]

Although stocks should earn a significant premium, economists do not have a fully satisfactory explanation of why stocks have yielded so much more than bonds historically, a fact that has been called the equity-premium puzzle (Mehra and Prescott 1985; Cochrane 1997). Ongoing research is trying to develop more satisfactory explanations, but the theory still has inadequacies.[21] Nevertheless, to explain why the future may be different from the past, one needs to rely on some theoretical explanation of the past in order to have a basis for projecting a different future.

Commentators have put forth three reasons as to why future returns may be different from those in the historical record. First, past and future long-run trends in the capital market may imply a decline in the equity premium. Second, the current valuation of stocks, which is historically high relative to various benchmarks, may signal a lower future rate of return on equities. Third, the projection of slower economic growth may suggest a lower long-run marginal product of capital, which is the source of returns to financial assets. The first two issues are discussed in the context of financial markets; the third, in the context of physical assets. One should distinguish between arguments that suggest a lower equity premium and those that suggest lower returns to financial assets generally.

Equity Premium and Developments in the Capital Market

The capital market has experienced two related trends—the decrease in the cost of acquiring a diversified portfolio of stocks and the spread of stock ownership more widely in the economy. The relevant equity premium for investors is the equity premium net of the costs of investing. Thus, if the cost of investing in some asset decreases, that asset should have a higher price and a lower expected return gross of investment costs. The availability of mutual funds and the decrease in the cost of purchasing them should lower the equity premium in the future relative to long-term historical values. Arguments have also been raised about investors' time horizons and their understanding of financial markets, but the implications of those arguments are less clear.

Mutual Funds. In the absence of mutual funds, small investors would need to make many small purchases in different companies in order to acquire a widely diversified portfolio. Mutual funds provide an opportunity to acquire a diversified portfolio at a lower cost by taking advantage of the economies of scale in investing. At the same time, these funds add another layer of intermediation, with its costs, including the costs of marketing the funds.

Nevertheless, as the large growth of mutual funds indicates, many investors find them a valuable way to invest. That suggests that the equity premium should be lower in the future than in the past, since greater diversification means less risk for investors. However, the significance of the growth of mutual funds depends on the importance in total equity demand of "small" investors who purchase them, since this argument is much less important for large investors, particularly large institutional investors. According to recent data, mutual funds own less than 20 percent of U.S. equity outstanding (Investment Company Institute 1999).

A second development is that the average cost of investing in mutual funds has decreased. Rea and Reid (1998) report a drop of 76 basis points (from 225 to 149) in the average annual charge of equity mutual funds from 1980 to 1997. They attribute the bulk of the decline to a decrease in the importance of front-loaded funds (funds that charge an initial fee when making a deposit in addition to annual charges). The development and growth of index funds should also reduce costs, since index funds charge investors considerably less on average than do managed funds while doing roughly as well in gross rates of return. In a separate analysis, Rea and Reid (1999) also report a decline of 38 basis points (from 154 to 116) in the cost of bond mutual funds over the same period, a smaller drop than with equity mutual funds. Thus, since the cost of stock funds has fallen more than the cost of bond funds, it is plausible to expect a decrease in the equity premium relative to historical values. The importance of that decline is limited, however, by the fact that the largest cost savings do not apply to large institutional investors, who have always faced considerably lower charges.

A period with a declining required equity premium is likely to have a temporary increase in the realized equity premium. Assuming no anticipation of an ongoing trend, the divergence occurs because a greater willingness to hold stocks, relative to bonds, tends to increase the price of stocks. Such a price rise may yield a realized return that is higher than the required return.[22] The high realized equity premium since World War II may be partially caused by a decline in the required equity premium over that period. During such a transition period, therefore, it would be a mistake to extrapolate what may be a temporarily high realized return.

Spread of Stock Ownership. Another trend that would tend to decrease the equity premium is the rising fraction of the American public investing in stocks either directly or indirectly through mutual funds and retirement accounts (such as 401(k) plans). Developments in tax law, pension provision, and the capital markets have expanded the base of the population who are sharing in the risks associated with the return to corporate stock. The share of households investing in stocks in any form increased from 32 percent in 1989 to 41 percent in 1995 (Kennickell, Starr-McCluer, and Sundén 1997). Numerous studies have concluded that widening the pool of investors sharing in stock market risk should lower the equilibrium risk premium (Mankiw and Zeldes 1991; Brav and Geczy 1996; Vissing-Jorgensen 1997; Diamond and Geanakoplos 1999; Heaton and Lucas 2000). The importance of that trend must be weighted by the low size of investment by such new investors.[23]

Investors' Time Horizons. A further issue relevant to the future of the equity premium is whether the time horizons of investors, on average, have changed or will change.[24] Although the question of how time horizons should affect demands for assets raises subtle theoretical issues (Samuelson 1989), longer horizons and sufficient risk aversion should lead to greater willingness to hold stocks given the tendency for stock prices to revert toward their long-term trend (Campbell and Viceira 1999).[25]

The evidence on trends in investors' time horizons is mixed. For example, the growth of explicit individual retirement savings vehicles, such as individual retirement accounts (IRAs) and 401(k)s, suggests that the average time horizons of individual investors may have lengthened. However, some of that growth is at the expense of defined benefit plans, which may have longer horizons. Another factor that might suggest a longer investment horizon is the increase in equities held by institutional investors, particularly through defined benefit pension plans. However, the relevant time horizon for such holdings may not be the open-ended life of the plan but rather the horizon of the plans' asset managers, who may have career concerns that shorten the relevant horizon.

Other developments may tend to lower the average horizon. Although the retirement savings of baby boomers may currently add to the horizon, their aging and the aging of the population generally will tend to shorten horizons. Finally, individual stock ownership has become less concentrated (Poterba and Samwick 1995), which suggests a shorter time horizon because less wealthy investors might be less concerned about passing assets on to younger generations. Overall, without detailed calculations that would go beyond the scope of this article, it is not clear how changing time horizons should affect projections.

Investors' Understanding. Another factor that may affect the equity premium is investors' understanding of the properties of stock and bond investments. The demand for stocks might be affected by the popular presentation of material, such as Siegel (1998), explaining to the general public the difference between short- and long-run risks. In particular, Siegel highlights the risks, in real terms, of holding nominal bonds. While the creation of inflation-indexed Treasury bonds might affect behavior, the lack of wide interest in those bonds (in both the United States and the United Kingdom) and the failure to fully adjust future amounts for inflation generally (Shafir, Diamond, and Tversky 1997) suggest that nominal bonds will continue to be a major part of portfolios. Perceptions that those bonds are riskier than previously believed would then tend to decrease the required equity premium.

Popular perceptions may, however, be excessively influenced by recent events—both the high returns on equity and the low rates of inflation. Some evidence suggests that a segment of the public generally expects recent rates of increase in the prices of assets to continue, even when those rates seem highly implausible for a longer term (Case and Shiller 1988). The possibility of such extrapolative expectations is also connected with the historical link between stock prices and inflation. Historically, real stock prices have been adversely affected by inflation in the short run. Thus, the decline in inflation expectations over the past two decades would be associated with a rise in real stock prices if the historical pattern held. If investors and analysts fail to consider such a connection, they might expect robust growth in stock prices to continue without recognizing that further declines in inflation are unlikely. Sharpe (1999) reports evidence that stock analysts' forecasts of real growth in corporate earnings include extrapolations that may be implausibly high. If so, expectations of continuing rapid growth in stock prices suggest that the required equity premium may not have declined.

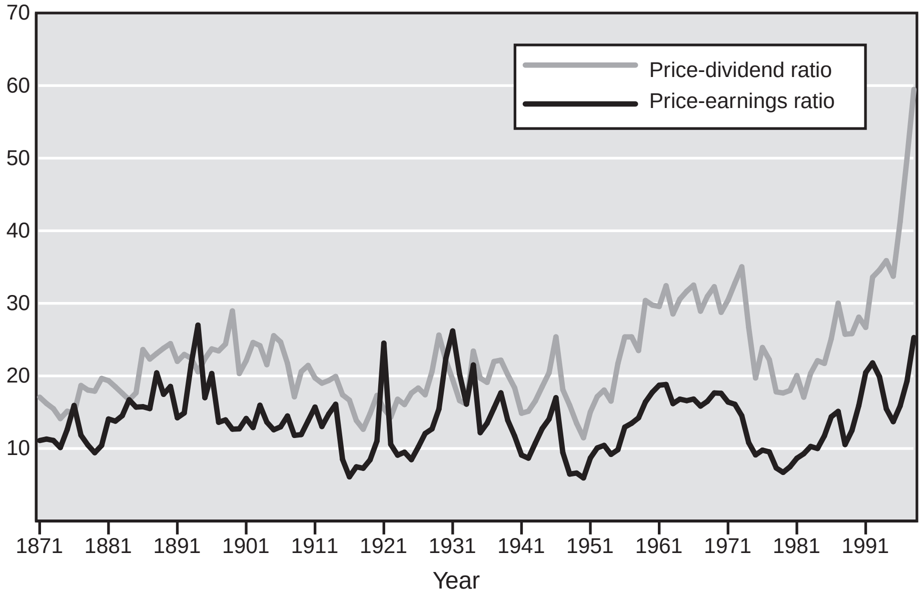
On balance, the continued growth and development of mutual funds and the broader participation in the stock market should contribute to a drop in future equity premiums relative to the historical premium, but the drop is limited.[26] Other factors, such as investors' time horizons and understanding, have less clear-cut implications for the equity premium.

Equity Premium and Current Market Values

At present, stock prices are very high relative to a number of different indicators, such as earnings, dividends, book values, and gross domestic product (GDP) (Charts 1 and 2). Some critics, such as Baker (1998), argue that this high market value, combined with projected slow economic growth, is not consistent with a 7.0 percent return. Possible implications of the high prices have also been the subject of considerable discussion in the finance community (see, for example, Campbell and Shiller 1998; Cochrane 1997; Philips 1999; and Siegel 1999).

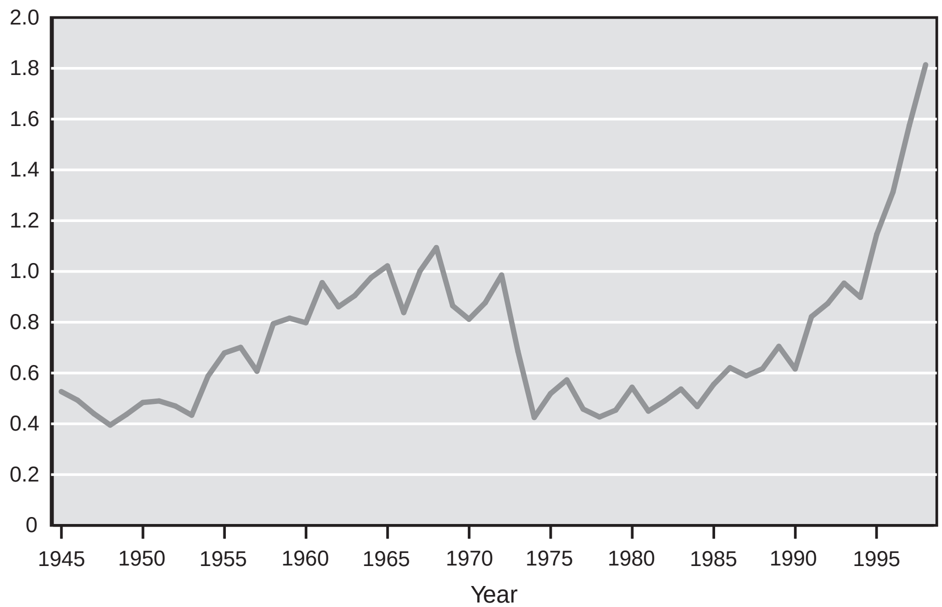
The inconsistency of current share prices and 7.0 percent real returns, given OCACT's assumptions for GDP growth, can be illustrated in two ways. The first way is to project the ratio of the stock market's value to GDP, starting with today's values and given assumptions about the future. The second way is to ask what must be true if today's values represent a steady state in the ratio of stock values to GDP.

Chart 1.
Price-dividend ratio and price-earnings ratio, 1871-1998



Source: Robert Shiller, Yale University. Available at www.econ.yale.edu/~shiller/data/chapt26.html.
Note: These ratios are based on Standard and Poor's Composite Stock Price Index.

Chart 2.
Ratio of market value of stocks to gross domestic product,1945-1998



Source: Bureau of Economic Analysis data from the national income and product accounts and federal flow of funds.

The first calculation requires assumptions for stock returns, adjusted dividends (dividends plus net share repurchases),[27] and GDP growth. For stock returns, the 7.0 percent assumption is used. For GDP growth rates, OCACT's projections are used. For adjusted dividends, one approach is to assume that the ratio of the aggregate adjusted dividend to GDP would remain the same as the current level. However, as discussed in the accompanying box, the current ratio seems too low to use for projection purposes. Even adopting a higher, more plausible level of adjusted dividends, such as 2.5 percent or 3.0 percent, leads to an implausible rise in the ratio of stock value to GDP—in this case, a more than 20-fold increase over the next 75 years. The calculation derives each year's capital gains by subtracting projected adjusted dividends from the total cash flow to shareholders needed to return 7.0 percent on that year's share values. (See Appendix A for an alternative method of calculating this ratio using a continuous-time differential equation.)

A second way to consider the link between stock market value, stock returns, and GDP is to look at a steady-state relationship. The Gordon formula says that stock returns equal the ratio of adjusted dividends to prices (or the adjusted dividend yield) plus the growth rate of stock prices.[28] In a steady state, the growth rate of prices can be assumed to equal that of GDP. Assuming an adjusted dividend yield of roughly 2.5 percent to 3.0 percent and projected GDP growth of 1.5 percent, the Gordon equation implies a stock return of roughly 4.0 percent to 4.5 percent, not 7.0 percent. Those lower values would imply an equity premium of 1.0 percent to 1.5 percent, given OCACT's assumption of a 3.0 percent yield on Treasury bonds. Making the equation work with a 7.0 percent stock return, assuming no change in projected GDP growth, would require an adjusted dividend yield of roughly 5.5 percent—about double today's level.[29]

For such a large jump in the dividend yield to occur, one of two things would have to happen—adjusted dividends would have to grow much more rapidly than the economy, or stock prices would have to grow much less rapidly than the economy (or even decline). But a consistent projection would take a very large jump in adjusted dividends, assuming that stock prices grew along with GDP starting at today's value. Estimates of recent values of the adjusted dividend yield range from 2.10 percent to 2.55 percent (Dudley and others 1999; Wadhwani 1998).[30]

Even with reasons for additional growth in the dividend yield, which are discussed in the box on projecting future dividends, an implausible growth of adjusted dividends is needed if the short- and long-term returns on stocks are to be 7.0 percent. Moreover, historically, very low values of the dividend yield and earnings-price ratio have been followed primarily by adjustments in stock prices, not in dividends and earnings (Campbell and Shiller 1998).

If the ratio of aggregate adjusted dividends to GDP is unlikely to change substantially, there are three ways out of the internal inconsistency between the market's current value and OCACT's assumptions for economic growth and stock returns. One can:

- Assume higher GDP growth, which would decrease the implausibility of the calculations described above for either the ratio of market value to GDP or the steady state under the Gordon equation. (The possibility of more rapid GDP growth is not explored further in this article.[31])

Projecting Future Adjusted Dividends

This article uses the concept of adjusted dividends to estimate the dividend yield. The adjustment begins by adding the value of net share repurchases to actual dividends, since that also represents a cash flow to stockholders in aggregate. A further adjustment is then made to reflect the extent to which the current situation might not be typical of the relationship between dividends and gross domestic product (GDP) in the future. Three pieces of evidence suggest that the current ratio of dividends to GDP is abnormally low and therefore not appropriate to use for projection purposes.

First, dividends are currently very low relative to corporate earnings—roughly 40 percent of earnings compared with a historical average of 60 percent. Because dividends tend to be much more stable over time than earnings, the dividend-earnings ratio declines in a period of high growth of corporate earnings. If future earnings grow at the same rate as GDP, dividends will probably grow faster than GDP to move toward the historical ratio.[1] On the other hand, earnings, which are high relative to GDP, might grow more slowly than GDP. But then, corporate earnings, which have a sizable international component, might grow faster than GDP.

Second, corporations are repurchasing their outstanding shares at a high rate. Liang and Sharpe (1999) report on share repurchases by the 144 largest (nonbank) firms in the Standard and Poor's 500. From 1994 to 1998, approximately 2 percent of share value was repurchased, although Liang and Sharpe anticipate a lower value in the future. At the same time, those firms were issuing shares because employees were exercising stock options at prices below the share values, thus offsetting much of the increase in the number of shares outstanding. Such transfers of net wealth to employees presumably reflect past services. In addition, initial public offerings (IPOs) represent a negative cash flow from stockholders as a whole. Not only the amount paid for stocks but also the value of the shares held by insiders represents a dilution relative to a base for long-run returns on all stocks. As a result, some value needs to be added to the current dividend ratio to adjust for net share repurchases, but the exact amount is unclear. However, in part, the high rate of share repurchase may be just another reflection of the low level of dividends, making it inappropriate to both project much higher dividends in the near term and assume that all of the higher share repurchases will continue. Exactly how to project current numbers into the next decade is not clear.

Finally, projected slow GDP growth, which will plausibly lower investment levels, could be a reason for lower retained earnings in the future. A stable level of earnings relative to GDP and lower retained earnings would increase the ratio of adjusted dividends to GDP.[2]

In summary, the evidence suggests using an "adjusted" dividend yield that is larger than the current level. Therefore, the illustrative calculations in this article use adjusted dividend yields of 2.0 percent, 2.5 percent, 3.0 percent, and 3.5 percent. (The current level of dividends without adjustment for share repurchases is between 1.0 percent and 2.0 percent.)

---

[1] For example, Baker and Weisbrot (1999) appear to make no adjustment for share repurchases or for current dividends being low. However, they use a dividend payout of 2.0 percent, while Dudley and others (1999) report a current dividend yield on the Wilshire 5000 of 1.3 percent.

[2] Firms might change their overall financing package by changing the fraction of net earnings they retain. The implications of such a change would depend on why they were making it. A long-run decrease in retained earnings might merely be increases in dividends and borrowing, with investment held constant. That case, to a first approximation, is another application of the Modigliani-Miller theorem, and the total stock value would be expected to fall by the decrease in retained earnings. Alternatively, a change in retained earnings might signal a change in investment. Again, there is ambiguity. Firms might be retaining a smaller fraction of earnings because investment opportunities were less attractive or because investment had become more productive. These issues tie together two parts of the analysis in this article. If slower growth is associated with lower investment that leaves the return on capital relatively unchanged, then what financial behavior of corporations is required for consistency? Baker (1999b) makes such a calculation; it is not examined here.

- Adopt a long-run stock return that is considerably less than 7.0 percent.
- Lower the rate of return during an intermediate period so that a 7.0 percent return could be applied to a lower market value base thereafter.

A combination of the latter two alternatives is also possible.

In considering the prospect of a near-term market decline, the Gordon equation can be used to compute the magnitude of the drop required over, for example, the next 10 years in order for stock returns to average 7.0 percent over the remaining 65 years of OCACT's projection period (see Appendix B). A long-run return of 7.0 percent would require a drop in real prices of between 21 percent and 55 percent, depending on the assumed value of adjusted dividends (Table 3).[32] That calculation is relatively sensitive to the assumed rate of return—for example, with a long-run return of 6.5 percent, the required drop in the market falls to a range of 13 percent to 51 percent.[33]

The two different ways of restoring consistency—a lower stock return in all years or a near-term decline followed by a return to the historical yield—have different implications for Social Security finances. To illustrate the difference, consider the contrast between a scenario with a steady yield of 4.25 percent derived by using current values for the Gordon equation as described above (the steady-state scenario) and a scenario in which stock prices drop by half immediately and the yield on stocks is 7.0 percent thereafter (the market-correction scenario).[34] First, dollars newly invested in the future (that is, after any drop in share prices) earn only 4.25 percent per year under the steady-state scenario, compared with 7.0 percent per year under the market-correction scenario. Second, even for dollars currently in the market, the long-run yield differs under the two scenarios when the returns on stocks are being reinvested.

Under the steady-state scenario, the yield on dollars currently in the market is 4.25 percent per year over any projected time period; under the market-correction scenario, the annual rate of return depends on the time horizon used for the calculation.[35] After one year, the latter scenario has a rate of return of –46 percent. By the end of 10 years, the annual rate of return with the latter scenario is –0.2 percent; by the end of 35 years, 4.9 percent; and by the end of 75 years, 6.0 percent. Proposals for Social Security generally envision a gradual buildup of stock investments, which suggests that those investments would fare better under the market-correction scenario. The importance of the difference between scenarios depends also on the choice of additional changes to Social Security, which affect how long the money can stay invested until it is needed to pay benefits.

Given the different impacts of these scenarios, which one is more likely to occur? The key issue is whether the current stock market is overvalued in the sense that rates of return are likely to be lower in the intermediate term than in the long run. Economists have divergent views on this issue.

Table 3.
Required percentage decline in real stock prices over the next 10 years to justify a return of 7.0, 6.5, and 6.0 percent thereafter

Adjusted dividend yield	Percentage decline to justify a long-run return of— 7.0	6.5	6.0
2.0	55	51	45
2.5	44	38	31
3.0	33	26	18
3.5	21	13	4

Source: Author's calculations.
Note: Derived from the Gordon formula. Dividends are assumed to grow in line with gross domestic product (GDP), which the Office of the Chief Actuary (OCACT) assumes is 2.0 percent over the next 10 years. For long-run GDP growth, OCACT assumes 1.5 percent.

One possible conclusion is that current stock prices signal a significant drop in the long-run required equity premium. For example, Glassman and Hassett (1999) have argued that the equity premium will be dramatically lower in the future than it has been in the past, so that the current market is not overvalued in the sense of signaling lower returns in the near term than in the long run.[36] Indeed, they even raise the possibility that the market is "undervalued" in the sense that the rate of return in the intermediate period will be higher than in the long run, reflecting a possible continuing decline in the required equity premium. If their view is right, then a 7.0 percent long-run return, together with a 4.0 percent equity premium, would be too high.

Others argue that the current stock market values include a significant price component that will disappear at some point, although no one can predict when or whether it will happen abruptly or slowly. Indeed, Campbell and Shiller (1998) and Cochrane (1997) have shown that when stock prices (normalized by earnings, dividends, or book values) have been far above historical ratios, the rate of return over the following decade has tended to be low, and the low return is associated primarily with the price of stocks, not the growth of dividends or earnings.[37] Thus, to project a steady rate of return in the future, one needs to argue that this historical pattern will not repeat itself. The values in Table 3 are in the range suggested by the historical relationship between future stock prices and current price-earnings and price-dividend ratios (see, for example, Campbell and Shiller 1998).

Therefore, either the stock market is overvalued and requires a correction to justify a 7.0 percent return thereafter, or it is correctly valued and the long-run return is substantially lower than 7.0 percent. (Some combination of the two is also possible.) Under either scenario, stock returns would be lower than 7.0 percent for at least a portion of the next 75 years. Some evidence suggests, however, that investors have not adequately considered that possibility.[38] The former view is more convincing, since accepting the "correctly valued" hypothesis implies an implausibly small long-run equity premium. Moreover, when stock values (compared with earnings or dividends) have been far above historical ratios, returns over the following decade have tended to be low. Since this discussion has no direct bearing on bond returns, assuming a lower return for stocks over the near- or long-term also means assuming a lower equity premium.

In short, given current stock values, a constant 7.0 percent return is not consistent with OCACT's projected GDP growth.[39] However, OCACT could assume lower returns for a decade, followed by a return equal to or about 7.0 percent.[40] In that case, OCACT could treat equity returns as it does Treasury rates, using different projection methods for the first 10 years and for the following 65. This conclusion is not meant to suggest that anyone is capable of predicting the timing of annual stock returns, but rather that this is an approach to financially consistent assumptions. Alternatively, OCACT could adopt a lower rate of return for the entire 75-year period.

Marginal Product of Capital and Slow Growth

In its long-term projections, OCACT assumes a slower rate of economic growth than the U.S. economy has experienced over an extended period. That projection reflects both the slowdown in labor force growth expected over the next few decades and the slowdown in productivity growth since 1973.[41] Some critics have suggested that slower growth implies lower projected rates of return on both stocks and bonds, since the returns to financial assets must reflect the returns on capital investment over the long run. That issue can be addressed by considering either the return to stocks directly, as discussed above, or the marginal product of capital in the context of a model of economic growth.[42]

For the long run, the returns to financial assets must reflect the returns on the physical assets that support the financial assets. Thus, the question is whether projecting slower economic growth is a reason to expect a lower marginal product of capital. As noted above, this argument speaks to rates of return generally, not necessarily to the equity premium.

The standard (Solow) model of economic growth implies that slower long-run economic growth with a constant savings rate will yield a lower marginal product of capital, and the relationship may be roughly point-for-point (see Appendix C). However, the evidence suggests that savings rates are not unaffected by growth rates. Indeed, growth may be more important for savings rates than savings are for growth rates. Bosworth and Burtless (1998) have observed that savings rates and long-term rates of income growth have a persistent positive association, both across countries and over time. That observation suggests that if future economic growth is slower than in the past, savings will also be lower. In the Solow model, low savings raise the marginal product of capital, with each percentage-point decrease in the savings rate increasing the marginal product by roughly one-half of a percentage point in the long run. Since growth has fluctuated in the past, the stability in real rates of return to stocks, as shown in Table 1, suggests an offsetting savings effect, preserving the stability in the rate of return.[43]

Focusing directly on demographic structure and the rate of return rather than on labor force growth and savings rates, Poterba (1998) does not find a robust relationship between demographic structure and asset returns. He does recognize the limited power of statistical tests based on the few "effective degrees of freedom" in the historical record. Poterba suggests that the connection between demography and returns is not simple and direct, although such a connection has been raised as a possible reason for high current stock values, as baby boomers save for retirement, and for projecting low future stock values, as they finance retirement consumption. Goyal (1999) estimates equity premium regressions and finds that changes in population age structure add significant explanatory power. Nevertheless, using a vector autoregression approach, his analysis predicts no significant increase in *average* outflows over the next 52 years. That occurs despite the retirement of baby boomers. Thus, both papers reach the same conclusion—that demography is not likely to effect large changes in the long-run rate of return.

Another factor to consider in assessing the connection between growth and rates of return is the increasing openness of the world economy. Currently, U.S. corporations earn income from production and trade abroad, and individual investors, while primarily investing at home, also invest abroad. It is not clear that putting the growth issue in a global context makes much difference. On the one hand, since other advanced economies are also aging, increased economic connections with other advanced countries do not alter the basic analysis. On the other hand, although investment in the less-developed countries may preserve higher rates, it is not clear either how much investment opportunities will increase or how to adjust for political risk. Increasing openness further weakens the argument for a significant drop in the marginal product of capital, but the opportunities abroad may or may not be realized as a better rate of return.

On balance, slower projected growth may reduce the return on capital, but the effect is probably considerably less than one-for-one. Moreover, this argument relates to the overall return to capital in an economy, not just stock returns. Any impact would therefore tend to affect returns on both stocks and bonds similarly, with no directly implied change in the equity premium.[44]

V. Other Issues

This paper has considered the gross rate of return to equities and the equity premium generally. Two additional issues arise in considering the prospect of equity investment for Social Security: how gross returns depend on investment strategy and how they differ from net returns; and the degree of risk associated with adding stock investments to a current all-bond portfolio.

Gross and Net Returns

A gross rate of return differs from a net return because it includes transactions costs such as brokerage charges, bid-ask spreads, and fees for asset management.[45]

If the Social Security trust fund invests directly in equities, the investment is likely to be in an index fund representing almost all of the equities outstanding in the United States. Thus, the analysis above holds for that type of investment. Although some critics have expressed concern that political influence might cause deviations from a broad-based indexing strategy, the evidence suggests that such considerations would have little impact on the expected rate of return (Munnell and Sundén 1999).

If the investment in stocks is made through individual accounts, then individuals may be given some choice either about the makeup of stock investment or about varying the mix of stocks and bonds over time. In order to consider the rate of return on stocks held in such individual accounts, one must consider the kind of portfolio choices individuals might make, both in the composition of the stock portfolio and in the timing of purchases and sales. Given the opportunity, many individuals would engage in numerous transactions, both among stocks and between stocks and other assets (attempts to time the market).

The evidence suggests that such transactions reduce gross returns relative to risks, even before factoring in transactions costs (Odean 1998). Therefore, both the presence of individual accounts with choice and the details of their regulation are likely to affect gross returns. On average, individual accounts with choice are likely to have lower gross returns from stocks than would direct trust fund investment.

Similarly, the cost of administration as a percentage of managed assets varies depending on whether there are individual accounts and how they are organized and regulated (National Academy of Social Insurance 1998; Diamond 2000). Estimates of that cost vary from 0.5 basis points for direct trust fund investment to 100 to150 basis points for individually organized individual accounts, with government-organized individual accounts somewhere in between.

Investment Risk of Stocks

The Office of the Chief Actuary's projections are projections of plausible long-run scenarios (ignoring fluctuations). As such, they are useful for identifying a sizable probability of future financial needs for Social Security. However, they do not address different probabilities for the trust fund's financial condition under different policies.[46] Nor are they sufficient for normative evaluation of policies that have different distributional or risk characteristics.

Although investment in stocks entails riskiness in the rate of return, investment in Treasury bonds also entails risk. Therefore, a comparison of those risks should consider the distribution of outcomes—concern about risk should not be separated from the compensation for bearing risk. That is, one needs to consider the probabilities of both doing better and doing worse as a result of holding some stocks. Merely observing that stocks are risky is an inadequate basis for policy evaluations. Indeed, studies of the historical pattern of returns show that portfolio risk decreases when some stocks are added to a portfolio consisting only of nominal bonds (Siegel 1998). Furthermore, many risks affect the financial future of Social Security, and investing a small portion of the trust fund in stocks is a small risk for the system as a whole relative to economic and demographic risks (Thompson 1998).

As long as the differences in risk and expected return are being determined in a market and reflect the risk aversion of market participants, the suitability of the trust fund's portfolio can be considered in terms of whether Social Security has more or less risk aversion than current investors. Of course, the "risk aversion" of Social Security is a derived concept, based on the risks to be borne by future beneficiaries and taxpayers, who will incur some risk whatever portfolio Social Security holds. Thus, the question is whether the balance of risks and returns looks better with one portfolio than with another. The answer is somewhat complex, since it depends on how policy changes in taxes and benefits respond to economic and demographic outcomes. Nevertheless, since individuals are normally advised to hold at least some stocks in their own portfolios, it seems appropriate for Social Security to also hold some stocks when investing on their behalf, at least in the long run, regardless of the rates of return used for projection purposes (Diamond and Geanakoplos 1999).[47]

VI. Conclusion

Of the three main bases for criticizing OCACT's assumptions, by far the most important one is the argument that a constant 7.0 percent stock return is not consistent with the value of today's stock market and projected slow economic growth. The other two arguments—pertaining to developments in financial markets and the marginal product of capital—have merit, but neither suggests a dramatic change in the equity premium.

Given the high value of today's stock market and an expectation of slower economic growth in the future, OCACT could adjust its stock return projections in one of two ways. It could assume a decline in the stock market sometime over the next decade, followed by a 7.0 percent return for the remainder of the projection period. That approach would treat equity returns like Treasury rates, using different short- and long-run projection methods for the first 10 years and the following 65 years. Alternatively, OCACT could adopt a lower rate of return for the entire 75-year period. That approach may be more acceptable politically, but it obscures the expected pattern of returns and may produce misleading assessments of alternative financing proposals, since the appropriate uniform rate to use for projection purposes depends on the investment policy being evaluated.

Notes

Peter Diamond is Institute Professor at the Massachusetts Institute of Technology, where he has taught since 1966. He is a member of the Board of the National Academy of Social Insurance, where he has been President, Chair of the Board, and Chair of the Panel on Privatization of Social Security. He has written on public finance, macroeconomics, and the economics of uncertainty.

Acknowledgments: The author is grateful to John Campbell, Alicia Munnell, and Jim Poterba for extended discussions and to Andy Abel, Dean Baker, Olivier Blanchard, John Cochrane, Andy Eschtruth, Steve Goss, Joyce Manchester, Peter Orszag, Bernie Saffran, Jeremy Siegel, Tim Smeeding, Peter Temin, and Joe White for helpful comments. The views and remaining errors are those of the author.

[1] This 7.0 percent real rate of return is gross of administrative charges.

[2] To generate short-run returns on stocks, the Social Security Administration's Office of the Chief Actuary (OCACT) multiplied the ratio of one plus the ultimate yield on stocks to one plus the ultimate yield on bonds by the annual bond assumptions in the short run.

[3] An exception was the use of 6.75 percent for the President's proposal evaluated in a memorandum on January 26, 1999.

[4] This report is formally called the *1999 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.*

[5] For OCACT's short-run bond projections, see Table II.D.1 in the 1999 Social Security Trustees Report.

[6] This article was written in the summer of 1999 and uses numbers appropriate at the time. The 2000 Trustees Report uses the same assumptions of 6.3 percent for the nominal interest rate and 3.3 percent for the annual percentage change in the consumer price index. The real wage is assumed to grow at 1.0 percent, as opposed to 0.9 percent in the 1999 report.

[7] See, for example, Baker (1999a) and Baker and Weisbrot (1999). This article only considers return assumptions given economic growth assumptions and does not consider growth assumptions.

[8] This article does not analyze the policy issues related to stock market investment either by the trust fund or through individual accounts. Such an analysis needs to recognize that higher expected returns in the U.S. capital market come with higher risk. For the issues relevant for such a policy analysis, see National Academy of Social Insurance (1998).

[9] Ideally, one would want the yield on the special Treasury bonds held by Social Security. However, this article simply refers to published long-run bond rates.

[10] Because annual rates of return on stocks fluctuate so much, a wide band of uncertainty surrounds the best statistical estimate of the average rate of return. For example, Cochrane (1997) notes that over the 50 years from 1947 to 1996, the excess return of stocks over Treasury bills was 8 percent, but, assuming that annual returns are statistically independent, the standard statistical confidence interval extends from 3 percent to 13 percent. Using a data set covering a longer period lowers the size of the confidence interval, provided one is willing to assume that the stochastic process describing rates of return is stable for the longer period. This article is not concerned with that uncertainty, only with the appropriate rate of return to use for a central (or intermediate) projection. For policy purposes, one must also look at stochastic projections (see, for example, Copeland, VanDerhei, and Salisbury 1999; and Lee and Tuljapurkar 1998). Despite the value of stochastic projections, OCACT's central projection plays an important role in thinking about policy and in the political process. Nevertheless, when making a long-run projection, one must realize that great uncertainty surrounds any single projection and the relevance of returns in any short period of time.

[11] Table 2 also shows the equity premiums relative to Treasury bills. Those numbers are included only because they arise in other discussions; they are not referred to in this article.

[12] For determining the equity premium shown in Table 2, the rate of return is calculated assuming that a dollar is invested at the start of a period and the returns are reinvested until the end of the period. In contrast to that geometric average, an arithmetic average is the average of the annual rates of return for each of the years in a period. The arithmetic average is larger than the geometric average. Assume, for example, that a dollar doubles in value in year 1 and then halves in value from year 1 to year 2. The geometric average over the 2-year period is zero; the arithmetic average of +100 percent and –50 percent annual rates of return is +25 percent. For projection purposes, one looks for an estimated rate of return that is suitable for investment over a long period. Presumably the best approach would be to take the arithmetic average of the rates of return that were each the geometric average for different historical periods of the same length as the average investment period within the projection period. That calculation would be close to the geometric average, since the variation in 35- or 40-year geometric

rates of return, which is the source of the difference between arithmetic and geometric averages, would not be so large.

[13] In considering recent data, some adjustment should be made for bond rates being artificially low in the 1940s as a consequence of war and postwar policies.

[14] Also relevant is the fact that the real rate on 30-year Treasury bonds is currently above 3.0 percent.

[15] Finance theory relates the willingness to hold alternative assets to the expected risks and returns (in real terms) of the different assets, recognizing that expectations about risk and return are likely to vary with the time horizon of the investor. Indeed, time horizon is an oversimplification, since people are also uncertain about when they will want to have access to the proceeds of those investments. Thus, finance theory is primarily about the difference in returns to different assets (the equity premium) and needs to be supplemented by other analyses to consider the expected return to stocks.

[16] With Treasury bonds, investors can easily project future nominal returns (since default risk is taken to be virtually zero), although expected real returns depend on projected inflation outcomes given nominal yields. With inflation-protected Treasury bonds, investors can purchase bonds with a known real interest rate. Since those bonds were introduced only recently, they do not play a role in interpreting the historical record for projection purposes. Moreover, their importance in future portfolio choices is unclear.

[17] In theory, for determining asset prices at which markets clear, one wants to consider marginal investments. Those investments are made up of a mix of marginal portfolio allocations by all investors and by marginal investors who become participants (or nonparticipants) in the stock and/or bond markets.

[18] This conclusion does not contradict the Modigliani-Miller theorem. Different firms with the same total return distributions but different amounts of debt outstanding will have the same total value (stock plus bond) and so the same total expected return. A firm with more debt outstanding will have a higher expected return on its stock in order to preserve the total expected return.

[19] Consideration of equilibrium suggests an alternative approach to analyzing the historical record. Rather than looking at realized rates of return, one could construct estimates of expected rates of return and see how they have varied in the past. That approach has been taken by Blanchard (1993). He concluded that the equity premium (measured by expectations) was unusually high in the late 1930s and 1940s and, since the 1950s, has experienced a long decline from that unusually high level. The high realized rates of return over this period are, in part, a consequence of a decline in the equity premium needed for people to be willing to hold stocks. In addition, the real expected returns on bonds have risen since the 1950s, which should have moderated the impact of a declining equity premium on expected stock returns. Blanchard examines the importance of inflation expectations and attributes some of the recent trend to a decline in expected inflation. He concluded that the premium in 1993 appeared to be around 2 percent to 3 percent and would probably not move much if inflation expectations remain low. He also concluded that decreases in the equity premium were likely to involve both increases in expected bond rates and decreases in expected rates of return on stocks.

[20] If current cash returns to stockholders are expected to grow at rate g, with projected returns discounted at rate r, this fundamental value is the current return divided by $(r - g)$. If r is smaller, fluctuations in long-run projections of g result in larger fluctuations in the fundamental value.

[21] Several explanations have been put forth, including: (1) the United States has been lucky, compared with stock investment in other countries, and realized returns include a premium for the possibility that the U.S. experience might have been different; (2) returns to actual investors are considerably less than the returns on indexes that have been used in analyses; and (3) individual preferences are different from the simple models that have been used in examining the puzzle.

[22] The timing of realized returns that are higher than required returns is somewhat more complicated, since recognizing and projecting such a trend will tend to boost the price of equities when the trend is recognized, not when it is realized.

[23] Nonprofit institutions, such as universities, and defined benefit plans for public employees now hold more stock than in the past. Attributing the risk associated with that portfolio to the beneficiaries of those institutions would further expand the pool sharing in the risk.

[24] More generally, the equity premium depends on the investment strategies being followed by investors.

[25] This tendency, known as mean reversion, implies that a short period of above-average stock returns is likely to be followed by a period of below-average returns.

[26] To quantify the importance of these developments, one would want to model corporate behavior as well as

investor behavior. A decline in the equity premium reflects a drop to corporations in the "cost of risk" in the process of acquiring funds for risky investment. If the "price per unit of risk" goes down, corporations might respond by selecting riskier investments (those with a higher expected return), thereby somewhat restoring the equity premium associated with investing in corporations.

[27] In considering the return to an individual from investing in stocks, the return is made up of dividends and a (possible) capital gain from a rise in the value of the shares purchased. When considering the return to all investment in stocks, one needs to consider the entire cash flow to stockholders, including dividends and net share repurchases by the firms. That suggests two methods of examining the consistency of any assumed rate of return on stocks. One is to consider the value of all stocks outstanding. If one assumes that the value of all stocks outstanding grows at the same rate as the economy (in the long run), then the return to all stocks outstanding is that rate of growth plus the sum of dividends and net share repurchases, relative to total share value. Alternatively, one can consider ownership of a single share. The assumed rate of return minus the rate of dividend payment then implies a rate of capital gain on the single share. However, the relationship between the growth of value of a single share and the growth of the economy depends on the rate of share repurchase. As shares are being repurchased, remaining shares should grow in value relative to the growth of the economy. Either approach can be calculated in a consistent manner. What must be avoided is an inconsistent mix, considering only dividends and also assuming that the value of a single share grows at the same rate as the economy.

[28] Gordon (1962). For an exposition, see Campbell, Lo, and MacKinlay (1997).

[29] The implausibility refers to total stock values, not the value of single shares—thus, the relevance of net share repurchases. For example, Dudley and others (1999) view a steady equity premium in the range of 1.0 percent to 3.0 percent as consistent with current stock prices and their projections. They assume 3.0 percent GDP growth and a 3.5 percent real bond return, both higher than the assumptions used by OCACT. Wadhwani (1998) finds that if the S&P 500 is correctly valued, he has to assume a negative risk premium. He considers various adjustments that lead to a higher premium, with his "best guess" estimate being 1.6 percent. That still seems too low.

[30] Dudley and others (1999) report a current dividend yield on the Wilshire 5000 of 1.3 percent. They then make an adjustment that is equivalent to adding 80 basis points to that rate for share repurchases, for which they cite Campbell and Shiller (1998). Wadhwani (1998) finds a current expected dividend yield of 1.65 percent for the S&P 500, which he adjusts to 2.55 percent to account for share repurchases. For a discussion of share repurchases, see Cole, Helwege, and Laster (1996).

[31] Stock prices reflect investors' assumptions about economic growth. If their assumptions differ from those used by OCACT, then it becomes difficult to have a consistent projection that does not assume that investors will be surprised.

[32] In considering these values, note the observation that a fall of 20 percent to 30 percent in advance of recessions is typical for the U.S. stock market (Wadhwani 1998). With OCACT assuming a 27 percent rise in the price level over the next decade, a 21 percent decline in real stock prices would yield the same nominal prices as at present.

[33] The importance of the assumed growth rate of GDP can be seen by redoing the calculations in Table 3 for a growth rate that is one-half of a percent larger in both the short and long runs. Compared with the original calculations, such a change would increase the ratios by 16 percent.

[34] Both scenarios are consistent with the Gordon formula, assuming a 2.75 percent adjusted dividend yield (without a drop in share prices) and a growth of dividends of 1.5 percent per year.

[35] With the steady-state scenario, a dollar in the market at the start of the steady state is worth 1.0425^t dollars t years later, if the returns are continuously reinvested. In contrast, under the market-correction scenario, a dollar in the market at the time of the drop in prices is worth $(1/2)(1.07^t)$ dollars t years later.

[36] The authors appear to assume that the Treasury rate will not change significantly, so that changes in the equity premium and in the return to stocks are similar.

[37] One could use equations estimated on historical prices to check the plausibility of intermediate-run stock values with the intermediate-run values needed for plausibility for the long-run assumptions. Such a calculation is not considered in this article. Another approach is to consider the value of stocks relative to the replacement cost of the capital that corporations hold, referred to as Tobin's q. That ratio has fluctuated considerably and is currently unusually high. Robertson and Wright (1998) have analyzed the ratio and concluded that a cumulative real decline in the stock market over the first decades of the 21st century has a high probability.

[38] As Wadhwani (1998, p. 36) notes, "Surveys of individual investors in the United States regularly suggest that they expect returns above 20 percent, which is obviously unsustainable. For example, in a survey conducted by Montgomery Asset Management in 1997, the typical mutual fund investor expected annual returns from the

stock market of 34 percent over the next 10 years! Most U.S. pension funds operate under actuarial assumptions of equity returns in the 8-10 percent area, which, with a dividend yield under 2 percent and nominal GNP growth unlikely to exceed 5 percent, is again, unsustainably high."

[39] There is no necessary connection between the rate of return on stocks and the rate of growth of the economy. There is a connection among the rate of return on stocks, the current stock prices, dividends relative to GDP, and the rate of growth of the economy.

[40] The impact of such a change in assumptions on actuarial balance depends on the amount that is invested in stocks in the short term relative to the amount invested in the long term. The levels of holdings at different times depend on both the speed of initial investment and whether stock holdings are sold before very long (as would happen with no other policy changes) or whether, instead, additional policies are adopted that result in a longer holding period, possibly including a sustained sizable portfolio of stocks. Such an outcome would follow if Social Security switched to a sustained level of funding in excess of the historical long-run target of just a contingency reserve equal to a single year's expenditures.

[41] "The annual rate of growth in total labor force decreased from an average of about 2.0 percent per year during the 1970s and 1980s to about 1.1 percent from 1990 to 1998. After 1998 the labor force is projected to increase about 0.9 percent per year, on average, through 2008, and to increase much more slowly after that, ultimately reaching 0.1 percent toward the end of the 75-year projection period" (Social Security Trustees Report, p. 55). "The Trustees assume an intermediate trend growth rate of labor productivity of 1.3 percent per year, roughly in line with the average rate of growth of productivity over the last 30 years" (Social Security Trustees Report, p. 55).

[42] Two approaches are available to answer this question. Since the Gordon formula, given above, shows that the return to stocks equals the adjusted dividend yield plus the growth of stock prices, one needs to consider how the dividend yield is affected by slower growth. In turn, that relationship will depend on investment levels relative to corporate earnings. Baker (1999b) makes such a calculation, which is not examined here. Another approach is to consider the return on physical capital directly, which is the one examined in this article.

[43] Using the Granger test of causation (Granger 1969), Carroll and Weil (1994) find that growth causes saving but saving does not cause growth. That is, changes in growth rates tend to precede changes in savings rates but not vice versa. For a recent discussion of savings and growth, see Carroll, Overland, and Weil (2000).

[44] One can also ask how a change in policy designed to build and maintain a larger trust fund in a way that significantly increases national saving might affect future returns. Such a change would plausibly tend to lower rates of return. The size of that effect depends on the size of investment increases relative to available investment opportunities, both in the United States and worldwide. Moreover, it depends on the response of private saving to the policy, including the effect that would come through any change in the rate of return. There is plausibly an effect here, although this article does not explore it. Again, the argument speaks to the level of rates of return generally, not to the equity premium.

[45] One can also ask how changed policies might affect future returns. A change in portfolio policy that included stocks (whether in the trust fund or in individual accounts) would plausibly lower the equity premium somewhat. That effect could come about through a combination of a rise in the Treasury rate (thereby requiring a change in tax and/or expenditure policy) and a fall in expected returns on stocks. The latter depends on both the underlying technology of available returns to real investments and the effect of portfolio policy on national saving. At this time, research on this issue has been limited, although it is plausible that the effect is not large (Bohn 1998; Abel 1999; Diamond and Geanakoplos 1999).

[46] For stochastic projections, see Copeland, VanDerhei, and Salisbury (1999); and Lee and Tuljapurkar (1998). OCACT generally provides sensitivity analysis by doing projections with several different rates of return on stocks.

[47] Cochrane (1997, p. 32) reaches a similar conclusion relative to individual investment: "We could interpret the recent run-up in the market as the result of people finally figuring out how good an investment stocks have been for the last century, and building institutions that allow wise participation in the stock market. If so, future returns are likely to be much lower, but there is not much one can do about it but sigh and join the parade."

Bibliography

Abel, Andrew B. 1999. *The Social Security Trust Fund, The Riskless Interest Rate, and Capital Accumulation.* NBER Working Paper 6991. Cambridge, Mass.: National Bureau of Economic Research.

Advisory Council on Social Security. 1997. *Report of the 1994-1996 Advisory Council on Social Security*. Vol. 1, *Findings and Recommendations*. Washington, D.C.

Baker, Dean. 1998. "Saving Social Security in Three Steps." Briefing Paper, Economic Policy Institute, Washington, D.C.

_____. 1999a. "Two Cheers for Clinton's Social Security Plan." *The American Prospect,* Vol. 10, No. 44, pp. 82-85.

_____.1999b. Letter to Martin Feldstein, May 15. Available at www.preamble.org.

Baker, Dean, and Mark Weisbrot. 1999. *Social Security: The Phony Crisis*. Chicago, Ill.: University of Chicago Press.

Blanchard, Olivier J. 1993. "Movements in the Equity Premium." *Brookings Papers on Economic Activity*, No. 2, pp. 75-138.

Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds. 1999. *1999 Annual Report.* Washington, D.C.: U.S. Government Printing Office.

Bohn, Henning. 1998. "Social Security Reform and Financial Markets." In Federal Reserve Bank of Boston, *Social Security Reform: Links to Saving, Investment, and Growth*. Boston, Mass.: Federal Reserve Bank of Boston.

Bosworth, Barry, and Gary Burtless. 1998. "Social Security Reform in a Global Context." In Federal Reserve Bank of Boston, *Social Security Reform: Links to Saving, Investment, and Growth*. Boston, Mass.: Federal Reserve Bank of Boston.

Brav, Alon, and Christopher C. Geczy. 1996. "An Empirical Resurrection of the Simple Consumption CAPM with Power Utility." Working Paper, University of Chicago.

Campbell, John Y.; Andrew W. Lo; and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, N.J.: Princeton University Press.

Campbell, John Y., and Robert J. Shiller. 1998. “Valuation Ratios and the Long-Run Stock Market Outlook.” *Journal of Portfolio Management*, Vol. 24, No. 2, pp. 11-26.

Campbell, John Y., and Luis M. Viceira. 1999. “Consumption and Portfolio Decisions When Expected Returns Are Time Varying.” *Quarterly Journal of Economics* (May), pp. 433-495.

Carroll, Christopher D.; Jody Overland; and David N. Weil. 2000. “Saving and Growth with Habit Formation.” *American Economic Review*, Vol. 90, No. 3, pp. 341-355.

Carroll, Christopher D., and David N. Weil. 1994. “Saving and Growth: A Reinterpretation.” *Carnegie-Rochester Conference Series on Public Policy,* No. 40, pp. 133-192.

Case, Karl E., and Robert J. Shiller. 1988. “The Behavior of Home Buyers in Boom and Post-Boom Markets.” *New England Economic Review* (November/December), pp. 29-46.

Cochrane, John H. 1997. “Where is the Market Going? Uncertain Facts and Novel Theories.” *Economic Perspectives,* Vol. 21, Federal Reserve Bank of Chicago. (Also published as NBER Working Paper 6207, Cambridge, Mass.: National Bureau of Economic Research.)

Cole, Kevin; Jean Helwege; and David Laster. 1996. “Stock Market Valuation Indicators: Is This Time Different?” *Financial Analysts Journal,* Vol. 52, No. 3, pp. 56-64.

Copeland, Craig; Jack VanDerhei; and Dallas L. Salisbury. 1999. *Social Security Reform: Evaluating Current Proposals: Latest Results of the EBRI SSASIM2 Policy Simulation Model*. EBRI Issue Brief 210. Washington, D.C.: Employee Benefit Research Institute.

Diamond, Peter. 1999. *Issues in Privatizing Social Security*. Cambridge, Mass.: MIT Press.

_______. 2000. “Administrative Costs and Equilibrium Charges with Individual Accounts.” In *Administrative Aspects of Investment-Based Social Security Reform*, edited by John B. Shoven. Chicago and London: University of Chicago Press, pp. 137-162.

Diamond, Peter, and John Geanakoplos. 1999. *Social Security Investment in Equities I: Linear Case*. Working Paper 2. Center for Retirement Research at Boston College, Chestnut Hill, Mass.

Dudley, William; Jan Hatzius; Joanne Hill; Rajib Pal; and Maria Tsu. 1999. “An Analysis of Social Security Trust Fund Investment into Equities.” Goldman Sachs Global Economics Paper 18, New York.

Glassman, James K., and Kevin A. Hassett. 1999. *Dow 36,000*. New York: Times Business.

Gordon, Myron. 1962. *The Investment, Financing, and Valuation of the Corporation.* Homewood, Ill.: Irwin.

Gordon, Roger. 1985. "Taxation of Corporate Capital Income: Tax Revenues versus Tax Distributions." *Quarterly Journal of Economics,* Vol. 100, No. 1, pp. 1-27.

Goyal, Amit. 1999. "Demographics, Stock Market Flows, and Stock Returns." Anderson Graduate School of Management, University of California at Los Angeles. Unpublished.

Granger, C.W.J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, Vol. 37, No. 3, pp. 424-438.

Heaton, John, and Deborah Lucas. 2000. "Stock Prices and Fundamentals." In *NBER Macroeconomics Annual,* edited by Ben S. Bernanke and Julio Rotemberg. Cambridge, Mass.: MIT Press.

Investment Company Institute. 1999. *Mutual Fund Fact Book* (May), p. 22. Available at www.ici.org.

Kaplow, Louis. 1994. "Taxation and Risk-taking: A General Equilibrium Perspective." *National Tax Journal,* Vol. 47, pp. 789-798.

Kennickell, Arthur B.; Martha Starr-McCluer; and Annika Sundén. 1997. "Family Finances in the U.S.: Recent Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin,* Vol. 83 (June), pp. 1-24.

Lee, Ronald, and Shripad Tuljapurkar. 1998. "Stochastic Forecasts for Social Security." In *Frontiers in the Economics of Aging*, edited by David Wise. Chicago, Ill.: University of Chicago Press, pp. 393-420.

Liang, J. Nellie, and Steven A. Sharpe. 1999. "Share Repurchases and Employee Stock Options and Their Implications for S&P 500 Share Retirements and Expected Returns." Division of Research and Statistics, Federal Reserve Board, Washington, D.C. Unpublished.

Mehra, Rajnish, and Edward Prescott. 1985. "The Equity Premium Puzzle." *Journal of Monetary Economics*, Vol. 15, pp. 145-161.

Mankiw, N.G., and S. Zeldes. 1991. "The Consumption of Stockholders and Non-Stockholders." *Journal of Financial Economics*, Vol.17, pp. 211-219.

Munnell, Alicia, and Annika Sundén (with the assistance of Cynthia Perry and Ryan Kling). 1999. *Investment Practices of State and Local Pension Funds: Implications for Social Security*. Working Paper 1. Center for Retirement Research at Boston College, Chestnut Hill, Mass.

National Academy of Social Insurance. 1998. *Evaluating Issues in Privatizing Social Security.* Report of the Panel on Privatization of Social Security. Available at www.nasi.org and as Diamond 1999.

Odean, Terrance. 1998. "Do Investors Trade Too Much?" Graduate School of Management,University of California, Davis. Unpublished.

Philips, Thomas K. 1999. "Why Do Valuation Ratios Forecast Long-Run Equity Returns?" *Journal of Portfolio Management*, Vol. 25, No. 3, pp. 39-44.

Poterba, James M. 1998. "Population Age Structure and Asset Returns: An Empirical Investigation." Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass. Unpublished.

Poterba, James M., and Andrew A. Samwick. 1995. "Stock Ownership Patterns, Stock Market Fluctuations, and Consumption." *Brookings Papers on Economic Activity*, No. 2, pp. 295-372.

Rea, John D., and Brian K. Reid. 1998. *Trends in the Ownership Cost of Equity Mutual Funds*. Perspective, Vol. 4, No. 3. Washington, D.C.: Investment Company Institute.

_______.1999. *Total Shareholder Cost of Bond and Money Market Mutual Funds*. Perspective, Vol. 5, No. 3. Washington, D.C.: Investment Company Institute.

Robertson, Donald, and Stephen Wright. 1998. "The Good News and the Bad News about Long-Run Stock Market Returns." Faculty of Economics and Politics, University of Cambridge. Unpublished.

Samuelson, Paul A. 1989. "A Case at Last for Age-Phased Reduction in Equity." *Proceedings of the National Academy of Sciences*, Vol. 86, pp. 9048-9051.

Shafir, Eldar; Peter Diamond; and Amos Tversky. 1997. "Money Illusion." *Quarterly Journal of Economics*, Vol. 112, No. 2, pp. 341-374.

Sharpe, Steven. 1999. "Stock Prices, Expected Returns, and Inflation." Federal Reserve Board, Washington, D.C. Unpublished.

Siegel, Jeremy J. 1998. *Stocks for the Long Run*. New York: McGraw Hill, p. 318.

_______. 1999. "The Shrinking Equity Premium: Historical Facts and Future Forecasts." Wharton School, University of Pennsylvania, Philadelphia, Pa. Unpublished.

Thompson, Lawrence. 1998. *Older and Wiser: The Economics of Public Pensions*. Washington, D.C.: Urban Institute Press.

Vissing-Jorgensen, Annette. 1997. "Limited Stock Market Participation." Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass. Unpublished.

Wadhwani, Sushil B. 1998. "The U.S. Stock Market and the Global Economic Crisis." Tudor Investment Corporation, New York. Unpublished.

Appendix A:

Alternative Method for Determining the Ratio of Stock Value to GDP

Variables

r …… rate of return on stocks
g …… rate of growth of both GDP and dividends
a …… adjusted dividend yield at time 0
$P(t)$ … aggregate stock value at time t
$Y(t)$ … GDP at time t
$D(t)$ … dividends at time t

Equations

$$Y(t) = Y(0)e^{gt}$$

$$D(t) = D(0)e^{gt} = aP(0)e^{gt}$$

$$dP(t)/dt = rP - D(t) = rP - aP(0)e^{gt}$$

Solving the differential equation, we have:

$$\begin{aligned} P(t) &= P(0)\{(r-g-a)e^{rt} + ae^{gt}\}/(r-g) \\ &= P(0)\{e^{rt} - (a/(r-g))(e^{rt} - e^{gt})\} \end{aligned}$$

Taking the ratio of prices to GDP, we have:

$$\begin{aligned} P(t)/Y(t) &= \{P(0)/Y(0)\}\{(r-g-a)e^{(r-g)t} + a\}/(r-g) \\ &= \{P(0)/Y(0)\}\{(e^{(r-g)t} - (a/(r-g))(e^{(r-g)t} - 1)\} \end{aligned}$$

Consistent with the Gordon formula, a constant ratio of P/Y (that is, a steady state) follows from $r = g + a$. As a non-steady-state example—with values of .07 for r, .015 for g, and .03 for a—$P(75)/Y(75) = 28.7P(0)/Y(0)$.

Appendix B:

Calculation Using the Gordon Equation

In discrete time, once we are in a steady state, the Gordon growth model relates a stock price P at time t to the expected dividend D in the following period, the rate of growth of dividends G, and the rate of return on the stock R. Therefore, we have:

$$P_t = D_{t+1}/(R-G) = (1+G)D_t/(R-G)$$

We denote values after a decade (when we are assumed to be in a steady state) by P' and D' and use an "adjusted" initial dividend that starts at a ratio X times current stock prices. Thus, we assume that dividends grow at the rate G from the "adjusted" current value for 10 years, where G coincides with GDP growth over the decade. We assume that dividends grow at G' thereafter, which coincides with long-run GDP growth. Thus, we have:

$$\begin{aligned} P'/P &= (1+G')D'/((R-G')P) \\ &= (1+G')D(1+G)^{10}/((R-G')P) \\ &= X(1+G')(1+G)^{10}/(R-G') \end{aligned}$$

For the basic calculation, we assume that R is .07, G is .02, G' is .015. In this case, we have:

$$P'/P = 22.5X$$

Thus, for initial ratios of adjusted dividends to stock prices of .02, .025, .03, and .035, P'/P equals .45, .56, .67 and .79, respectively. Subtracting those numbers from 1 yields the required decline in the real value of stock prices as shown in the first column of Table 3. Converting them into nominal values by multiplying by 1.27, we have values of .57, .71, and .86. If the long-run stock return is assumed to be 6.5 percent instead of 7.0 percent, the ratio P'/P is higher and the required decline is smaller. Increasing GDP growth also reduces the required decline. Note that the required declines in stock values in Table 3 is the decline in real values; the decline in nominal terms would be less.

Appendix C:

A Cobb-Douglas Solow Growth Model in Steady State

Variables

Y .... output
K .... capital
L .... labor
a .... growth rate of Solow residual
g .... growth rate of both K and Y
n .... growth rate of labor
b .... share of labor
s .... savings rate
c .... depreciation rate
$MP(K)$ ... marginal product of capital

Equations

$$\log[Y] = at + b\log[L] + (1 - b)\log[K]$$
$$(dL/dt)/L = n$$
$$(dY/dt)/Y = (dK/dt)/K = g$$
$$dK/dt = sY - cK$$
$$(dK/dt)/K = sY/K - c$$
$$Y/K = (g + c)/s$$
$$MP(K) = (1 - b)Y/K = (1 - b)(g + c)/s$$
$$g = a + bn + (1 - b)g$$
$$g = (a + bn)/b$$
$$MP(K) = (1 - b)\{(a + bn)/(bs) + c/s\}$$
$$dMP(K)/da = (1 - b)/(bs)$$
$$dg/da = 1/b$$

Assume that the share of labor is .75 and the gross savings rate is .2. Then the change in the marginal product of capital from a change in the growth rate is:

(Note that these are gross savings, not net savings. But the corporate income tax reduces the return to savers relative to the return to corporate capital, so the derivative should be multiplied by roughly 2/3.)

$$dMP(K)/dg = (dMP(K)/da)/(dg/da) = (1 - b)/s == .25/.2$$

Similarly, we can consider the effect of a slowdown in labor force growth on the marginal product of capital:

$$dMP(K)/dn = (1 - b)/s$$

$$dg/dn = 1$$

$$dMP(K)/dg = (dMP(K)/dn)/(dg/dn) = (1 - b)/s == .25/.2$$

(This is the same expression as when the slowdown in economic growth comes from a drop in technical progress.)

Turning to the effects of changes in the savings rate, we have:

$$dMP(K)/ds = -MP(K)/s == .5$$

Thus, the savings rate has a large impact on the marginal product of capital as well.

Both of these effects are attenuated to the extent that the economy is open and rates of return in the United States change less because some of the effect occurs abroad.

What Are Reasonable Long-Run Rates of Return to Expect on Equities?

John B. Shoven, Professor of Economics
Stanford University
July 20, 2001

I. Introduction

The average inflation-adjusted rate of return on large capitalization stocks from 1926-2000 was 9.7 percent (Ibbotson (2001)). Over the same period of time, the average real return on Treasury Bills was 0.8 percent while it was 2.7 percent on long-term U.S. government bonds. The premium of stocks over long-term government bonds was 7.0 percent.[1]

The question of interest is not what happened in the past, but what is likely to happen over the next fifty or seventy-five years. Will stocks once again outperform bonds by 7 percent? One needs to be humble when predicting the stock market, although ironically it may be easier to look further into the future than it is to predict what will happen over the next few months or years. In the very long-run, stock returns are more likely to be driven by fundamentals, while in the short-run price movements can appear to have a life of their own.

There are a number of reasons to expect the return on stocks and the premium of the return of stocks over bonds to be lower than over the last three-fourths of the twentieth century. This paper reviews those reasons and concludes with an estimate of the expected long-run real rate of return for equities and an implied equity premium.

II. Dividends Are Obsolete

Traditional equity valuation models (Gordon(1962)) are based on the value of shares being equal to the present value of future dividends. This leads to the result that the expected return to holding stocks is equal to the current dividend yield plus the growth rate in dividend payments. This basic structure is behind most analysis of long-run stock returns today (see, for example, Campbell and Shiller (2001)). The problem with this framework is that dividends are only one way for the corporate sector to transfer money to shareholders and a particularly tax inefficient way at that (Shoven (1987)). Dividend payments are fully taxable for investors who do not have their equity sheltered in pension accounts or other tax deferred or exempt vehicles. In contrast, companies can buy their own shares from their shareholders and achieve the same cash transfer with much lower taxation. With a share repurchase, some of the money is treated as a return of basis and the rest is treated as a capital gain. The tax saving can be enormous. Companies began to take advantage of share repurchases in a significant way in the mid-1980s. In recent years the

[1] All of these numbers are arithmetic averages. The geometric mean real return on large capitalization stocks was 7.7%, whereas it was 2.2% on long-term government bonds. The geometric premium of stocks over long-term government bonds was thus 5.5%.

aggregate amount of share repurchases has exceeded dividends and is currently running at about \$150 billion per year (Liang and Sharpe (1999)). Clearly share repurchases can no longer be treated as a footnote in a story primarily concerned with dividends as a mechanism for transferring cash to shareholders. Companies can also buy the shares of other companies. The extreme form of this is a cash merger. Once again, cash is transferred from companies to shareholders, affecting the valuation of shares. While it is hard to get precise information on the amounts involved, the cash transferred to shareholders via cash mergers is almost certainly even larger than the amount in share repurchases. The point of this is to emphasize that dividends are a choice variable and dividend-price ratios should not be a fundamental building block of share valuation or long-run shareholder return. In fact, it is not clear that companies founded in the 1980s and later will ever pay dividends in the same way as older companies.

III. The Model

The original Gordon model had the intrinsic value of the firm depending on dividends and the growth rate of dividends such that

$$V = \frac{D}{k - g}$$

$$or$$

$$k = \frac{D}{V} + g$$

where $\boldsymbol{V}$ is the intrinsic value of the equity, $\boldsymbol{D}$ is the cash dividends, $\boldsymbol{k}$ is capital asset pricing model required rate of return for equity of this risk class, and $\boldsymbol{g}$ is the growth rate of dividends.

The modernized Gordon model can be represented as

$$k = \theta\frac{E}{P} + (1 - \theta)\rho$$

where $\boldsymbol{k}$ is the expected real return to equity, θ is the fraction of earnings paid out to shareholders via dividends or share repurchases, $\boldsymbol{E}$ is earnings per share, $\boldsymbol{P}$ is the current share price and ρ is the ROE (return on equity).[2] The first right hand side term replaces the dividend yield of the Gordon model with the cash-from-earnings yield including share repurchases. The second term on the right hand side is simply the growth rate of future cash flows and indicates that it depends on the amount of retained earnings and the rate of return associated with those retained earnings.[3] This equation is an identity if the various parameters in it remain constant. On the other hand, the observed realized rate of return to holding equity can deviate widely from the value given in the equation if the parameters (particularly the earnings-price ratio) change.

[2] Share repurchases can be added to the cash flow yield as in the equation in the paper or added to the growth rate term, but not both. Investors who don't participate in a share repurchase benefit from owning a growing fraction of the company. Investors taken as a group receive the cash from a share repurchase just like a dividend. The company's opportunities are the same after the payment of an equivalent amount in dividends or share repurchases.

[3] I have not required ρ to equal k in the long-run steady state, although an argument could be made that they should be equated. If they are equal, then the expected return to equity is independent of payout policy and is simply equal to the reciprocal of the P-E ratio.

IV. Steady State Returns

The model just presented gives the steady state real returns that investors can expect to receive from equity markets. The steady state assumption is that aggregate corporate earnings, aggregate dividends, the total market capitalization of stocks, the total money used for share repurchases, and GDP all grow at the same long-run rate. In such a scenario, the price-earnings ratio would remain stable. However, the role of share repurchases would continue to be very important. Due to the declining number of shares, stock prices, dividends per share, and earnings per share would all grow at a rate faster than GDP and the other aggregates. The equilibrium real rate of return to owning stock would be the total of three terms: the dividend rate, the share repurchase rate, and the steady-state growth rate of aggregates in the economy including GDP. That is,

$$k = \frac{D}{P} + \frac{S}{P} + g$$

where S is share repurchases and g is the common steady-state growth rate of economic aggregates. This is simply a different way to write the equation of the previous section. It does highlight that real share prices would go up at the rate of g plus the rate of net share repurchases. To make the equivalence with the previous formulation clear note that

$$\theta \frac{E}{P} = \frac{D}{P} + \frac{S}{P} \; and \; (1-\theta)\rho = g$$

V. The Big Question: Future P-E Ratios

The very difficult question is whether the current price-earnings ratio of roughly 25 represents a new steady-state level. Of course, no one would assume that fluctuations in price-earnings ratios will cease, but will 25 be the average level for the next 50 or 75 years? My guess is that the long-run steady state level for the price-earnings ratio will be somewhere between its current level (24 as I write this on July 20, 2001) and its average level over the past 75 years of approximately 15. A reasonable guess would be that P-E ratios might average 20 over the next 50 to 75 years. What would be the consequences of a steady-state P-E ratio of 20 on real expected stock returns? That means that ***(E/P)*** would average .05. Firms pay out somewhere between half and three-fourths of their earnings as dividends and net share repurchases, so a reasonable value for θ is 0.625. The ROE of retained earnings is approximately 8 percent, so ρ can be set at that level. [4] Substituting these values into the model gives

$$k = (.625)(.05) + (.375)(.08) = .03125 + .03 = .06125$$

This model and these parameters predict the expected long-run real return to equity to be 6.125 percent.

---

[4] This value is roughly consistent with the rate of return to corporate capital reported in Poterba (1997).

From its current levels, the S&P 500 would not have to crash to reach a P-E level of 20. In fact, the current S&P forecast for next year's earnings of the S&P 500 is $62.88, so the market is currently selling at 19.3 times next year's predicted earnings. That means that if the market were to go up 3.5 percent over the next year and the 2002 earnings forecasts panned out exactly, then by mid-2002 the market would be selling for exactly 20 times earnings. Obviously, there are other combinations of earnings realizations and price appreciation that would allow the market to equilibrate at a P-E of 20 over the next couple of years.

What would be the consequences of a long run average price-earnings ratio of 15 rather than 20? This would put the P-E ratio close to its average level for the past 75 years. In the short-run this implies that the current market is almost 40 percent overvalued and would indicate that near-term stock returns might be quite poor. On the other hand, once the correction is completed and the equilibrium P-E ratio of 15 is established the real rate of return to equities could average slightly better than 7 percent. If we stick with the assumption that ρ is .08, the expected real return to equity would be in the 7 to 7.5 percent range for all reasonable cash-payout rates (i.e. for all reasonable values of θ).

So, we see that the assumed equilibrium price-earnings rate is important. It should be noted that a near-term market correction to bring about a P-E ratio of 15 would not hurt the proposed Social Security individual accounts as long as it occurred before they had accumulated significant balances. In general, the fact that the individual accounts do not yet exist and will have small balances over the next several years even if they are established soon means that the timing of returns matters a lot. Low returns over the next several years followed by high returns would be much better for the balances in these new Social Security individual accounts than high returns first followed by low ones. There is a big difference between the circumstances of someone who has a lot of wealth but is not saving and someone who is just starting to systematically accumulate assets. The non-saving wealth holder is indifferent to the order of returns. However, the systematic saver has little at stake early in his or her accumulation period, but much more at stake later. Even if real stock returns average 6.0 percent over the next 50 years, the Social Security individual account holders would prefer a pattern where the real returns averaged 2.0 percent for the first decade and 7.0 percent thereafter rather than a pattern of 10.0 percent in the first decade and 5.0 percent thereafter.

VI. The Long-Run Outlook for Equity Rates of Return

My own estimate for the long-run real return to equities looking forward is 6 to 6.5 percent. I come to that using roughly the parameters chosen above. If the P-E ratio fluctuates around 20, the cash payouts to shareholders should range from 3 to 3.5 percent. I am relatively optimistic about the possible steady-state growth rate of GDP and would choose 3 percent for that number.[5]

[5] It should be noted that the Trustees are projecting long-run average growth in aggregate labor income of slightly less than 2 percent. If 2 percent were the steady-state growth rate rather than three percent, then that would lower my prediction for equilibrium real stock returns by 0.5 percent. The reason that a one-percent drop in the economy wide growth rate would not lower stock returns by a full one percent is that the lower growth rate would require lower retained earnings and permit a higher rate of payout of earnings. For example, you then could support a value of θ of .75 with an E-P ratio of .05 and a value of ρ of .08.

That leads me to my 6 to 6.5 percent real rate of return range. While this is the range that I would choose as the expected return to equities, it does not indicate the degree of uncertainty about actual outcomes over the next 50-75 years. I think there is a great deal of uncertainty about long-run equity returns. A range of outcomes as wide as 2.0 to 10.0 percent would not strike me as unreasonable. Even this wide range of possible outcomes indicates that the 9.7 percent real return that stocks actually earned over the 1926-2000 period is quite unlikely to be repeated.

VII. Why Won't Equity Returns Be As Good in the 21st Century?

Why is it somewhat unlikely that the future returns will be as favorable as the past returns? There actually are quite a few reasons. First, share prices went up faster in the last twenty years than the value of the underlying capital. This relative price appreciation of paper claims to real assets is unlikely to continue over the long haul. Second, of the entire world's equity markets, the American market was the strongest over the last 75 years (see, Jorion and Goetzmann (1999)). While we might come in first again over the next half or three-quarters of a century, one shouldn't count on it. Third, the nature of stockholders has changed dramatically over the last few decades, with far more of the market being held by pension accounts. Whereas stock holdings used to be concentrated amongst the superrich, there has been a noticeable democratization of shareholding over the post World War II period. While it is speculative to be sure, one could argue that the degree of risk aversion displayed in the market has decreased as the market has become more democratic. Fourth, the changing demographics with the increase in the number of elderly relative to the number of working age adults can dampen the demand for financial assets (Schieber and Shoven (1997) and Abel (2001)).[6] Fifth, stock returns in the past may have been enhanced due to low *ex-post* real returns of long-term bonds. These low real returns were due to unexpectedly high inflation, particularly in the 1960s and 1970s. The total impact of these and other arguments is an equity premium that is likely to be considerably smaller than that observed since 1926.

VIII. The Equity Premium Will Be Lower Because Real Interest Rates Are Higher

The real return on long-run (30-year) inflation-indexed Treasury securities (TIPS) today is about 3.5 percent. Presumably the expected real return on regular nominal Treasury bonds is at least as high. If one uses my central guess for the average real return on equity markets of 6.0 to 6.5 percent, that leaves an equity premium on the order of 2.5 to 3.0 percent. Of course, real interest rates may drift down from current levels, increasing the equity premium. In fact, Social Security currently assumes that long-term government bonds will yield 3.0 percent in the future. That strikes me as reasonable and would not cause me to materially change my 6.0 to 6.5 percent range for the expected long-run real return on equities. Obviously, that leaves an equity premium of 3.0 to 3.5 percent, far lower than experienced during the last three-fourths of the 20th Century.

[6] For a skeptical view on the impact of demographics on asset prices see Poterba (2001).

IX. Which Rate To Use for Projections?

The next issue is whether one should use the expected equity returns to estimate the future balance of an equity portfolio or should one use the return on safe inflation-indexed government securities. On balance, I favor using the safe bond return on the argument that the extra expected return on equities is compensated for by the extra variance in the outcomes. Both the expected and median return for equities is almost certainly greater than for safe bonds. However, in order for markets to be in equilibrium, the poor equity outcomes must be worse than bond returns. Therefore, a scenario analysis for equity investments would, in my opinion, have to include outcomes worse than bonds as well as those better than for a bond portfolio. I find it preferable to simply calculate the outcomes with a safe investment strategy such as 100 percent Treasury Inflation-Protected Securities and then state that the expected outcome would be higher with stocks in the portfolio but that the risk would be correspondingly greater. The "no free lunch" saying is as true in finance as in the rest of the economy. The extra return of a stock heavy portfolio is matched by the extra riskiness (MaCurdy and Shoven (2000)).

One aside that the discussion of equity premium brings up is the useful role that government bonds play in anchoring financial returns and in providing a relatively risk-free asset alternative. The discussion in Washington of eliminating the publicly held federal debt should at least consider the value of such debt to financial markets. Another point worth remembering is that the traditional pay-as-you-go defined benefit structure is not without risk. The risks of a PAYGO system depend on fertility rates, immigration rates, mortality rates, labor force participation, and worker productivity. The risks of the defined benefit program are not perfectly correlated with the risks of individual accounts invested in private securities. One of the strongest arguments in favor of individual accounts is risk diversification. Clearly more work should be done to quantify the covariance between financial returns and the factors influencing the sustainability of a PAYGO system.

X. Conclusions

My best guess for a real equity return over a long-horizon is 6.0 to 6.5 percent per year. I suggest that Social Security lower its intermediate assumption for real equity returns from its current level of 7.0 percent to 6.5 percent or slightly lower. The narrowness of my range for the expected return does not represent a high degree of certainty about the actually realized real return on equities over the next 50-75 years. Throughout this note I have used terms like "best guess." That was totally intentional. Even if forecasting stock returns is easier over long horizons, it still isn't science. To put this concretely, I think that there is something like a 5 percent chance that real stock returns over the next 50 years will be worse than 2.5 percent and there is similarly something like a 5 percent chance that they will exceed 9.5 percent. While it is possible that stocks will underperform bonds over that horizon, it is quite unlikely. However, I think there is only a very slight chance that stocks will outperform bonds in the future by as much as they have in the past. That is, the equity premium is likely to be lower than it has been. My own best guess for the equity premium (stock return over the return on long-term government bonds) is 3.0 to 3.5 percent.

References

Abel, Andrew B., "Will Bequests Attenuate the Predicted Meltdown in Stock Prices When Baby Boomers Retire?" National Bureau of Economic Research Working Paper No. 8131, February 2001.

Campbell, John Y. and Robert J. Shiller, "Valuation Ratios and the Long-Run Stock Market Outlook: An Update," National Bureau of Economic Research Working Paper No. 8221, April 2001.

Gordon, Myron J., *The Investment, Financing and Valuation of the Corporation*, Irwin, Homewood, Illinois, 1962.

Ibbotson Associates, *Stocks, Bonds, Bills and Inflation – 2001 Yearbook*, Ibbotson Associates, Inc., Chicago, 2001.

Jorion, Philippe and William N. Goetzmann, "Global Stock Markets in the Twentieth Century," *Journal of Finance*, Volume 54, No. 3, 1999, pp.953-980.

MaCurdy, Thomas E. and John B. Shoven, "Asset Allocation and Risk Allocation: Can Social Security Improve Its Future Solvency Problem by Investing in Private Securities?" in *Risk Aspects of Investment-Based Social Security Reform*, John Y. Campbell and Martin Feldstein, eds., NBER-University of Chicago Press, 2001, pp. 11-40.

Liang, J. Nellie and Steven A. Sharpe, "Share Repurchases and Employee Stock Options and Their Implications for S&P 500 Share Retirements and Expected Returns," Federal Reserve Board Working Paper, 1999.

Poterba, James M., "The Rate of Return to Corporate Capital and Factor Shares: New Estimates Using Revised National Income Accounts and Capital Stock Data." National Bureau of Economic Research Working Paper No. 6263, November 1997.

Poterba, James M., "Demographic Structure and Asset Returns," *Review of Economic and Statistics*, forthcoming, 2001

Schieber, Sylvester and John Shoven, "The Consequences of Population Aging on Private Pension Fund Saving and Asset Markets," in *Public Policy Toward Pensions*, Sylvester Schieber and John Shoven, eds., A Twentieth Century Fund Book, 1997, pp. 219-246.

Shoven, John B., "The Tax Consequences of Share Repurchases and Other Non-dividend Cash Payments to Equity Owners," in *Tax Policy and the Economy*, Lawrence H. Summers, ed., Volume 1, 1987, pp. 29-54.

Biographies of Authors

John Y. Campbell

John Y. Campbell grew up in Oxford, England, and received a B.A from Oxford in 1979. He came to the United States to attend graduate school, earning his Ph. D. from Yale in 1984. He spent the next ten years teaching at Princeton, moving to Harvard in 1994 to become the first Otto Eckstein Professor of Applied Economics. Campbell has co-edited the *American Economic Review* and currently edits the *Review of Economics and Statistics*; he is a Fellow of the Econometric Society and the American Academy of Arts and Sciences, and a Research Associate and former Director of the Program in Asset Pricing at the National Bureau of Economic Research. His research concerns asset markets, the macroeconomy, and the links between them. His graduate-level textbook on empirical finance, *The Econometrics of Financial Markets*, written with Andrew Lo and Craig MacKinlay, was published by Princeton University Press in 1997. His latest book on *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, with Luis Viceira, will be published by Oxford University Press in 2001. Campbell is also a founding partner of Arrowstreet Capital, LP, a quantitative asset management firm in Cambridge, Massachusetts.

Peter A. Diamond

Peter Diamond is an Institute Professor at the Massachusetts Institute of Technology, where he has taught since 1966. He received his B.A. in Mathematics from Yale University in 1960 and his Ph.D. in Economics from M.I.T. in 1963. He has been President of the Econometric Society and Vice-President of the American Economic Association. He is a Founding Member and member of the Board of the National Academy of Social Insurance, where he has been President and Chair of the Board. He is a Fellow of the American Academy of Arts and Sciences and a Member of the National Academy of Sciences. He has written on behavioral economics, public finance, social insurance, uncertainty and search theories, and macroeconomics. He was Chair of the Panel on Privatization of Social Security of the National Academy of Social Insurance, whose report, *Issues in Privatizing Social Security* has been published by M.I.T. Press. He has written about social security in Chile, Germany, Italy, the Netherlands, Sweden and the U.S.

John B. Shoven

John Shoven is a member of Stanford University's Economics Department, where he holds the Charles R. Schwab Professorship. The holder of a Ph.D. in economics from Yale University, Dr. Shoven has been at Stanford since 1973, serving as Chairman of the Economics Department from 1986 to 1989, as Director of the Center for Economic Policy Research from 1989 to 1993, and as Dean of the School of Humanities and Sciences form 1993 to 1998. An expert on tax policy, Dr. Shoven was a consultant for the U.S. Treasury Department from 1975 to 1988. The author of approximately eighty professional articles and ten books, he has been a visiting professor at Harvard University, the London School of Economics, Kyoto University and Monash University. In 1995 he was elected a fellow of the American Academy of Arts and Sciences. Dr. Shoven has participated in various Hoover Programs and conferences, including the 1997 symposium "Facing the Age Wave," at which he addressed the taxing of pensions as an illustration of tax policy that seems to have gone awry and that may limit the most important form of savings in America. He also contributed a chapter to the book that resulted from the symposium.

Appendix

Equity Yield Assumptions Used by the Office of the Chief Actuary, Social Security Administration, to Develop Estimates for Proposals with Trust Fund and/or Individual Account Investments

Stephen C. Goss
Chief Actuary
May 8, 2001

Initial Assumptions in 1995

The Office of the Chief Actuary (OCACT) has been making estimates for proposals including investments in equities since 1995. A memorandum dated May 12, 1995 presented estimates for the Kerrey-Simpson proposal which included both individual accounts (with the opportunity for equity investment) and provision for investment of 25 percent of OASDI trust fund assets in equities. The assumed average real annual yield on equities for these estimates was 7 percent, consistent with the assumption developed for estimates being produced concurrently for the 1994-96 Advisory Council on Social Security.

Historical analysis of equity yields during the 20th century using Ibbottson data was provided to the Council by Joel Dickson of the Vanguard Group. Based on this analysis, the Advisory Council members and the OCACT agreed that the 7-percent average annual real yield experienced for the 20th century, particularly for the period beginning 1926, seemed to represent a reasonable assumption for an average real yield over long periods in the future as had occurred in the past. It was recognized that this average yield level was recorded rather consistently over long periods of time in the past which incorporated complete market cycles. The work of Dr. Jeremy Siegel of the Wharton School was also noted as supporting a long-term average yield on equities of about 7 percent.

Council Chairman Edward Gramlich noted that the equity market was then currently priced at a level above the historical average, as indicated by relatively high price-to-earnings (PE) ratios. However, it was agreed that in the future market cycles would continue, likely resulting in yields for investments made in successive future years that would average close to the average yields of the past. Estimates produced for the three proposals developed for the Advisory Council (included in Appendix 2 of Volume 1 of the Council's Report) used a 7-percent average real equity yield as an intermediate assumption. Estimates were also produced assuming that equities would achieve a long-term average yield no higher than the yield on long-term U.S. Government marketable securities (Treasury securities), in order to illustrate both the sensitivity of estimates to this assumption and the uncertainty about the likely average yield on equities for even very long periods of time in the future. For individual account proposals, analysis of expected benefit levels and money's worth was also provided using a higher average real annual equity-yield assumption of about 9.6 percent. This higher average yield reflected the arithmetic mean, rather than the geometric mean (which was 7 percent), of historical data for annual yields. It was suggested by Dr. Dickson that financial analysts generally use the arithmetic mean yield as a basis for illustrating likely expected yield on investments. It was observed that this approach was consistent with assuming that future annual yields would occur as if drawn at random, independently from the distribution of past annual yields.

Estimates for the Kerrey-Simpson proposal and for the Advisory Council proposals were based on the intermediate assumptions of the 1995 Trustees Report, including an assumption of an average annual future real yield of 2.3 percent for Treasury securities. Thus, an equity premium over long-term Treasury securities of 4.7 percentage points was implicitly assumed. It was noted that the historical average equity premium was higher, because the average real yield on Treasury securities was lower than 2.3 percent for the past.

Assumptions Since 1995

Since 1995, the OCACT has continued to use an assumption that average annual real yield on equities will be about 7 percent for investments made in future years. Because the Trustees have gradually increased their assumption for the average future real yield on Treasury securities from 2.3 to 3.0 percent, the implicit equity premium has been reduced from 4.7 to 4 percentage points. In addition, OCACT has continued to provide estimates using lower assumed equity yields for all proposals, in order to illustrate the uncertainty and sensitivity of these estimates.

While it has been recognized that the equity market has continued to be priced at levels above the historical average (as indicated by PE ratios) since 1995, future cycles have been assumed to continue as in the past, so that the average real yield on equity investments made in future years will vary but will still average at a level similar to the past. While an “overpriced” current market suggests that current equity investments may be expected to achieve lower than average real yield, investments made in future years, when the price of stocks may have dropped to a cyclical low, may be expected to achieve a higher than average real yield. Market trends for 2000 and 2001 suggest that the equity market is no longer as “overpriced” as it had been in late 1999, supporting the assumption that future market cycles and average PE ratios may indeed continue to mirror the past.

OCACT has recognized that future equity yields will depend on the future return to capital and many other factors, as it has in the past. Based on the Trustees assumptions in the 2001 Trustees Report, labor productivity is projected to continue to increase in the future at a rate similar to past average growth over long periods of time. This assumption implies that capital deepening (increasing ratio of capital to labor) in the U.S. economy will also continue to trend at about the same rate as in the past. This is believed to be consistent with the assumption that real equity returns and the return to capital will be similar in the future to those in the past. On this basis, OCACT believes that assumption of a future average real equity yield of about 7 percent is consistent with the Trustees assumptions.

Other Views

Some have suggested that slower growth in the U.S. labor force in the future may result in accelerated capital deepening based on an assumed continuation in the historical rate of growth in domestic capital investment, and thus a lower future return to capital (and lower equity yields) in the U.S. economy. Specifically, this would imply that capital investment would grow to levels higher than could be accommodated with current technology while maintaining the marginal product of capital at a maximum. While this may be plausible (if investors have nowhere else to invest and are willing to accept a lower return), it would also imply a higher rate of growth in labor productivity than in the past, and thus would be inconsistent with current Trustees assumptions.

A more compelling argument may be that the general investor may see equities as less risky in the future than in the past, or may be less averse to the level of risk that is present. This attitude would be consistent with a higher level of equity prices, higher PE ratios, lower dividend ratios (to price), and thus a lower real yield on equities (see Diamond 1999). However, OCACT believes that the perception in 1999 that equities will be consistently less risky in the future than in the past may already have been dispelled by price changes since 1999. In the future, OCACT believes that it is likely that stocks will be viewed as risky to about the same extent as in the past, over long periods of time.

Growth in the Total Value of the Equity Market

The assumption that future PE ratios will average at about the same level as in the past implies that the AGGREGATE price of all equities outstanding will grow at the same rate as for aggregate corporate earnings, and thus for GDP. This means that a slower future rate of growth in labor force and GDP (as projected by the Trustees) implies a slower future growth rate for aggregate stock value. In order to be consistent with a continuation of the past equity yield of 7 percent, this would imply that the dividend ratio will be higher in the future, offsetting the lower growth in corporate sales (GDP) and earnings, and thus share values. This would seem to be a reasonable consequence of slower labor force growth. Slower growth in employment from one year to the next means that the share of each year's corporate earnings that must be retained for investment in a growing workforce is reduced. These corporate earnings may reasonably be assumed to be distributed in the form of dividends, providing an equity yield that compensates for the slower increase in equity price.

An alternative assumption might be that corporate earnings that would be retained for a faster growing work force might be invested by the corporation abroad, thus effectively expanding labor and output offshore. This would result in increases in corporate output (although not in domestic GDP) and corporate earnings that would in turn support higher increases in equity prices, and thus total equity yield.

THE SOCIAL SECURITY ADVISORY BOARD

Establishment of the Board

In 1994, when the Congress passed legislation establishing the Social Security Administration as an independent agency, it also created a 7-member bipartisan Advisory Board to advise the President, the Congress, and the Commissioner of Social Security on matters relating to the Social Security and Supplemental Security Income (SSI) programs. The conference report on this legislation passed both Houses of Congress without opposition. President Clinton signed the Social Security Independence and Program Improvements Act of 1994 into law on August 15, 1994 (P.L. 103-296).

Advisory Board members are appointed to 6-year terms, made up as follows: 3 appointed by the President (no more than 2 from the same political party); and 2 each (no more than one from the same political party) by the Speaker of the House (in consultation with the Chairman and Ranking Minority Member of the Committee on Ways and Means) and by the President pro tempore of the Senate (in consultation with the Chairman and Ranking Minority Member of the Committee on Finance). Presidential appointees are subject to Senate confirmation. Board members serve staggered terms. There is currently one vacancy on the Board.

The Chairman of the Board is appointed by the President for a 4-year term, coincident with the term of the President, or until the designation of a successor.

Members of the Board

Stanford G. Ross, Chairman

Stanford Ross is a partner in the law firm of Arnold & Porter, Washington, D.C. He has dealt extensively with public policy issues while serving in the Treasury Department, on the White House domestic policy staff, as Commissioner of Social Security, and as Public Trustee of the Social Security and Medicare Trust Funds. He is a Founding Member and a former Director and President of the National Academy of Social Insurance. He has provided technical assistance on Social Security and tax issues under the auspices of the International Monetary Fund, World Bank, and U.S. Treasury Department to various foreign countries. He has taught at the law schools of Georgetown University, Harvard University, New York University, and the University of Virginia, and has been a Visiting Fellow at the Hoover Institution, Stanford University. He is the author of many papers on Social Security and Federal taxation subjects. Term of office: October 1997 to September 2002.

Jo Anne Barnhart

Jo Anne Barnhart is a political consultant and public policy consultant to State and local governments on welfare and social services program design, policy, implementation, evaluation, and legislation. From 1990 to 1993 she served as Assistant Secretary for Children and Families, Department of Health and Human Services, overseeing more than 65 programs, including Aid to Families with Dependent Children, the Job Opportunities and Basic Skills Training program, Child Support Enforcement, and various child care programs. Previously, she was Minority Staff Director for the U.S. Senate Committee on Governmental Affairs, and legislative assistant for domestic policy issues for Senator William V. Roth. Ms. Barnhart served as Political Director for the National Republican Senatorial Committee. First term of office: March 1997 to September 1998; current term of office: October 1998 to September 2004.

Martha Keys

Martha Keys served as a U.S. Representative in the 94th and 95th Congresses. She was a member of the House Ways and Means Committee and its Subcommittees on Health and Public Assistance and Unemployment Compensation. Ms. Keys also served on the Select Committee on Welfare Reform. She served in the executive branch as Special Advisor to the Secretary of Health, Education, and Welfare and as Assistant Secretary of Education. She was a member of the 1983 National Commission (Greenspan) on Social Security Reform. Martha Keys is currently consulting on public policy issues. She has held executive positions in the non-profit sector, lectured widely on public policy in universities, and served on the National Council on Aging and other Boards. Ms. Keys is the author of *Planning for Retirement: Everywoman's Legal Guide*. First term of office: November 1994 to September 1999; current term of office: October 1999 to September 2005.

David Podoff

David Podoff is visiting Associate Professor at the Department of Economics and Finance at the Baruch College of the City University of New York. Recently, he was Minority Staff Director and Chief Economist for the Senate Committee on Finance. Previously, he also served as the Committee's Minority Chief Health and Social Security Counselor and Chief Economist. In these positions on the Committee he was involved in major legislative debates with respect to the long-term solvency of Social Security, health care reform, the constitutional amendment to balance the budget, the debt ceiling, plans to balance the budget, and the accuracy of inflation measures and other government statistics. Prior to serving with the Finance Committee he was a Senior Economist with the Joint Economic Committee and directed various research units in the Social Security Administration's Office of Research and Statistics. He has taught economics at the University of Massachusetts and the University of California at Santa Barbara. He received his Ph.D. in economics from the Massachusetts Institute of Technology and a B.B.A. from the City University of New York. Term of office: October 2000 to September 2006.

Sylvester J. Schieber

Sylvester Schieber is Director of the Research and Information Center at Watson Wyatt Worldwide, where he specializes in analysis of public and private retirement policy issues and the development of special surveys and data files. From 1981 to 1983, Mr. Schieber was the Director of Research at the Employee Benefit Research Institute. Earlier, he worked for the Social Security Administration as an economic analyst and as Deputy Director at the Office of Policy Analysis. Mr. Schieber is the author of numerous journal articles, policy analysis papers, and several books including: *Retirement Income Opportunities in An Aging America: Coverage and Benefit Entitlement; Social Security: Perspectives on Preserving the System;* and *The Real Deal: The History and Future of Social Security*. He served on the 1994-1996 Advisory Council on Social Security. He received his Ph.D. from the University of Notre Dame. Term of office: January 1998 to September 2003.

Gerald M. Shea

Gerald M. Shea is currently assistant to the president for Government Affairs at the AFL-CIO. He previously held several positions within the AFL-CIO, serving as the director of the policy office with responsibility for health care and pensions, and also in various executive staff positions. Before joining the AFL-CIO, Mr. Shea spent 21 years with the Service Employees International Union as an organizer and local union official in Massachusetts and later on the national union's staff. He was a member of the 1994-1996 Advisory Council on Social Security. Mr. Shea serves as a public representative on the Joint Commission on the Accreditation of Health Care Organizations, is a founding Board member of the Foundation for Accountability, Chair of the RxHealth Value Project, and is on the Board of the Forum for Health Care Quality and Measurement. He is a graduate of Boston College. First term of office: January 1996 to September 1997; current term of office: October 2000 to September 2004.

Members of the Staff

Margaret S. Malone, Staff Director

Michael Brennan
Beverly Rollins
George Schuette
Wayne Sulfridge
Jean Von Ancken
David Warner

Social Security Advisory Board
400 Virginia Avenue, SW
Suite 625
Washington, D.C. 20024
Tel: (202) 475-7700
Fax: (202) 475-7715
www.ssab.gov

by Willard T. Carleton and Josef Lakonishok

Risk and Return on Equity: The Use and Misuse of Historical Estimates

The task of estimating a company's expected return typically involves an initial estimate of the market's expected return. This, in turn, is usually based on summary statistics about risk premiums drawn from historical average returns. The approach appears simple, but the underlying complexities may trip up unwary analysts.

The authors demonstrate how choice of measurement period, averaging method, portfolio weighting and risk-free rate can cause the equity risk premium to vary from 0.9 to 24.9 per cent. Over the 1926-80 period, for example, the arithmetic mean annual return on an equally weighted portfolio was 17.1 per cent; the geometric mean annual return on a corresponding value-weighted portfolio was 9.1 per cent. Furthermore, differences in historical returns between industries, and company size effects within industries, are also substantial.

FINANCIAL ANALYSTS HAVE come to rely heavily on summary statistics drawn from historical returns on common stocks.[1] Typically, these returns, aggregated over time and over securities, have been compared with historical returns on lower-risk assets such as Treasury bills or U.S. government bonds to provide estimates of the stock market's average risk premium on equities.[2] The considerable complexity underlying the aggregate data seems to have been ignored, for the most part, in practice.

The consequences of ignoring complexity can be substantial in dollar terms. For example, the book value of Duke Power Company's common equity is about $2.4 billion. Each percentage point in estimates of its cost of equity capital thus translates into $24 million of earnings per year, when applied as an earnings rate on book equity. And the differences between estimates of costs of equity generated by different "readings" of historical returns could easily amount to several percentage points—or multiples of $24 million per year—in required earnings.

This article attempts to introduce some caution into the uncritical acceptance and use of aggregated historical return differentials. Using return data for the period 1926–80, we present tables showing how mean or risk-adjusted stock returns are affected by the following dimensions of historical return measurement and presentation:

- geometric vs. arithmetic mean returns,
- equally weighted vs. value-weighted stock portfolios,
- time periods chosen,
- bills vs. bonds as the base for the market risk premium,
- industry risk-adjusted return differentials,
- effect of data point intervals on industry risk adjustments,
- the significance of some industry "alphas,"
- size effects within industries.

We used as our main data base the monthly

Willard Carleton is Karl Eller Professor of Finance at the College of Business and Public Administration of the University of Arizona. Josef Lakonishok is Associate Professor of Finance at The Leon Recanati Graduate School of Business Administration of Tel Aviv University.

1. Footnotes appear at end of article.

Table I Annualized Historical Returns and Standard Deviations on Market Portfolios

	Geometric Mean		Arithmetic Mean		Standard Deviation	
Period	*Val. Wtd.*	*Eq. Wtd.*	*Val. Wtd.*	*Eq. Wtd.*	*Val. Wtd.*	*Eq. Wtd.*
1926–80	9.1%	12.5%	11.4%	17.1%	21.9%	33.1%
1931–80	9.5	14.4	11.7	18.7	21.3	32.7
1936–80	10.2	13.4	11.8	16.6	18.7	26.8
1941–80	11.4	14.9	12.8	17.7	17.6	25.4
1946–80	10.6	12.2	12.0	14.7	17.7	23.8
1951–80	10.8	13.0	12.3	15.6	18.3	24.7
1956–80	8.9	11.9	10.3	14.7	18.0	25.4
1961–80	8.7	12.2	10.1	15.1	17.9	25.4
1966–80	7.2	11.2	8.9	14.6	19.3	28.2
1971–80	9.1	13.3	11.1	16.9	21.3	29.0
1976–80	15.9	26.3	16.7	27.1	15.2	15.0

CRSP tape, which contains monthly stock returns for all NYSE companies and for various monthly stock indexes. We used the Compustat tape, which provides summaries of financial statements of all major U.S. corporations, to construct firm size measures.[3] The monthly returns on Treasury bills and long-term government bonds constructed by Ibbotson and Sinquefield were also used.

Overall Equity Market Results

Assume that our analytical task is to forecast the expected rate of return (alternatively, the required rate of return) on a given stock. Most such forecasts involve estimation of the expected return on the market and the return on some "risk-free" asset (or, alternatively, the difference between the two as the market's risk premium) and the risk of the particular stock. We therefore start by estimating the expected return on the market as a whole, defining the market portfolio conventionally as a portfolio that includes only common stock.[4]

Table I presents data on annual historical returns and standard deviations for two widely used market portfolios—the value-weighted Fisher index and the equally weighted Fisher index.[5] The results are presented for various periods, all of which have 1980 as an ending date. We selected 1980 to reflect the point of view of an analyst today who is trying to decide how far back into historical data he must go to develop averages that validly represent current investors' beliefs about the future.

Computing Average Returns

The annual returns in Table I are aggregated across time based on both geometric mean and arithmetic mean computations. For example, the value-weighted geometric mean of 9.1 per cent for the 1926–80 period is derived in the following way:

$$[(1 + r_{1926})(1 + r_{1927}) \cdot \cdot \cdot (1 + r_{1980})]^{1/55} - 1,$$

where r denotes the annual rate of return. The comparable arithmetic mean of 11.4 per cent is derived as:

$$(r_{1926} + r_{1927} + \cdot \cdot \cdot \; r_{1980})/55.$$

The difference between the two means of 2.3 per cent is substantial and is directly related to the variability of the return series. The differences between the means would be more pronounced in the case of individual securities, because of their higher variability.

Which of the two means should be used? The truth is, each is appropriate under particular circumstances. The geometric mean measures changes in wealth over more than one period on a buy and hold (with dividends reinvested) strategy. If the average investor rebalanced his portfolio every period, the geometric mean would not be a correct representation of his portfolio's performance over time. The arithmetic mean would provide a better measure of typical performance over a single historical period (in the example, one year).

Portfolio Weights

The differences between returns on a value-weighted index, or portfolio, and those on an equally weighted index are even more striking than the differences between arithmetic and geometric means. For the 1926–80 period, the equally weighted market portfolio had an average mean return of 17.1 per cent versus 11.4 per cent for the value-weighted portfolio. The geometric means of the two portfolios are closer

Table II Annualized Historical Returns and Standard Deviations on Long-Term Government Bonds and Treasury Bills

	Bonds		*Bills*		*Standard Deviation*	
Period	*Geo. Mean*	*Arith. Mean*	*Geo. Mean*	*Arith. Mean*	*Bonds*	*Bills*
1926–80	3.0%	3.2%	2.8%	2.8%	5.7%	2.7%
1931–80	2.8	3.0	2.7	2.8	5.9	2.8
1936–80	2.6	2.7	3.0	3.0	5.6	2.8
1941–80	2.3	2.4	3.4	3.4	5.8	2.8
1946–80	2.0	2.2	3.8	3.9	6.0	2.7
1951–80	2.2	2.3	4.3	4.4	6.4	2.6
1956–80	2.3	2.5	4.9	4.9	6.8	2.5
1961–80	2.6	2.8	5.5	5.6	6.4	2.4
1966–80	2.6	2.9	6.3	6.4	7.3	2.2
1971–80	4.0	4.2	6.8	6.8	6.9	2.5
1976–80	1.9	2.1	7.8	7.8	8.3	2.9

(12.5 versus 9.1 per cent) because the equally weighted portfolio has a higher standard deviation than the value-weighted portfolio (33.1 vs. 21.9 per cent).[6]

Again, which index should be used? The value-weighted index obviously provides a better measure of stock market performance in general, hence of the experience of investors as a whole. The difference between AT&T and a small NYSE company cannot be ignored; investors have committed more funds to AT&T than they have to many smaller companies. Equally weighted indexes are very simple to construct and understand, but they probably make no more sense than an index constructed by weighting companies according to the length of their names. Nonetheless, equally weighted indexes may have their uses in determining expected rates of return for specific companies.

Equally weighted indexes give much more weight to smaller companies, and smaller companies are in general riskier than larger companies, so part of the average return difference between the two types of indexes can be explained by risk differences. However, only part of the small firm-large firm return difference can be explained by the conventional measures of risk, beta and unsystematic risk; for reasons still not fully understood, stocks of small companies have outperformed those of large companies on a risk-adjusted basis.[7] (Note that any use of historical return characteristics for forward-looking purposes requires a belief that history tends to repeat itself.) In determining expected rates of return, company size cannot therefore be ignored, and an equally weighted index may be appropriate for certain companies and for particular uses of expected market return estimates.[8] Clearly, investment strategies based on portfolios of small firms fall into this category.

Finally, Table I shows that, with the exception of the 1976–80 results, choice of starting year makes a difference of up to about 4 per cent per year in average equity return for each of the four portfolio measures. The 1976–80 period represents a special case noted by many analysts: During the later part of the decade, probably because of unanticipated changes in inflation and interest rates, average stock returns *and* their variability substantially exceeded their average long-term values.

Choice of Risk-Free Rates

To estimate the equity market's *expected* risk premium (or forward-looking average), one usually computes the *historical* average return on lower-risk securities such as Treasury bills or U.S. government bonds.[9] The difference between the equity and bill or bond historical average provides an estimate of the market risk premium.

The logic of this procedure is straightforward: Expected rates of return on bills, bonds and stocks vary over time, reflecting common underlying changes in interest rates. Over short periods of time, realized return differences between stocks and bills, or between stocks and bonds, will vary because of random and unanticipated repricing of assets. Over a sufficiently large number of observations (number of years), however, investors realize, on average, the return differential consistent with the greater risk of common stocks—i.e., an amount equal to the expected risk premium.

Table II provides historical returns on Treasury bills and long-term U.S. government bonds. For these fixed income securities, the differences between geometric and arithmetic

Table III Annualized Equity Premium Estimates

	Arithmetic Means				Geometric Means			
	– Bonds		– Bills		– Bonds		– Bills	
Period	Val Wtd	Eq Wtd	Val Wtd	Eq Wtd	Val Wtd	Eq Wtd	Val Wtd	Eq Wtd
1926–80	8.2%	13.9%	8.6%	14.3%	6.1%	9.3%	6.3%	9.7%
1931–80	8.7	15.7	8.9	15.9	6.7	11.4	6.8	11.7
1936–80	9.1	13.9	8.8	13.6	7.6	10.7	7.2	8.2
1941–80	10.4	15.2	9.4	14.2	9.1	10.4	8.0	8.0
1946–80	9.7	12.5	8.0	10.8	8.6	10.0	6.8	6.8
1951–80	9.9	13.3	7.8	11.2	8.6	10.7	6.5	6.5
1956–80	7.8	12.2	5.4	9.8	6.6	9.4	4.0	4.0
1961–80	7.3	12.3	4.5	9.5	6.1	9.4	3.2	3.2
1966–80	6.0	11.7	2.5	8.2	4.6	7.4	0.9	0.9
1971–80	6.9	12.7	4.3	10.1	5.1	9.1	2.3	2.3
1976–80	14.6	24.9	8.9	19.2	14.0	24.2	8.1	8.1

mean rates of return are very small, reflecting the small variability of the return series. For the total 1926–80 period, the arithmetic mean return on long-term government bonds is 3.2 per cent, versus 2.8 per cent for Treasury bills. For any period starting after 1936, however, Treasury bills show higher returns.

The superior performance of Treasury bills is especially striking in the more recent periods. From 1971 through 1980, for example, the average return on long-term government bonds was 4.2 per cent, versus 6.8 per cent for Treasury bills. The main contributor to this behavior was unexpected inflation, which led to higher than expected interest rates, hence lower bond prices. Unanticipated capital losses on bonds offset coupon income, producing lower realized returns.

Assuming that more history is better than less for purposes of estimating the market risk premium, there still remains the serious question of whether to base the premium on Treasury bills or on long-term government bonds. Again, the means will depend on the ends.

Advocates of the Capital Asset Pricing Model (CAPM) routinely employ the stock-bill average return differential. Aside from questions relating to the model's conceptual validity, the stock-bill spread is appropriate for uses involving short-term investment horizons. But the one-period CAPM is valid for multiperiod environments only under implausible and rigid assumptions. And expected market return estimates based on risk premium computations may be used to value expenditures for irreversible, long-term investments (nuclear power generating plants, for example); in these cases, the stock-bond return differential may provide a more appropriate measure of the average long-term risk premium.[10]

Table III presents annual risk premium estimates for equally weighted and value-weighted market portfolios based on Treasury bills and long-term government bonds. There are a number of choices and the differences between them are not trivial. Depending on the particular time period, method of weighting, method of averaging, and risk-free rate used, the market equity risk premium ranges from 0.9 to 24.9 per cent per year.[11]

Equity Returns and Risk Adjustments by Industry

Now that we have estimated the equity market portfolio's risk premium, we can make some adjustments for the difference in risk between our company and a typical company in the market portfolio. The CAPM relates return to risk as follows:

$$E(R_i) = R_f + [E(R_m) - R_f]\beta_i,$$

where:

$E(R_i)$ = the expected return on company i,
R_f = the risk-free rate,
$E(R_m)$ = the expected return on the market portfolio, and
β_i = the company's systematic risk, or beta.

The remaining task, under the CAPM, is to determine the company's beta. Our confidence in choice of any given historical data representation to estimate the market risk premium is at this point somewhat shaken, however. A natural step may be to examine the return experiences of similar firms, given that we are not sure about how to determine a market risk premium, hence expected return. In addition, even in the CAPM framework, it may be appropriate to look at groups of companies or industries, rather than at individual companies.

Thus, rather than concentrate on various issues critical in the case of individual securities (such as measurement error and coefficient instability), we will focus our analysis on the industry level. This will facilitate the presentation of results and enable us to demonstrate better the possible reason for differences in return experiences.[12]

We grouped the sample companies into 15 industries based on their two-digit Standard Industrial Classification codes. Table IV gives the number of companies in each industry. Table V provides for each industry annual geometric returns, arithmetic returns and standard deviations of returns for the 1926–80 period. Three beta coefficients, three intercept (alpha) coefficients, and three coefficients of determination (R-squares) are also presented. Table VI shows the same results for the 1971–80 period. These coefficients were estimated from the following regression:

$$R_{it} - R_{ft} = \alpha_i + \beta_i[R_{mt} - R_{ft}) + e_{it},$$

where R_{it}, R_{ft} and R_{mt} are the period t returns for industry i (each security received the same weight), the risk-free rate (Treasury bill returns), and the return on the market portfolio (equally weighted Fisher index), respectively. Thus the differences between the three sets of coefficients result from differences in the estimation intervals (monthly, quarterly or annual).[13]

Table IV Industry Classifications

Industry	*SIC Code*
1. Mining	10-14
2. Construction	15-17
3. Food	20-21
4. Textile	22-23
5. Paper	24-27
6. Chemicals	28
7. Petroleum	29
8. Rubber	30-31
9. Metals	32-34
10. Machinery	35-39
11. Transportation	40-49
12. Wholesale Trade	50-51
13. Retail Trade	52-59
14. Finance	60-67
15. Services	70-89

Beta and Estimation Intervals

For the 1971–80 period, 10 of the 15 industries exhibit differences in betas of at least 0.1. For the mining industry, the monthly beta is 0.83, the annual 0.63; for the petroleum industry, the quarterly beta is 0.50, the annual 0.73. Assuming an annual risk premium of about 8 per cent, a 0.1 difference in betas will create a 0.8 per cent difference in expected returns; not much in the abstract, perhaps, but one that translates into $1.9 million per year in earnings for Duke Power if beta is used to determine its return on book equity.

The coefficients of determination at the indus-

Table V Returns and Risk Measures by Industries, 1926–1980

Industry	*Geo. Mean*[a]	*Arith. Mean*[a]	*Stan. Dev.*[a]	*Beta (1)*[b]	*Beta (3)*[b]	*Beta (12)*[b]	*Alpha (1)*[a,b]	*Alpha (3)*[a,b]	*Alpha (12)*[a,b]	*R^2 (1)*[b]	*R^2 (3)*[b]	*R^2 (12)*[b]
Mining	16.1	21.7	38.7	1.02	1.10	1.03	3.54[c]	2.91[d]	4.10	0.87	0.92	0.78
Construction	7.2	20.1	62.0	1.43	1.72	1.53	−3.17	−6.09	−4.80	0.60	0.78	0.66
Food	11.9	15.0	27.6	0.75	0.71	0.80	1.33[d]	1.45[d]	0.83	0.92	0.94	0.92
Textile	10.6	16.8	38.7	1.04	1.13	1.11	−1.69[d]	−2.22[c]	−1.93	0.90	0.95	0.89
Paper	13.0	18.4	37.6	1.01	1.07	1.10	0.60	0.12	−0.12	0.92	0.96	0.93
Chemicals	12.7	16.1	28.6	0.86	0.82	0.83	1.33[d]	1.61[c]	1.55[c]	0.92	0.96	0.92
Petroleum	14.7	18.9	31.3	0.80	0.74	0.81	4.28[c]	4.35[c]	4.65[c]	0.71	0.82	0.73
Rubber	10.6	16.8	39.2	1.06	1.10	1.12	−1.94	−2.02[d]	−2.10	0.89	0.95	0.89
Metals	12.2	17.8	38.9	1.11	1.13	1.13	−0.72	−0.96	−1.30	0.96	0.98	0.93
Machinery	12.5	18.4	37.6	1.09	1.07	1.11	−0.24	0.04	−0.40	0.97	0.98	0.96
Transportation	10.4	14.5	29.9	0.99	0.95	0.81	−1.33	−0.68	0.37	0.89	0.91	0.80
Wholesale Trade	11.4	16.7	35.9	0.83	0.91	1.02	1.33	0.28	−0.82	0.69	0.84	0.89
Retail Trade	10.7	16.3	36.1	0.90	0.87	1.01	−0.60	−0.28	−1.03	0.88	0.91	0.86
Finance	11.4	15.8	30.1	0.99	0.94	0.85	−0.60	0.00	1.02	0.94	0.95	0.84
Services	13.0	19.9	40.6	1.04	1.03	1.09	0.84	1.45	1.47	0.86	0.91	0.79
Average	11.9	17.5	36.8	0.99	1.02	1.02	0.24	0.08	0.10	0.86	0.92	0.85

[a] Annualized percentages.
[b] The number in parentheses is the length of the estimation interval—monthly, quarterly or yearly
[c] Statistical significance of 5 per cent for a two-tailed test.
[d] Statistical significance of 10 per cent for a two-tailed test.

Table VI Returns and Risk Measures by Industry, 1971–1980

	Geo. Mean[a]	Arith Mean[a]	Stan. Dev.[a]	Beta (1)[b]	Beta (3)[b]	Beta (12)[b]	Alpha (1)[a,b]	Alpha (3)[a,b]	Alpha (12)[a,b]	R^2 (1)[b]	R^2 (3)[b]	R^2 (12)[b]
Mining	24.8	29.4	38.2	0.83	0.70	0.63	12.42[c]	13.43[c]	17.54	0.55	0.51	0.23
Construction	20.1	26.6	41.4	1.21	1.29	1.31	5.79[d]	6.01	6.65	0.86	0.88	0.83
Food	12.6	15.0	25.1	0.81	0.81	0.83	0.24	0.80	−0.15	0.92	0.92	0.91
Textile	7.6	14.3	41.9	1.13	1.17	1.34	−5.41[c]	−5.14[d]	−6.11	0.87	0.88	0.86
Paper	11.6	15.0	28.6	0.99	1.03	0.96	−1.33	−1.61	−1.64	0.94	0.96	0.95
Chemicals	13.7	15.4	20.0	0.81	0.77	0.66	1.33	1.29	1.94	0.86	0.91	0.91
Petroleum	20.7	24.4	31.5	0.69	0.50	0.73	9.25[d]	10.42[d]	10.16	0.49	0.40	0.45
Rubber	11.6	16.4	33.5	1.01	1.02	1.10	−1.45	−1.33	−1.53	0.88	0.89	0.90
Metals	14.8	17.3	25.0	1.01	0.94	0.83	1.33	1.89	2.02	0.94	0.95	0.93
Machinery	16.2	21.2	34.1	1.15	1.18	1.17	2.30	0.08	2.47[c]	0.96	0.96	0.99
Transportation	10.9	13.4	24.3	0.72	0.68	0.82	−0.84	−0.76	−1.83	0.87	0.87	0.97
Wholesale Trade	12.7	17.7	34.0	1.19	1.24	1.13	−1.09	−1.16	−0.50	0.94	0.94	0.92
Retail Trade	8.4	14.4	38.9	1.13	1.26	1.15	−4.91	−5.01[d]	−5.62	0.92	0.94	0.86
Finance	8.9	13.4	30.3	1.06	1.05	1.00	−4.41[d]	−4.06[d]	−3.46	0.89	0.92	0.91
Services	15.2	22.1	38.6	1.28	1.38	1.28	1.09	1.15	2.78	0.94	0.95	0.93
Average	14.0	18.4	32.4	1.00	1.00	1.00	0.84	0.96	1.52	0.86	0.86	0.84

[a] Annualized percentages.
[b] The number in parentheses is the length of the estimation interval—monthly, quarterly or yearly.
[c] Statistical significance of 5 per cent for a two-tailed test.
[d] Statistical significance of 10 per cent for a two-tailed test.

try level are extremely high. For the 1926–80 period, the averages across industry are 0.86, 0.92 and 0.85 for the monthly, quarterly and annual intervals, respectively. Although there is some indication of a better fit for quarterly data, the differences are not large enough to decide on the basis of statistical fit that quarterly data should be used to estimate betas.

We should note that the results in Tables V and VI probably underestimate the impact of estimation intervals on betas of individual companies. We used intervals of one month or longer. Betas estimated from daily or weekly data are subject to biases caused by trading patterns; there are no biases in estimated betas for NYSE securities when monthly data are used.[14] Furthermore, our betas are estimated at the level of industries, not individual securities; differences due to beta estimation intervals are partially suppressed when industry aggregates are employed.[15]

Estimation Intervals and Alpha

According to the CAPM, the theoretical intercept, or alpha, should be zero; estimated deviations from zero should be attributable to conventional estimation problems; and the intercept should be irrelevant in generating industry or company expected returns. Given that our beliefs in CAPM are somewhat shaken, however, the question is whether to retain or discard the intercept when expected returns are being generated.[16]

For the 1926–80 period and the monthly intercept, a two-tailed test shows two intercepts to be different from zero at the 5 per cent significance level and three at the 10 per cent level; 10 intercepts are not significantly different from zero. One approach to the development of an expected industry rate of return would be to discard the intercepts, especially the 10 that are not significantly different from zero, statistically. We feel that this procedure errs. What we want for an expected return estimate is an unbiased point estimate; if the regression equation were correctly specified, retaining estimated beta while discarding estimated alpha would obviously produce bias in estimated expected rate of return.[17]

Unfortunately, the size of the intercepts indicates that the effect on expected industry returns is substantial. For the rubber industry, for example, the monthly intercept is −1.94 per cent per year. Also, Table V indicates that differences in estimation intervals produce differences in intercepts. For the finance industry, the monthly intercept is −0.6 per cent, while the annual intercept is 1.02 per cent per year.

There is one other problem. A high (low) intercept may simply result from a series of unexpectedly favorable (unfavorable) circumstances in the past. For the 1971–80 period, the intercept of the oil industry was 9.25 per cent per year—but a 9.25 per cent intercept for the industry in the future is not a proposition most analysts would accept. The high intercept reflects the misspecification of the return-generating process being used; the intercept captures factors omitted by the model. Unfortunately, the market model regression cannot provide additional insight about the size and origin of such factors.

The intercept can have a substantial effect on expected returns. Table VII presents estimates of the expected return for the construction industry, under a CAPM framework. The returns—based on the results of Table VI, an assumed market risk premium of 8 per cent and a risk-free rate of 9 per cent—range from 18.68 to 26.13 per cent. At the level of individual securities, the effects will be even greater.

Table VII Expected Return Estimates for the Construction Industry

	Without Intercept	*With Intercept*
Monthly Data Interval	18.68%	24.47%
Quarterly Data Interval	19.32%	25.33%
Annual Data Interval	19.48%	26.13%

Industry Size and Risk Effects

Our examination of equally weighted and value-weighted portfolios suggested the existence of a company size effect on stock returns. Are the effects of size on historical return experience present within industries? The presence of size effects within industries would vastly complicate the estimation of company expected returns.

Tables VIII, IX and X describe in some detail the role of company size within industries. We analyzed the periods 1961–80, 1966–80, 1971–80 and 1976–80, but given the similarity of results, we present here only those for the whole period (Table VIII) and for the last 10 years (Table IX). We measured size by the market value of the common stock as of December 31, and estimated its effect by dividing the companies within the 13 given industries into four size groups, based on their size at the end of the previous year.[18]

Table VIII indicates an almost perfect relation between size and return. For *all* 13 industries, the smallest companies (designated size Group 1) had higher annual returns (on the basis of both arithmetic and geometric means) than the largest companies (size Group 4). Based on the summary in Table X, the difference between Groups 1 and 4 in arithmetic mean across industries for 1961–80 amounts to 11.1 per cent per year (22.3–11.2 per cent).

An almost perfect monotonic relation exists, not only between size and returns, but also between size and risk, as the betas and standard deviations in Tables IX and X indicate. From Table X, the average beta and standard deviation for the smallest companies are 1.14 and 36.7 per cent, respectively, for 1961–80; the corresponding numbers for the largest companies are 0.79 and 23.8 per cent.

Table VIII Returns and Risk Measures by Industries and Size, 1961–1980

Industry	*Size Group*	*Size*	*Geo. Mean*	*Arith. Mean*	*Stan. Dev.*	*Beta*	*Alpha*
Metals	1	29	16.9	20.3	28.9	1.17	0.31*
	2	66	12.4	15.2	25.2	1.04	0.02
	3	169	8.1	10.7	24.3	0.98	−0.28*
	4	822	7.2	8.8	19.0	0.86	−0.30*
Machinery	1	27	17.0	23.5	41.0	1.36	0.27*
	2	78	11.9	16.3	31.9	1.23	−0.08
	3	220	10.9	14.4	28.7	1.09	−0.11
	4	2356	9.1	11.9	24.6	0.88	−0.16**
Transportation	1	63	15.3	17.6	23.5	0.83	0.31*
	2	170	10.9	12.6	20.3	0.73	0.03
	3	396	8.1	9.6	18.1	0.66	−0.14
	4	1800	5.8	7.0	16.8	0.60	−0.28*
Trade	1	23	14.2	21.0	41.9	1.26	0.10
	2	62	12.4	18.0	36.9	1.16	−0.01
	3	157	10.2	14.9	33.8	1.02	−0.13
	4	1186	7.4	11.1	28.8	0.87	−0.28*
Finance	1	29	14.4	19.6	34.3	1.36	0.16
	2	88	14.2	18.9	33.9	1.06	0.18
	3	272	10.3	13.0	23.9	0.95	−0.09
	4	1362	10.3	12.0	19.7	0.78	−0.01
Services	1	36	16.6	22.9	38.9	1.33	0.31*
	2	74	12.0	18.1	37.7	1.28	−0.05
	3	141	12.0	17.0	32.9	1.21	−0.02
	4	381	7.9	14.8	40.9	1.14	−0.30*

(Table continued)

Table VIII continued

Industry	Size Group	Size	Geo. Mean	Arith. Mean	Stan. Dev.	Beta	Alpha
Mining	1	40	25.6	34.2	55.1	1.06	1.11*
	2	121	22.2	26.0	32.3	0.79	0.94*
	3	292	18.7	21.8	29.4	0.84	0.63*
	4	1341	16.6	19.5	26.7	0.77	0.49*
Food	1	29	16.6	19.9	29.3	0.92	0.40*
	2	101	13.9	17.0	27.2	0.90	0.19*
	3	363	9.4	12.0	25.0	0.81	−0.11
	4	1428	8.8	10.3	18.2	0.62	−0.07
Textile	1	18	13.1	20.8	45.4	1.22	0.07
	2	43	11.0	16.2	36.1	1.13	−0.08
	3	87	9.1	15.0	36.8	1.01	−0.18**
	4	265	7.9	13.0	33.2	0.96	−0.26*
Paper	1	34	17.4	22.4	38.4	1.18	0.36*
	2	91	11.0	14.4	27.5	1.02	−0.07
	3	300	10.6	13.1	24.2	0.94	−0.06
	4	1344	6.7	8.6	21.0	0.83	−0.32*
Chemicals	1	50	16.4	19.8	28.8	1.11	0.30*
	2	184	11.7	13.8	21.6	0.94	0.01
	3	565	12.3	13.8	18.6	0.80	0.12
	4	2537	6.3	7.2	14.2	0.61	−0.23*
Petroleum	1	134	19.6	24.4	34.5	0.94	0.67*
	2	906	20.4	23.3	26.2	0.72	0.81*
	3	2763	15.2	17.7	25.0	0.55	0.55**
	4	8369	13.5	15.6	22.9	0.50	0.43**
Rubber	1	25	19.1	24.4	37.1	1.12	0.54*
	2	57	9.0	12.9	27.9	1.06	−0.20**
	3	212	10.3	14.5	32.9	0.93	−0.07
	4	847	2.5	5.2	23.5	0.85	−0.63*

* Statistical significance of 5 per cent for a two-tailed test.
** Statistical significance of 10 per cent for a two-tailed test.

Table IX Returns and Risk Measures by Industries and Size, 1971–1980

Industry	Size Group	Size	Geo. Mean	Arith. Mean	Stan. Dev.	Beta	Alpha
Metals	1	27	18.6	21.2	27.2	1.22	0.35*
	2	64	17.1	19.4	24.2	1.00	0.30*
	3	162	10.5	13.6	26.7	0.96	−0.18
	4	730	9.8	11.6	21.1	0.83	−0.17
Machinery	1	24	20.8	27.1	40.0	1.40	0.47*
	2	77	16.4	21.4	34.4	1.22	0.18
	3	229	13.6	18.3	33.2	1.06	0.02
	4	2517	9.9	13.3	27.6	0.83	−0.16
Transportation	1	61	14.9	18.1	28.2	0.85	0.19
	2	163	12.0	14.7	25.9	0.72	0.03
	3	387	8.3	10.4	22.7	0.66	−0.22
	4	1660	6.1	8.0	20.7	0.57	−0.34**
Trade	1	22	12.2	19.5	43.2	1.35	−0.14
	2	63	12.3	18.7	40.9	1.25	−0.13
	3	167	9.1	14.9	38.8	1.04	−0.31
	4	1171	4.0	8.8	34.1	0.90	−0.64*
Finance	1	31	15.1	20.8	35.0	1.54	0.09
	2	91	10.3	15.5	33.2	1.06	−0.22
	3	299	8.3	12.2	28.6	0.94	−0.32**
	4	1352	9.3	11.5	22.0	0.74	−0.16
Services	1	27	17.1	24.5	40.8	1.35	0.25
	2	64	12.3	20.1	40.4	1.40	−0.13
	3	148	13.7	20.1	36.6	1.21	0.03
	4	502	11.0	18.5	41.2	1.13	−0.16
Mining	1	50	27.9	36.2	57.9	1.03	1.26*
	2	149	26.3	31.0	37.9	0.82	1.16*
	3	396	24.0	28.0	35.4	0.80	0.99*
	4	2039	18.2	21.9	30.8	0.69	0.58

(Table continued)

Table IX continued

Industry	Size Group	Size	Geo. Mean	Arith. Mean	Stan. Dev.	Beta	Alpha
Food	1	29	18.9	22.1	30.2	0.94	0.46*
	2	118	17.6	20.2	27.1	0.90	0.37*
	3	436	7.9	11.2	29.3	0.79	-0.30*
	4	1733	8.4	10.1	19.9	0.60	-0.17
Textile	1	17	11.5	20.9	52.0	1.30	-0.12
	2	40	4.5	9.9	38.5	1.10	-0.64*
	3	83	2.1	7.9	37.3	0.98	-0.80*
	4	276	4.5	10.8	37.2	0.97	-0.61*
Paper	1	34	15.2	18.9	30.3	1.21	0.12
	2	97	10.5	15.4	32.9	0.99	-0.18
	3	326	12.4	15.5	28.8	0.89	0.00
	4	1500	6.9	9.6	25.4	0.79	-0.36*
Chemicals	1	50	18.7	22.2	30.2	1.08	0.40*
	2	211	13.0	15.3	23.0	0.87	0.05
	3	682	13.8	15.7	21.0	0.73	0.18
	4	2969	5.9	7.0	15.9	0.56	-0.30
Petroleum	1	158	22.0	29.1	42.0	0.95	0.77**
	2	1134	20.4	24.5	32.0	0.73	0.75**
	3	3526	22.5	25.5	29.5	0.47	1.07*
	4	9044	16.2	19.2	28.3	0.49	0.57
Rubber	1	23	22.9	30.6	46.7	1.18	0.74*
	2	52	9.9	14.7	30.4	1.05	-0.20
	3	210	10.8	15.7	37.3	0.94	-0.12
	4	739	-0.6	3.2	28.9	0.85	-0.98*

* Statistical significance of 5 per cent for a two-tailed test.
** Statistical significance of 10 per cent for a two-tailed test.

Table X Returns and Risk Measures Averaged Across Industries, by Size Groups

Period	Size	Geo. Mean	Arith. Mean	Stan. Dev.	Beta	Alpha
1961–80	41	17.1	22.3	36.7	1.14	0.38*
	157	13.3	17.1	29.6	1.01	0.13
	457	11.1	14.4	27.2	0.91	0.01
	1849	8.5	11.2	23.8	0.79	-0.15**
1971–80	43	18.1	23.9	38.8	1.18	0.37*
	179	14.1	18.5	32.3	1.01	0.10
	542	12.1	16.1	31.1	0.88	0.00
	2019	8.4	11.8	27.1	0.77	-0.22*

* Statistical significance of 5 per cent by two-tailed test.
** Statistical significance of 10 per cent by two-tailed test.

Does Alpha Depend on Size?

Did small companies outperform large companies on a risk-adjusted basis? The last column in each table presents the industry alphas, which should theoretically equal zero. Higher intercepts for the smaller companies would suggest superior performance on a risk-adjusted basis. For both 1961–80 and 1971–80 periods, the smallest companies in all 13 industries outperformed the largest. The 1961–80 difference in intercepts between the smallest and the largest group sizes, summarized over all industries in Table X, is 0.53 per cent per month, which translates to 6.55 per cent per year (statistically significant at the 5 per cent level). For 1971–80, the difference is 7.31 per cent per year (also significant at the 5 per cent level).

Our results regarding the effect of size on industry returns are consistent with results of previous studies that did not examine differential returns within industries.[19] As noted, the presence of intraindustry size effects vastly complicates estimation of expected returns for individual companies. Whether the purpose is capital budgeting, rate of return regulation, or investment strategy, the analyst has to decide to include or ignore the size effect. We have no theory that adequately explains the phenomenon, so it is tempting to assume that it will not persist in the future. But discarding it is to deny historical reality and, in the framework of CAPM-based market model regressions, to produce biased return estimates.

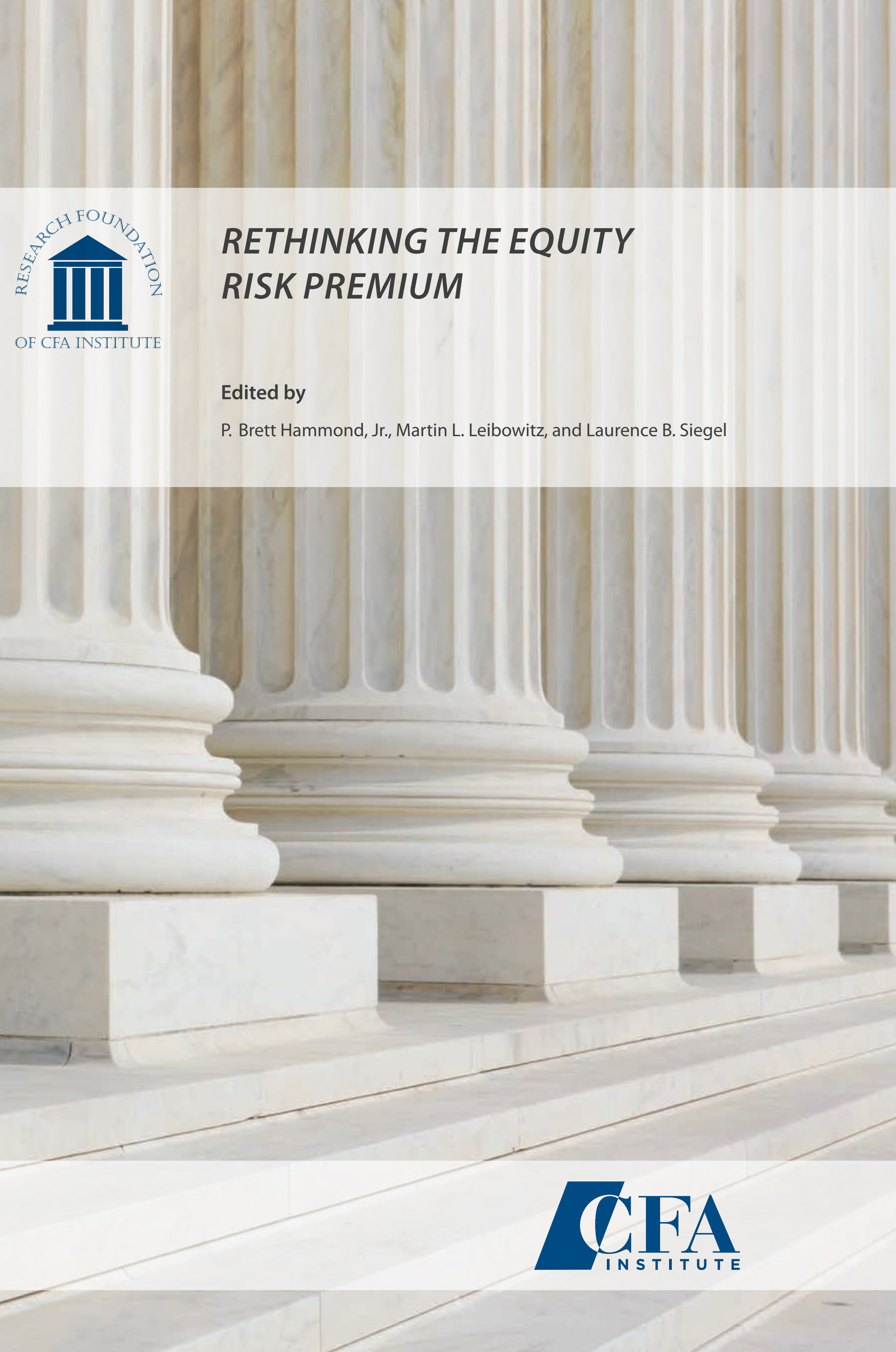
Implications for Analysts

The practical applications of expected return estimates entail serious financial consequences (especially in the case of utility regulation). Given our incomplete understanding of how stock returns are determined, we think it is delusionary and misleading not to acknowledge the complexities just under the surface of simple historical average returns. On empirical grounds, if no other, it would appear that the popular recipe of, say, 8 per cent times company beta, added to a bill yield, may not be robust enough for general use.

Footnotes

1. For among other tasks, development of capital budgeting discount rates; estimation of equilibrium stock prices in order to measure deviations against which speculative trading can take place; and estimation of costs of equity capital for utilities, to be employed in rate hearings.
2. See, for example, R.G. Ibbotson and R.A. Sinquefield, *Stocks, Bonds, Bills, and Inflation: The Past (1926–1976) and the Future (1977–2000)* (Charlottesville, Va.: The Financial Analysts Research Foundation, 1977); *Stocks, Bonds, Bills, and Inflation: Historical Returns (1926–1978)* (Charlottesville, Va.: The Financial Analysts Research Foundation, 1979); and *Stocks, Bonds, Bills and Inflation: The Past and the Future* (Charlottesville, Va.: The Financial Analysts Research Foundation, 1982).
3. The Compustat tape provides data only for companies that exist currently. For example, the 1980 Compustat tape provides data only for companies that existed in 1980. The Research Compustat tape was used to provide data on companies that went out of existence.
4. For purposes of this article, we will not deal with the well known problems associated with the validity of a portfolio that excludes such important assets as bonds and real estate. For a comprehensive discussion of these issues see R.R. Roll, "A Critique of the Asset Pricing Theory's Tests, Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, March 1977, pp. 129–176.
5. For a complete description of the Fisher Index, see Lawrence Fisher and James Lorie, "Rates of Return on Investments in Common Stocks: The Year-by-Year Record, 1926–65," *Journal of Business*, July 1968, pp. 291-316. These indexes are available on the CRSP tapes and are adjusted for all changes in capitalization.
6. The difference between the equally weighted and value-weighted indexes would be even larger if AMEX and OTC companies had been included.
7. For a discussion of these issues, see Richard Roll, "A Possible Explanation of the Small Firm Effect," *Journal of Finance*, September 1981, pp. 879–888.
8. There is a further complication we do not pursue in this article, which arises in the context of estimation of expected rates of return for an average investor on an after-tax basis. Everything else constant, companies with high variability in returns provide investors with a higher tax subsidy. This subsidy is related to the distinction made by the IRS between long-term and short-term capital gains. These issues are discussed by George Constantinides, "Optimal Stock Trading with Personal Taxes: Implications for Prices and the Abnormal January Returns" (July 1982).
9. Note the greater returns of equities (Table I) over bonds (Table II) and bonds over bills (Table II), historically consistent with conventional descriptions of their relative risks.
10. For a discussion, see W.T. Carleton, "A Highly Personal Note on the Use of the CAPM in Public Utility Rate Cases," *Financial Management*, Autumn 1978, pp. 57–59, and W.T. Carleton, D.R. Chambers and J. Lakonishok, "Inflation Risk and Regulatory Lag," *Journal of Finance*, May 1983, pp. 419–436.
11. A further complication in the search for a market risk premium is that the variance of the market realized return series changes over time. We do not pursue this topic, as this article is addressed to the fairly typical user of historical returns observed in practice. For an exploration of the issues, see R.C. Merton, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, December 1980, pp. 323–361.
12. It should be pointed out at this stage that a popular alternative to the CAPM for deriving expected returns is based on observing the past performance of similar companies—companies from the same industry.
13. All the computations were repeated for the various time intervals discussed in Table I. Because the results were qualitatively similar we present only the findings for the total period, 1926–80, and the last 10 years, 1971–80.
14. The biases arise from trading patterns and are discussed by E. Dimson, "Risk Measurement When Shares are Subject to Infrequent Trading," *Journal of Financial Economics*, June 1979, pp. 197–226 and M. Scholes and J. Williams, "Estimating Betas from Non-Synchronous Data," *Journal of Financial Economics*, December 1977, pp. 309–327. H. Stoll and R. Whaley ("Transactions Costs and

(continued on page 62)



RESEARCH FOUNDATION OF CFA INSTITUTE

RETHINKING THE EQUITY RISK PREMIUM

Edited by

P. Brett Hammond, Jr., Martin L. Leibowitz, and Laurence B. Siegel

CFA INSTITUTE

Named Endowments

The Research Foundation of CFA Institute acknowledges with sincere gratitude the generous contributions of the Named Endowment participants listed below.

Gifts of at least US$100,000 qualify donors for membership in the Named Endowment category, which recognizes in perpetuity the commitment toward unbiased, practitioner-oriented, relevant research that these firms and individuals have expressed through their generous support of the Research Foundation of CFA Institute.

Ameritech
Anonymous
Robert D. Arnott
Theodore R. Aronson, CFA
Asahi Mutual Life
Batterymarch Financial Management
Boston Company
Boston Partners Asset Management, L.P.
Gary P. Brinson, CFA
Brinson Partners, Inc.
Capital Group International, Inc.
Concord Capital Management
Dai-Ichi Life Company
Daiwa Securities
Mr. and Mrs. Jeffrey J. Diermeier
Gifford Fong Associates
John A. Gunn, CFA
Jon L. Hagler Foundation
Investment Counsel Association of America, Inc.
Jacobs Levy Equity Management
Long-Term Credit Bank of Japan, Ltd.
Lynch, Jones & Ryan
Meiji Mutual Life Insurance Company
Miller Anderson & Sherrerd, LLP
John B. Neff, CFA
Nikko Securities Co., Ltd.
Nippon Life Insurance Company of Japan
Nomura Securities Co., Ltd.
Payden & Rygel
Provident National Bank
Frank K. Reilly, CFA
Salomon Brothers
Sassoon Holdings Pte. Ltd.
Scudder Stevens & Clark
Security Analysts Association of Japan
Shaw Data Securities, Inc.
Sit Investment Associates, Inc.
Standish, Ayer & Wood, Inc.
State Farm Insurance Companies
Sumitomo Life America, Inc.
T. Rowe Price Associates, Inc.
Templeton Investment Counsel Inc.
Travelers Insurance Co.
USF&G Companies
Yamaichi Securities Co., Ltd.

Senior Research Fellows

Financial Services Analyst Association

For more on upcoming Research Foundation publications and webcasts, please visit www.cfainstitute.org/about/foundation/.

Research Foundation monographs are online at www.cfapubs.org.

Edited by
P. Brett Hammond, Jr.
TIAA-CREF

Martin L. Leibowitz
Morgan Stanley

Laurence B. Siegel
Research Foundation of CFA Institute
Ounavarra Capital LLC

Rethinking the Equity Risk Premium

RESEARCH FOUNDATION OF CFA INSTITUTE

Statement of Purpose

The Research Foundation of CFA Institute is a not-for-profit organization established to promote the development and dissemination of relevant research for investment practitioners worldwide.

Neither the Research Foundation, CFA Institute, nor the publication's editorial staff is responsible for facts and opinions presented in this publication. This publication reflects the views of the author(s) and does not represent the official views of the Research Foundation or CFA Institute.

The Research Foundation of CFA Institute and the Research Foundation logo are trademarks owned by The Research Foundation of CFA Institute. CFA®, Chartered Financial Analyst®, AIMR-PPS®, and GIPS® are just a few of the trademarks owned by CFA Institute. To view a list of CFA Institute trademarks and the Guide for the Use of CFA Institute Marks, please visit our website at www.cfainstitute.org.

©2011 The Research Foundation of CFA Institute

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright holder.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

ISBN 978-1-934667-44-6

23 December 2011

Editorial Staff

Maryann Dupes
Book Editor

Mary-Kate Hines
Assistant Editor

Christina Hampton
Publishing Technology Specialist

Cindy Maisannes
Manager, Publications Production

Lois Carrier
Production Specialist

Contents

CFA Institute
CE Qualified Activity

This publication qualifies for 5 CE credits under the guidelines of the CFA Institute Continuing Education Program.

Rethinking the Equity Risk Premium: An Overview and Some New Ideas

P. Brett Hammond, Jr.
Managing Director and Chief Investment Strategist
TIAA-CREF

Martin L. Leibowitz
Managing Director, Research
Morgan Stanley

Many investors regard the past decade as an unusual one for market returns. This view is no doubt based on their having experienced a sea change in equity market behavior, including much-lower-than-average returns, much higher volatility, two of the biggest bubbles (and their subsequent bursting) in stock market history, and rising correlations—cross-asset, cross-country, cross-sector, and intra-sector. Any longtime investment market participant will have encountered more extreme trends and events in the past 10 years than during any other 10-year period in the past seven decades.

One of the key features of this turbulent period is renewed uncertainty about what may be the most important measure in all of finance—namely, the equity risk premium, or the expected return for equities in excess of a risk-free rate:

$$ERP = E(re) - E(rf).$$

The equity risk premium, or ERP, plays a critical role for any investor in that it affects savings and spending behavior as well as the all-important allocation decision between riskless and risky assets. In that sense, it is an equilibrium concept that looks beyond any given period's specific circumstances to develop a fundamental, long-term estimate of return trends.

It should be noted that the equity risk premium, as the term is used here, is not identical to the historical excess return. For example, for the 10 years beginning in the middle of 2001, annualized geometric mean U.S. equity returns significantly trailed U.S. TIPS (Treasury Inflation-Protected Securities)—roughly 3 percent versus 6 percent. So, one measure of the historical excess return is –3 percent.[1] In this volume, Robert Arnott shows that, using rolling 20-year returns, the historical excess return has ranged from +20 percent to –10 percent,

[1]Please note that, by convention, the return is often expressed as a "percentage" rather than "percentage points."

©2011 The Research Foundation of CFA Institute

a range that is not very helpful in forming a historical average. But these numbers do not say much about the equity risk premium, which is a forward-looking expectations-driven estimate of stock returns. In other words, what premium do we *expect* stocks to provide over a risk-free rate? This forward-looking premium is critical to fundamental activities in investing, especially strategic and tactical asset allocation but also in portfolio management, hedging, investment product development, and the formation of saving and spending plans.

The problem posed by recent history for all these activities is whether we can be confident in our understanding of equity risk. After several decades during which realized equity returns followed a welcome positive pattern, the past decade has seen a marked downturn in equities. This downturn has prompted some investors to suggest that we must permanently adjust our future expectations for equity returns versus other broad asset classes. Others argue that the same evidence suggests equities are poised for outstanding future excess returns. Which is it?

To investigate the ERP in more depth, we could evaluate forecasts, trends, and expected variations in forward-looking measures: P/Es, dividend payouts, debt, macroeconomic growth and inflation, investment horizon, demographic change, and other variables. We have at our disposal, arguably, more analytical techniques and sources of information than ever before that bear on asset class expectations and behavior, but we have less certainty than ever about the ERP.

This volume is the result of an effort to sort through and present some of the best recent thinking on the ERP in a way that practitioners may find useful in developing their own approach to the subject. It assembles leading practitioners and academics who have confronted the question of what the ERP might be going forward and, more importantly, what factors are the most important drivers of the premium.

Initial ERP Project

The present project arose out of an interest on the part of the Research Foundation of CFA Institute to revisit, in light of what has happened in asset markets, a similar but not identical effort that it sponsored in late 2001. This earlier effort emerged as the "dot-com" bubble burst and investors confronted, for the first time in many years, the possibility of an extended period of lower equity returns. The 2001 forum gathered a wide range of experts to discuss the theoretical foundations of the ERP, historical results, then-current estimates of the size of the premium, and implications for asset management (Association for Investment Management and Research 2002). It featured lively discussions of the definition of the ERP, rational expectations versus behavioral explanations for its existence, specific factors and models that explain its size and

 ©2011 The Research Foundation of CFA Institute

stability (or lack thereof), the possibility of structural change–driven effects on the premium, and ways in which institutions and individuals incorporate views on the ERP into asset allocation.

Rather than a firm consensus, a strong sense of diversity arose from this earlier forum regarding views on the ERP and possible explanations for differences among those views. For example, **Exhibit 1** shows, as of 2001, a selected set of estimates of the ERP ranging from 0 to 7 percent, with an average of a little less than 4 percent.

Exhibit 1. Estimates as of 2001 of the ERP

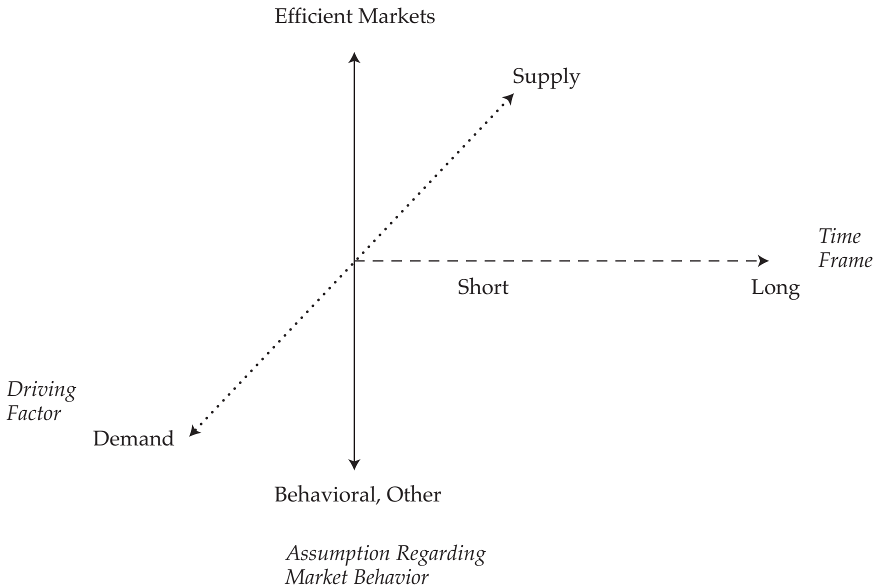
Source	ERP Estimate (%)
Arnott and Bernstein (2002)	0.0
Campbell and Shiller (2001)	0.0
McGrattan and Prescott (2001)	0.0
Ross, Goetzmann, and Brown (1995)	Low
Reichenstein (2001)	1.3
Campbell (2001)	1.5–2.5
Philips (2003)	1.0–3.0
Siegel (2002)	2.0
Bansal and Lundblad (2002)	2.5
Shoven (2001)	3.0
Siegel (1994)	3.0–4.0
Asness (2000)	4.0
Graham and Harvey (2001)	4.0
Ibbotson and Chen (2003)	4.0
Goyal and Welch (2002)	3–5
Fama and French (2002)	4.3
Cornell (1999)	5.0
Ibbotson and Sinquefield (1976)	5.0
Welch (2000)	6.0–7.0
Average	3.7
Range	0.0–7.0

Note: ERP estimates are the expected long-term geometric return of equities in excess of the real risk-free rate.

Figure 1 summarizes, in schematic form, some of the key dimensions that can help explain these estimates. On one dimension, differences in ERP estimates can be caused by the weight given to short-term versus long-term investment horizons, including an emphasis on mean reversion or cyclicality. (A related dimension, not shown here, for different regimes or macro environments could

©2011 The Research Foundation of CFA Institute

Figure 1. Three-Dimensional Array of Views on the ERP



also be added—for example, whether prevailing interest rates are high or low.) ERP estimates can also vary according to whether supply or demand considerations are the dominant influence. Some investigators focus on the demand for a return that will compensate investors for the extra risk of equities, whereas others look at the supply of cash flows that companies can inject into the market.

Perhaps most fundamentally, the forum exposed different views on investor behavior, specifically whether markets exhibit rational expectations or suffer from behavioral distortions, such as myopic loss aversion (which can be nonlinear or noncontinuous). One area of general agreement was that, to their detriment, few institutions or individuals explicitly address these issues and even fail to consider the size of the equity premium itself in forming policy portfolios and determining asset allocation.

10th Anniversary Project

The current project started with leading academics and practitioners gathering for a daylong discussion on what new developments, if any, have occurred in thinking about the ERP as well as in estimating the size of the ERP that we can expect in the future. Following that discussion, participants were asked to set down their current thoughts in essay form. The result, contained in this volume, is a rich set of papers that illuminate the issues and speak to the

 ©2011 The Research Foundation of CFA Institute

conceptual and empirical sources of the various perspectives. What is interesting about the more recent effort is not only some commonality with respect to the emphasis on supply-driven considerations but also—quite naturally in light of recent history and theory—a great deal of variation among the authors on the stability and term structure of the ERP as well as on whether variations in the ERP, no matter what their source, matter much.

The opening paper by Roger Ibbotson lays out several ways of estimating the ERP, including supply, demand, historical extrapolation, and combinations thereof. Investors are not the only agents who are affected by the excess return on equities over bonds; corporations should consider the ERP as the most important ingredient in understanding their cost of capital, and equity analysts need to use the ERP as part of the discount rate when estimating the present value of a company's future cash flows. Moreover, although it may be the largest market premium, the ERP is not the only one. Other premiums are associated with investment horizon, company size, value, momentum, default risk, and inflation risk. Of particular interest is the liquidity premium, described by Ibbotson as the phenomenon in which unpopular stocks (those that do not trade much) can display significant excess returns compared with stocks traded more often. Most important, investors often fail to differentiate a short-term tactical view of the ERP from the more fundamental long-term supply-driven equilibrium equity premium, suggesting that short-term signals may not always provide accurate information about the "true" long-term ERP.

Focusing on the cyclical nature of returns and fundamental indicators, Clifford Asness notes that there is no evidence that high P/Es are an accurate forecast of high future earnings growth rates. Rather, the evidence runs in the opposite direction. Using his own estimates of earnings growth and drawing on the Shiller P/E, which is the current price divided by trailing 10-year average real earnings, Asness offers a future equity return estimate in the range of 4 percent. Because it is hard to agree on a benchmark for the risk-free rate, he does not make a specific forecast of the ERP.

Looking historically and adopting a broad geographical perspective, Elroy Dimson, Paul Marsh, and Mike Staunton report on their most recent update of realized excess equity returns, relative to both bills and bonds, in 19 different countries from 1900 to the start of 2011. Although they found considerable variation across countries, the realized excess return was substantial everywhere. For their world index, annualized geometric mean real returns were 5.5 percent, the excess return relative to Treasury bills was 4.5 percent, and the excess return relative to long-term government bonds was 3.8 percent. Based on a supply model of the ERP, with the addition of the change in the real exchange rate, they estimate that the forward-looking equity premium is lower, around 3–3.5 percent, largely because of lower expected dividend growth compared with the historical average. In addition, they suggest that mean reversion in the stock market may not be as strong a force as others would argue. And even if mean reversion is a force, it may not provide much comfort to an investor who still does not know what the average stock market return will be in the future, nor what the equity premium is today or what the other parameters of the return process are.

The paper by Richard Grinold, Kenneth Kroner, and Laurence Siegel develops and estimates a supply model of the ERP. It decomposes equity returns into three major components: income, earnings growth, and repricing:

$$R \underbrace{\frac{D}{P} - \Delta S}_{\text{Income}} + \underbrace{i + g}_{\text{Earnings growth}} + \underbrace{\Delta \text{PE}}_{\text{Repricing}},$$

where D/P is the dividend yield, ΔS is share repurchases net of (that is, minus) new issuance, i is inflation, g is real earnings growth (not earnings per share), and the last term is the change in the P/E multiple. To illustrate, if the current 10-year bond yield is 2 percent and the ERP is 4 percent, then income, earnings growth, and repricing components must sum to 6 percent. Looking forward, the authors estimate future income to be about 2 percent, composed of dividend yield of about 1.8 percent and net share repurchases at 0.2 percent (repurchases of 2.2 percent and dilution or new issues at 2 percent). Earnings growth is expected to be a little more than 5 percent, with 2.4 percent coming from inflation and a little less than 3 percent coming from real earnings growth (which they equate to real GDP growth). Finally, although repricing contributed significantly to equity returns in the 20th century, there is little reason to believe that it will continue to do so. If we put these figures together, equity returns are expected to be about 7.2 percent. If the long-term nominal bond yield is about 3 percent, then the ERP is in the range of 4 percent.

Robert Arnott supports a view of the ERP as cyclical, smaller, and more dynamic than the prevailing theory of a more stable and robust premium would suggest. He counters a series of "myths" by showing that bonds have outperformed stocks over a significant period, the realized excess return has often been lower than the forward-looking ERP, net stock buybacks are lower than is often assumed, lower earnings yields are empirically associated with lower subsequent stock returns and premiums, real earnings and stock prices grow with per capita GDP rather than total GDP, and dividend yields are lower now than ever before. When taking this more sobering evidence into account, he finds that the probability of future stock returns matching the 7 percent real historical average is slight. Arnott's estimate of the future ERP ranges from negative to slightly positive.

 ©2011 The Research Foundation of CFA Institute

Antti Ilmanen directly addresses the issue of the stability of the ERP over time by considering what the premium might look like for the next decade and well beyond, including periods with regime and term structure variations. After helpfully reviewing a wide variety of approaches to the ERP, he makes three major points. First, term structure effects are more obvious on the bond side of the premium, where short-dated TIPS yields are currently negative but longer-dated TIPS are higher, implying a 2.7 percent forward TIPS yield for the decade starting in 2021. Second, abnormally high (or low) starting valuations for equity markets and related mean-reversion potential have strong implications for expected stock market returns for the next few years. However, if we consider prospective equity returns *after* the next decade, we have no clue what the starting valuation levels will be in 2021. Thus, if we assume below-average equity market returns for the next decade because of an expected normalization of the currently high Shiller P/E, our best forecast for real equity market returns beyond 2021 should be closer to our "unconditional" long-term return forecasts. That is, these forward forecasts should largely ignore starting valuations (or at least allow future higher starting yields in 2021 than in 2011). And third, many indicators besides valuation measures can be used to predict stock market returns. Regressions and other econometric techniques can be used to forecast returns over any investment horizon (admittedly having fewer independent data points in longer horizon regressions). It is thus possible to estimate a full term structure of expected returns.

Using a variation on the supply-driven approach, Peng Chen looks at whether bonds might outperform stocks over the long run as they have over the past decade. Although the bulk of bond returns comes from their yield or income, the recent outperformance of bonds is based on the decline in yield (price increase). Currently, long-term bond yields are so low (estimated at the time of writing to be less than 3 percent) that they are unlikely to decline much further, so expected capital gains from bonds are low to negative. In contrast, stock returns depend on earnings growth and the change in the ratio of price to earnings as well as their yield. If expected earnings growth and yields remain at roughly historical averages (5 percent and 2 percent, respectively), then P/Es have to decline to 5 to produce overall future stock returns less than the 4 percent expected bond yield—an outcome that seems highly unlikely.

Looking at the information contained in the P/E that might bear on the ERP, Andrew Ang and Xiaoyan Zhang conclude that the ERP is relatively stable over time. They decompose companies' future earnings into those associated with a perpetual, no-growth component and a component associated with future growth opportunities. In effect, movements in P/E reflect changes in discount rates, which contain the ERP, as well as growth opportunities, which involve the cash flow and earnings-generating capacity of company

©2011 The Research Foundation of CFA Institute

investments. Therefore, P/Es can be high (low) because growth opportunities are favorable and/or because expected returns are low. Using more than 50 years of data from the S&P 500 Index, Ang and Zhang show that macro variables—especially risk-free rates, earnings growth, and payout ratios—are important in explaining variations in P/E. Most important, although discount rates (which contain the ERP) are variable, they are also mean reverting; thus, changes in growth opportunities, rather than in the total discount rate, explain 95 percent of the variation in P/E.

Adopting a historical emphasis, as several of the other authors have, Jeremy Siegel looks back even further to emphasize continuities in the numbers that underlie the historical excess return and estimates of the ERP. He shows that the underperformance of real equity returns in the past 10 years relative to the historical average (6–7 percent) was just about offset by the outperformance of the previous 10 years. In addition, the average historical P/Es and earnings yields have changed very little in the past decade, further supporting the notion of stability in the forward-looking ERP. Siegel closes by observing, consistent with finance theory, that the dividend payout ratio has declined along with dividend yield but that it was offset by the growth of future earnings and dividends.

Rajnish Mehra looks back in a different way, asking whether the result of his original groundbreaking work, which predicted a very low ERP, is still warranted. Taking a long-term view that combines supply and demand considerations, he argues that higher estimates of the ERP typically depend on three basic assumptions that need rethinking because they lead to overestimations of aggregate risk. First, the risk-free rate of return should be matched to the duration of liabilities, which suggests using higher inflation-linked bond or mortgage returns rather than the more commonly used T-bill rate. Second, most estimates ignore the idea that households borrow considerably more than they lend, thus inflating the ERP. Third, younger investors have a higher demand for equities than middle-aged and older investors, but younger investors find it harder than older investors to borrow. These life-cycle and borrowing constraints artificially raise the ERP and the bond yield. Taken together, these corrections greatly reduce forward ERP estimates. One consequence of this analysis is that as the Baby Boomers retire and raise the demand for bonds, it is possible that the ERP will be higher in the future.

In sum, the papers collected in this volume share a general emphasis on supply factors and models for the historical excess return as well as the forward-looking equity risk premium. After 10 years of low and highly volatile equity returns, there is little consensus about the stability of the ERP over changing regimes and time horizons. Interestingly, the group appears to be in agreement more on the actual size of the ERP over the next few years (most agree that it is in the 4 percent range) than on its stability.

 ©2011 The Research Foundation of CFA Institute

Another Perspective: Regimes and Circumstantial Drivers

Rather than try to resolve what may be unresolvable differences in perspective on the ERP, and given the understandable challenges of evidence, inference, and prediction in this area, it may be useful to adopt a different approach—one that acknowledges and reflects the inherent multiplicity and diversity among (1) interest rate and market regimes and (2) investor perspectives.

The ERP is typically discussed as an expected return increment needed to compensate a universal or typical investor for accepting equity risk. This simple, and thus attractive, definition tempts us to think of a single investor deciding, on the margin, whether to move from a "riskless" fixed-income base into equities. The higher the ERP, the more the investor can expect to gain from a move from fixed income to equities and the higher the expected allocation to stocks. The lower the risk premium, the lower the expected gain and the lower the allocation to equities.

One implication of this single-premium concept is the assumption that it is possible to forecast a single "headline" ERP. This assumption is built into most discussions of the risk premium and most applications. Of course, these discussions and applications must take into account variables that affect the headline number. **Exhibit 2** is a far-from-exhaustive list of these "objective" drivers, including the selection of the risk-free asset base, the type of equities under consideration, real interest rate regimes, inflation expectations, other macro trends, earnings expectations, variations in the premium over time, and other considerations that can affect the forecast of a risk premium.

Each of these important variables can drive differences in calculations of the ERP. These variables have received considerable attention from analysts as well as from academics in search of the actual risk premium, including many of the contributors to this volume. Some of the differences in perspectives may be better understood by noting that the dynamics among macroeconomic and valuation factors, and their effects on the ERP, may be nonlinear. This nonlinearity can be seen in an admittedly simplistic form in **Exhibit 3**, in which the analysis is tied to interest rate regimes, which are nonlinearly associated with equity valuations. In other words, one can observe a sweet spot in P/Es and other valuations associated with moderate real long-term interest rates (2–3 percent), with a drop in valuations for lower and higher interest rate regimes. The relationships among some of the factors listed here display loosely connected tendencies rather than strong tight unities (e.g., inflation).

Exhibit 2. Objective Drivers of ERP Differences

Risk-Free Asset	Equity Class	Real Interest Rate Trend	Inflation Expectations	Other Macro Assumptions	Earnings Expectations	Dividend Trend	ERP Variations
Treasury bills	U.S. equities	High	High	Macroeconomy	High	Rising	Volatility
Treasury notes	Global equities	Medium	Medium	Demographics	Medium	Falling	Volatility of volatility
Inflation-linked bonds	Large cap	Low	Low	Globalization	Low		
	Other:						
	Size						
	Value						
	Geography						
	Sector						

©2011 The Research Foundation of CFA Institute

Exhibit 3. Real Interest Rate Regimes and the ERP

Factor	Low Rates 0–1%	Sweet Spot 2–3%	High Rates 6%+
Equity risk premium	High (6%)	Low (4% or less)	High (5%)
Probability of occurrence	Low	High	Low
Financial/economic environment	Dismal	Balanced	Overheated
Inflation expectations	Low (1–2%)	Low/medium (2–3%)	High (4%+)
Discount rate/cost of capital	Medium (7%)	Medium (7%)	High (11%)
Real growth rate	Very low (2.5%)	Good (4%)	Too high (7%)
Regime persistence	Hopefully brief	Sustainable	Almost surely brief
Sustainability of current earnings	Fair (0.4)	Fair (0.4)	Good (0.7)
New investment profitability	Good when available (6%)	Good (6%)	Squeezed (2%)
"Franchise" value (FV)	Low (4.8)	High (11.4)	Low (3.2)
"Ongoing" or "tangible" value (TV)	Fair (5.7)	Fair (5.7)	Fair (6.4)
Theoretical P/E (FV + TV)	Low (10.5)	Peak (17.1)	Low (9.6)

Notes: Specific functional values have no empirical validity. They are illustrative of relative values that might be associated with P/E and other valuation components corresponding to the three growth regimes.

Source: Based on Leibowitz and Bova (2007).

The main point is the relationship between the ERP and other economic and valuation factors. Note that although the middle, or medium, interest rate regime is the sweet spot for the economy and the equity market, the ERP could remain low in these circumstances. Whether we focus on supply or demand forces, excess return expectations may be low compared with those in more uncertain times when economies are troubled or overheated. So, some of the differences in views of the ERP could be attributed to specific regime forecasts or to whether regimes play a strong or weak role in determining the ERP.

One implication of looking at these sorts of objective determinants is that they are all, at least in theory, reducible. In other words, let's imagine it is possible to gather investors together to obtain a general agreement on selection of the risk-free asset, equity index, earnings and inflation expectations, and even the pattern by which the ERP varies over time or the list of forces that cause such variation. Although agreement on these matters might not be easy to obtain, discussions would focus on issues that are subject to measurement, analysis, and objective inference. With such a general agreement, some or maybe even a great portion of the differences among investors in their ERP estimates would be reduced. But not completely.

The differences in investors' ERP estimates would not, in the end, be eliminated. These differences are not fully reducible even with agreement on measurement and benchmarks. What remains are irreducible differences based on investors' varying conditions or circumstances. Each investor might have a unique combination of circumstances that differentiates her from all other investors, not in terms of her views on how to calculate the ERP but in terms of the circumstances in which she finds herself as an investor. In turn, those unique circumstances can then affect what we might call a "personal" or "institutional" ERP, one that is specific to an individual or institution. As shown in **Exhibit 4**, these circumstances could include investment horizon, need for liquidity, rebalancing requirement, sensitivity to changing market valuations, the capacity to evaluate those changing valuations, risk tolerance, and buyer or seller orientation.

All these circumstantial drivers of investor perceptions can affect the size of the equity premium that an investor might expect or experience at any point in time. Furthermore, this expected ERP is different from a "required" ERP in that it reflects what the investor actually experiences based on his or her individual circumstances (as opposed to an ERP that is required for the investor to act). For example, investment horizon can range from nearly perpetual (some foundations and endowments) to nearly immediate (an individual investor's current living expenses). A short-term investor might not experience the same ERP as a long-term investor, either in terms of expected return or expected volatility of that return. Similarly, liquidity needs can affect the return an investor can expect; sometimes there may be a positive or negative illiquidity premium built into the ERP. And rebalancing requirements can influence return, especially if we are aware that a large set of investors must rebalance in the same direction at the same time. In turn, the ERP may vary depending on whether one is a buyer or seller (such as during late 2008 in the equity markets, when bid–ask spreads or the differential returns required by buyers and sellers froze some markets and nearly destroyed others).

Take, for example, some combinations of these dimensions as illustrated in Exhibit 4. Many long-term investors are relatively premium insensitive in that they are interested in holding rather than buying or selling. Others, such as the LSB (long-horizon valuation-sensitive buyer), may be looking to add to positions if the price (premium) is right, although the LSS (long-horizon valuation-sensitive seller) is looking to lighten holdings based on receiving an adequate premium.[2] In contrast, a liquidity-sensitive investor (e.g., hedge funds in mid-2007 and late 2008), denoted by LLS, may need to sell at nearly any

[2]See the notes to Exhibit 4 for a full explanation of the acronyms used in this discussion.

 ©2011 The Research Foundation of CFA Institute

Exhibit 4. Circumstantial Drivers of Investors' Perceptions of the ERP

Investor Type	Investment Horizon	Liquidity Bias	Rebalancing Requirement	Valuation Sensitivity	Ability to Evaluate Market	Risk Tolerance	Trade Orientation	Example
Long horizon								
LSB	Long			Sensitive	High		Buyer	Discretionary buyer looking for low premium
LSS	Long			Sensitive	Low		Seller	Discretionary seller looking for extra premium
LLB	Long	Liquidity bias					Buyer	Buyer at nearly any price
LLS	Long	Liquidity bias					Seller	Seller at nearly any price
LRB or LRS	Long		Rebalance				Buyer	Must rebalance when market moves
LCB or LCS	Long				High	Constant		Constant risk tolerance but evaluates and acts on changing market opportunities
LVB or LVS	Long				High	Variable		Risk tolerance depends on market conditions or changing personal circumstances
LRB or LRS	Long					Range bound		Constant risk tolerance, except in extreme market move
Short horizon								
SSB or SSS	Short			Sensitive				Daily, weekly, monthly, quarterly performance evaluation
SLB or SLS	Short	Liquidity bias						Must remain liquid

Notes: First letter: L = long horizon, S = short horizon. Second letter: S = valuation sensitive, L = liquidity bias, C = constant risk tolerance, V = variable risk tolerance, R = has rebalancing requirement. Third letter: B = buyer, S = seller.

price in order to raise cash. Other investors, such as pension funds, may need to put cash to work quickly as contributions come in the door (LLB). Still others may need to rebalance systematically as the market pushes their allocations away from a policy portfolio (LRB or LRS), and therefore, they may be relatively premium insensitive. Of course, the same individual or institution may exhibit more than one of these behaviors depending on the circumstances. The point is that these circumstances can influence the size and character of the ERP investors experience or require.

Shorter-term investors may be a smaller part of the overall equity market but may receive an outsize portion of media attention. If we put aside share repurchases and new issues, as well as the supply of equity substitutes, the term structure of the ERP and its volatility may be such that both variables have very different values over the short and long term. A high short-term volatility may look much more acceptable to a long-term investor because of his ability to ride it out. Similarly, a high short-term premium can coexist with a dreary long-term premium.

So, long-term and short-term investors might share a sensitivity to valuation metrics but in very different ways. Long-term valuation-sensitive investors (LSB and LSS) might respond to a sufficiently high long-term ERP (that is, the ERP in excess of the long-term fixed-income yield) by selling bonds to buy stocks in the belief that such an action will compensate them for long-term nominal as well as real risk. In contrast, short-term valuation-sensitive investors (SSB and SSS) may be more inclined to judge the ERP either on an absolute stand-alone basis or relative to returns from various fixed-income durations given expectations regarding yield curve movements. In these cases, price volatility looms large as a risk factor, so short-term investors need a much greater premium inducement to get them to prefer equities to bonds over their short horizon.

One should also consider not just the effects of circumstantial ERP on investor behavior but also the effects of investor behavior on the ERP. As buyers and sellers meet in the marketplace, the transaction size, urgency, other asset holdings, and other circumstances could dampen or exacerbate equity premium movements. Rebalancers and especially liquidity-sensitive sellers may be relatively insensitive to price and premium and thus have a moderating effect on ERP variations. Both valuation-sensitive and valuation-insensitive investors could affect the equity premium. Valuation-sensitive investors are looking for a desired or required price or premium, so their actions will tend to move the market in that direction. The impact of actions by valuation-insensitive investors may be unpredictable because they purchase or sell shares at times that could inadvertently push the equity premium up or down.

 ©2011 The Research Foundation of CFA Institute

Some transactions, however, might have little effect on the marginal ERP. In general, the marginal ERP value is likely to be determined by one type of buyer interacting with one type of seller. Although we often think of both the marginal buyer and seller as savvy and valuation sensitive, an equally savvy investor on one side may not be able to exercise valuation sensitivity. For example, a long-term liquidity-sensitive buyer (LLB) might be content buying at a price set by a short-term valuation-sensitive seller (SSS) who thinks that equities are currently overpriced. The sum of all such forces would theoretically combine into a pair of supply and demand curves, which could be smooth, lumpy, kinked, and certainly multidimensional (e.g., with term structure characteristics and regime dependency). Thus, we can see how the interplay of these multiple circumstantial forces can lead to a risk premium that is far more multifaceted and complex than is typically envisioned in the standard discount models, even when we take into account structural and cyclical changes in the more objective factors cited in Exhibit 2.

Overlaid on all these issues may be behavioral effects, such as systematic investor misperceptions and behavioral anomalies, that affect buying and selling behavior (the behavioral versus efficient markets dimension in Figure 1). But these forces are in addition to the objective and circumstantial forces just described, and they may be more invariant. Finally, our investor categories are not all mutually exclusive, and depending on circumstances, investors may shift from one type to another.

Conclusion

The past 10 years have shown that the ERP, far from being a settled matter, continues to challenge analysts. The research and observations in this volume have a number of implications for investment practice and theory. First, investors and analysts should take care to be explicit about their estimates of the ERP. We still too often use different definitions of, assumptions about, and approaches to the ERP, or leave it altogether implicit in our analyses of asset markets and valuations. Further clarity may help reduce the number of occasions when we are talking past each other. Second, we should be clear about what model we are using when we offer a forecast or explanation of the ERP. We have seen that variations in our estimates can be the result of different approaches to objective, circumstantial, and behavioral factors. Third, differing circumstances among investors lead to true, irreducible differences in the ERP that each investor may face at any given time. This final consideration underscores how the interplay of these multiple circumstantial forces can lead to a risk premium that is far more multifaceted and complex than typically envisioned in the standard discount models, even when we take into account structural and cyclical changes in the more objective factors. The papers contained in this volume richly illustrate this interplay.

References

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal*, vol. 58, no. 2 (March/April):64–85.

Asness, Clifford S. 2000. "Stocks versus Bonds: Explaining the Equity Risk Premium." *Financial Analysts Journal*, vol. 56, no. 2 (March/April):96–113.

Association for Investment Management and Research. 2002. *Equity Risk Premium Forum*. Charlottesville, VA: Association for Investment Management and Research (www.cfapubs.org/toc/cp.1/2002/2002/7).

Bansal, Ravi, and Christian Lundblad. 2002. "Market Efficiency, Asset Returns, and the Size of the Risk Premium in Global Equity Markets." *Journal of Econometrics*, vol. 109, no. 2 (August):195–237.

Campbell, John Y. 2001. "Forecasting U.S. Equity Returns in the 21st Century." In *Estimating the Real Rate of Return on Stocks Over the Long Term*. Washington, DC: Social Security Advisory Board (August).

Campbell, John Y., and Robert J. Shiller. 2001. "Valuation Ratios and the Long-Run Stock Market Outlook: An Update." NBER Working Paper 8221 (April).

Cornell, Bradford. 1999. *The Equity Risk Premium: The Long-Run Future of the Stock Market.* New York: John Wiley & Sons.

Fama, Eugene F., and Kenneth R. French. 2002. "The Equity Premium." *Journal of Finance*, vol. 57, no. 2 (April):637–659.

Goyal, Amit, and Ivo Welch. 2002. "Predicting the Equity Premium with Dividend Ratios." NBER Working Paper 8788 (February).

Graham, John R., and Campbell R. Harvey. 2001. "The Theory and Practice of Corporate Finance: Evidence from the Field." *Journal of Financial Economics*, vol. 60, nos. 2–3 (May):187–243.

Ibbotson, Roger G., and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):88–98.

Ibbotson, Roger G., and Rex A. Sinquefield. 1976. "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)." *Journal of Business*, vol. 49, no. 1 (January):11–47.

Leibowitz, Martin L., and Anthony Bova. 2007. "P/Es and Pension Funding Ratios." *Financial Analysts Journal*, vol. 63, no. 1 (January/February):84–96.

McGrattan, Ellen R., and Edward C. Prescott. 2001. "Taxes, Regulations, and Asset Prices." NBER Working Paper 8623 (December).

Philips, Thomas K. 2003. "Estimating Expected Returns." *Journal of Investing*, vol. 12, no. 3 (Fall):49–57.

Reichenstein, William. 2001. "The Investment Implications of Lower Stock Return Prospects." *AAII Journal*, vol. 23, no. 9 (October):4–7.

Ross, Stephen A., William N. Goetzmann, and Stephen J. Brown. 1995. "Survival." *Journal of Finance*, vol. 50, no. 3 (July):853–873.

 ©2011 The Research Foundation of CFA Institute

Shoven, John B. 2001. "What Are Reasonable Long-Run Rates of Return to Expect on Equities?" In *Estimating the Real Rate of Return on Stocks Over the Long Term*. Washington, DC: Social Security Advisory Board (August).

Siegel, Jeremy J. 1994. *Stocks for the Long Run: A Guide to Selecting Markets for Long-Term Growth*. Burr Ridge, IL: Irwin.

———. 2002. *Stocks for the Long Run: The Definitive Guide to Financial Market Returns and Long-Term Investment Strategies*. 3rd ed. New York: McGraw-Hill.

Welch, Ivo. 2000. "Views of Financial Economists on the Equity Premium and on Professional Controversies." *Journal of Business*, vol. 73, no. 4 (October):501–537.

The Equity Risk Premium

Roger G. Ibbotson
Professor in Practice, Yale School of Management
Chairman, Zebra Capital Management

The equity risk premium (ERP) is a concept that seems to mean different things to different people. Some people treat it as the equilibrium long-run return, whereas others treat it as their own personal estimate of the long-run return. Some discuss it as a future return, whereas others discuss it as a realized return. Some compare equity returns with long-term bond returns or yields, whereas others compare equity returns with short-term bond returns or yields. There are various ways to estimate the ERP, whether we are talking about equilibrium or personal estimates and whether we are making forecasts or measuring past realizations. In this paper, I will clarify the terminology, compare the various ways of estimating and measuring the equity risk premium, and discuss some of the other premiums that exist in both equity and other capital markets.

What is the equity risk premium? I consider it a long-run equilibrium concept that gives an estimate of the future excess return of the stock market over and above the bond market. There are several advantages to thinking of the ERP as an equilibrium concept. It provides the *market's* estimate of the excess return on stocks relative to bonds. It is neutral in the sense that it does not take advantage of any particular investor's expertise but, rather, tries to determine what the market thinks. In this way, it can be used as a benchmark for more active or dynamic forecasts of the stock market. It can also be used for long-term planning purposes in setting a long-term asset allocation or in estimating the returns that a portfolio can provide to meet various future obligations.

I have already established that from an investor's perspective, the ERP is the expected return that investors can earn on stocks in excess of bonds. From a corporation's perspective, however, the ERP is part of the cost of equity capital. When looking at a company's entire weighted average cost of capital, the ERP is usually the most important ingredient. From a valuation perspective, the ERP is used as part of the discount rate when estimating the present value of a set of future cash flows. The expected return of equity is used in all three of these contexts, and they are all equivalent to each other after taking into account certain market imperfections, such as taxes and transaction costs.

 ©2011 The Research Foundation of CFA Institute

Methods of Estimating the Equity Risk Premium

How should we estimate the equity risk premium in equilibrium over the long run? There are four primary ways. The first is to look at the historical ERPs that we get from comparing past stock returns with past bond returns. These realizations give us an idea as to the magnitude of payoffs that investors have received for taking on the extra risk of being in the stock market rather than the various bond markets. A second way is to use a consensus estimate of the opinions of all the participants in the marketplace. Because these market participants are setting the price, they must also be the investors who are buying or selling stocks to reflect their long-term outlook. A third method is to look at the demand side of the equation. In this case, we are trying to determine how much extra return an investor would demand for taking on the extra risk of buying stocks rather than bonds. The last way is to look at the supply side of the equation. Here we consider what the economy and corporations supply to the market in the form of earnings or cash flow.

Historical. Let us start with the historical perspective. **Table 1** lists the returns over the period 1926 through 2010 for the following Ibbotson indices: Large Company Stocks, Small Company Stocks, Long-Term Corporate Bonds, Long-Term Government Bonds, Intermediate-Term Government Bonds, U.S. Treasury Bills, and Inflation. The geometric mean annualized return from Large Company Stocks was 9.9 percent, and the arithmetic mean return was 11.9 percent. The Long-Term Government Bond geometric mean return was 5.5 percent, and the arithmetic mean return was 5.9 percent. The U.S. Treasury Bill geometric mean return was 3.6 percent, and the arithmetic mean return was 3.7 percent. The table demonstrates that there can be many

Table 1. Ibbotson Index Series: Summary Statistics of Annual Total Return, 1926–2010

Series	Geometric Mean	Arithmetic Mean	Standard Deviation
Large Company Stocks	9.9%	11.9%	20.4%
Small Company Stocks	12.1	16.7	32.6
Long-Term Corporate Bonds	5.9	6.2	8.3
Long-Term Government Bonds	5.5	5.9	9.5
Intermediate-Term Government Bonds	5.4	5.5	5.7
U.S. Treasury Bills	3.6	3.7	3.1
Inflation	3.0	3.1	4.2

Source: Ibbotson® SBBI®, *2011 Classic Yearbook: Market Results for Stocks, Bonds, Bills, and Inflation, 1926–2010* (Chicago: Morningstar, 2011).

ERPs even when using a single historical data period. At the high extreme, the arithmetic mean ERP of Large Company Stocks compared with U.S. Treasury Bills was 8.2 percent (11.9 percent − 3.7 percent). At the low extreme, the geometric mean ERP of Large Company Stocks compared with Long-Term Government Bonds was 4.4 percent (9.9 percent − 5.5 percent). Thus, researchers and investors often have confusing conversations with each other. Even when they might agree on the *same* historical time interval and dataset, the ERP historical measure can be anywhere in the range of 4.4–8.2 percent, depending on which definition of ERP is used.

Investors typically use the Large Company Stock geometric mean return minus the Long-Term Government Bond return as their characterization of the historical ERP, which for 1926–2010 is 4.4 percent. In corporate finance and in valuation discounting, arithmetic means are more often used. Even if a characterization of the ERP is agreed upon, however, a debate over what historical period is most representative of the future long-run return can occur. Some might want to use even longer historical periods to reduce the estimation error, which falls in proportion to the square root of time. Some might want to use shorter and more recent periods, which better reflect the current and future environment. Those who think the historical method should be used still have plenty to debate about. The historical method, however, has the great advantage that it measures what really happened. It reveals how much stocks have actually outperformed bonds over whatever interval is under investigation.

Consensus. The consensus method might appear to be a very good approach; when using this method, one attempts to obtain the estimates from the market participants themselves (i.e., the very investors who are setting the market prices). But there are a number of problems with this approach. Most of these investors have no clear opinion about the long-run outlook. Many of them have only very short-term horizons. Individual investors often exhibit extreme optimism or pessimism and make procyclical forecasts, and so following a boom, they can have ERP estimates that exceed 20 percent or 30 percent. Following a recession or a decline in stock market prices, their estimates of the ERP might even be negative. Academics and institutional investors may be more thoughtful, but any survey of their opinions would have to be very carefully designed. I have seen surveys, however, that do not seem to even clarify whether the questionnaire refers to arithmetic mean returns or geometric mean returns. Many surveys also do not make clear whether the ERP to which they refer is the excess return of stocks over government bonds or Treasury bills or some other type of bond. This lack of clarity makes the surveys very difficult to interpret. The most extensive surveys have been done by Pablo Fernández (see, for example, Fernández, Aguirreamalloa, and Corres 2011).

 ©2011 The Research Foundation of CFA Institute

Demand. The demand approach to estimating the ERP stems from the idea that investors demand an extra return for investing in stocks rather than bonds. In the capital asset pricing model (CAPM), the ERP is the central feature. The CAPM is derived from utility curves that characterize the risk–return trade-off. In the CAPM, all assets are held in the market portfolio, and the expected return of the market portfolio is sufficient to satisfy the investors' demand for stocks relative to their risk. Attempts to measure the ERP using the demand approach focus on analyzing utility functions. Mehra and Prescott (1985) first attempted to come up with reasonable measures of the ERP in this way. The ERP was very low and did not reasonably match any of the historical data. This mismatch came to be known as the "equity premium puzzle." Subsequently, many researchers have attempted to resolve the puzzle using behavioral finance, different types of utility curves, different distributional assumptions about stock returns, and risk aversion measures that are conditional on the state of the economy. In the end, the puzzle can be resolved in many ways, but the demand approach is not likely to provide a good estimate of the equity risk premium.

Supply. The supply approach attempts to estimate what the economy or the companies in the economy can supply to the market in the form of cash flows. This approach can be applied to the economy, using per capita or total GDP growth, net capital investment, and output provided to both capital and labor. It can also be applied at the corporate level, using company cash flows, earnings, dividends, payout ratios, stock share repurchases, and cash flow receipts from mergers and acquisitions. My co-authors and I used this approach in Diermeier, Ibbotson, and Siegel (1984) and in Ibbotson and Chen (2003), as did several of the authors in *The Equity Risk Premium: Essays and Explorations* (Goetzmann and Ibbotson 2006). The supply approach is a promising alternative for estimating the ERP.

Many Different Risk Premiums

Table 1 shows that the equity risk premium is not the only premium in the market. The following are some of the potential premiums:

- Long-horizon ERP (stocks – long-term government bonds)
- Short-horizon ERP (stocks – U.S. Treasury bills)
- Small-stock premium (large stocks – small stocks)
- Default premium (long-term corporate bonds – long-term government bonds)
- Horizon premium (long-term government bonds – U.S. Treasury bills)
- Real interest rate (U.S. Treasury bills – inflation)

The equity risk premium is the largest of these premiums, but all are important. We can forecast stock and bond returns of various types by restacking the various premiums. This approach is known as the "build-up method" and was first proposed in Ibbotson and Siegel (1988). **Exhibit 1** provides an example of the build-up method.

Exhibit 1. Components of Assets' Expected Returns

Stocks	Bonds	Cash	Real Estate	Small Stocks	Foreign Stocks	Foreign Bonds
				Small-stock premium	Foreign stock premium	
Equity risk premium				Equity risk premium	Equity risk premium	Foreign bond premium
Bond horizon premium	Bond horizon premium			Bond horizon premium	Bond horizon premium	Bond horizon premium
Real riskless rate	Real riskless rate	Real riskless rate	Real return on real estate	Real riskless rate	Real riskless rate	Real riskless rate
Inflation	Inflation	Inflation	Inflation	Inflation	Inflation	Inflation

Source: Ibbotson and Siegel (1988).

As Exhibit 1 shows, a small-stock return can be estimated from the following components: expected inflation, the expected real rate of interest, the bond horizon premium, the long-horizon ERP, and the small-stock premium. A corporate bond return can be estimated from the expected inflation rate, the expected real rate of interest, the horizon risk bond horizon premium, and the default risk premium. Often the first three terms (inflation, interest rate, and bond horizon premium) are combined into the long-term yield of a riskless bond because this yield is typically observed directly in the marketplace.

One reason that the ERP is so important is that it is often the largest number in the stack. The ERP is also the most important source of estimation error because it is not directly observable in the future. Instead, we have a historical record of past realizations and various other forecast methods. In this framework, the expected stock return is the sum of two components: the long-term riskless rate, which is the yield on bonds and is directly observable, and the long-horizon ERP, which can only be estimated.

©2011 The Research Foundation of CFA Institute

Other Premiums in the Market

The stock market is frequently characterized by investment styles. I have discussed the small-stock premium, and investing in small- versus large-capitalization stocks is considered an investment style. Fama and French (1993), among others, proposed the other prevalent style in the marketplace. They showed that value stocks outperform growth stocks over long periods of time. They defined value stocks as those of companies that have high book-to-market ratios. Others define value stocks as having high earnings-to-price ratios (or low price-to-earnings ratios). The premiums of value over growth stocks and small over large stocks are often characterized as risk premiums because they are long term in nature, have a positive payoff, and can be earned through passive rather than active management.

Another premium in the market that has been empirically observed is the momentum premium (see, for example, Jegadeesh and Titman 1993). Stocks that did well in the previous year tend to do well in the next year, whereas stocks that did poorly in the previous year tend to do poorly again. The momentum premium is not typically characterized as an investment style because momentum investing usually involves some form of active management to realize the excess returns. There is some evidence that momentum premiums are becoming more erratic and less predictable, perhaps because momentum is becoming so well known in the market. With so many investors taking advantage of the momentum premium, it may tend to disappear over time.

The liquidity premium is perhaps as important as any of the risk premiums. Ibbotson, Diermeier, and Siegel (1984) proposed that the three security characteristics that investors most wish to avoid and, therefore, need to be most compensated for in the long run are (1) risk, (2) lack of liquidity, and (3) taxation. This observation forms part of the demand approach to expected returns because investors demand a premium to take on risk, to give up liquidity, or to invest in a security that is heavily taxed. The liquidity premium is very well known and has been applied primarily in bond and alternative asset markets. Because a bond yield is observable, a less liquid bond can easily be seen to have a higher yield than a more liquid bond that is otherwise similar. This spread is the liquidity premium, and it can be used as another stack in the build-up method described previously. Real estate and private equity are examples of alternative investments for which investors would demand a higher return in order to compensate for the fact that they cannot easily liquidate their positions. These liquidity premiums are not observable, but it is generally accepted that a substantial portion of the return that investors receive from these types of investments must be a reward for taking on this lack of liquidity.

©2011 The Research Foundation of CFA Institute

Ibbotson, Chen, and Hu (2011) proposed a new equity investment style based on the concept of the liquidity premium. We restricted the investment universe to publicly traded stocks and found that cross-sectional differences in liquidity have a large impact on returns, even though almost every one of these stocks trades every day. Thus, the liquidity premium is important not only across asset classes but also in the continuum of liquidity within an asset class. In the case of stocks, there is a substantial difference between the returns of the most popular stocks, which are the most heavily traded, and the returns of the least popular stocks. These premiums are larger than small-stock premiums and are comparable in magnitude to value premiums. When compared with size, value, and momentum, liquidity premiums have a different but at least as powerful effect. **Table 2** provides a comparison of liquidity and size premiums.

Table 2. U.S. Equity Annual Return Quartiles, 1972–2010

	Liquidity			
Size	1 (lowest)	2	3	4 (highest)
1 (smallest)	18.17%	17.46%	13.51%	6.16%
2	16.87	15.15	11.68	6.52
3	15.15	14.36	12.87	9.56
4 (largest)	12.49	11.48	11.55	9.87

Source: Ibbotson, Chen, and Hu (2011).

Dynamic and Tactical ERP Forecasts

Most forecasts of the equity risk premium are not equilibrium forecasts. They are not attempts at estimating an ERP that can be used for long-term investment-planning purposes, the equity cost of capital in corporate finance, or the discount rate used in valuation. Rather, they are attempts to outperform the market by applying special expertise in determining whether the stock market is over- or undervalued today. Forecasts of high returns for the stock market are accompanied by recommendations to buy stocks instead of bonds, whereas low-return forecasts are accompanied by recommendations to reduce stock investments.

Of course, knowing when to buy stocks and when to sell them is very difficult, particularly at the macro level. At the individual stock level, thousands of stocks might be over- or underpriced. But at the market level, any mispricing must be systematic. For the stock market to be overpriced in aggregate, most of the individual stocks have to be overpriced, which means that the investors in aggregate must be systematically overconfident because the market price

 ©2011 The Research Foundation of CFA Institute

reflects their collective judgment. Most stock market forecasts implicitly say that the market is wrong in some way. The forecasters believe that their particular judgment is superior to the judgment of the marketplace.

In many cases, whether the forecaster is making an equilibrium forecast or a beat-the-market forecast is not very clear. The four approaches to the equity risk premium discussed in this paper are not always clearly classified as to whether they are being applied in an equilibrium context or for the purpose of beating the market. The historical approach is based on return realizations, but one can argue over whether they are representative of the future or are too high or low. The consensus approach is subject to incorrect measurement to such an extent that it may be difficult to apply in either context. The demand approach is usually more theoretical and is mostly useful in determining the broad direction—so that one can say that the ERP is a positive number and in equilibrium stocks should always be expected to outperform bonds in the long run. The supply approach has the most flexibility; investors can attempt to use it in an equilibrium context, or they can apply their special expertise in an attempt to outperform the market. For example, one might say that an aging population argues for lower returns in the future or that the increasing speed of technological change argues for higher returns in the future. Each expert places relative importance on a particular factor, which causes the experts to end up with a wide diversity of opinions.

Summary

I have defined what the equity risk premium is and how it can be used in equilibrium and beat-the-market contexts. The terminology is confusing to many investors and financial writers: They tend to mix up a future concept with a past realization, they assign a number to the ERP without clarifying which measurement of the ERP is being used, and they rarely clarify whether they are talking about the ERP in an equilibrium or a beat-the-market context.

I have also discussed various other premiums in the market. These premiums represent the differential returns of the many different asset classes and investment styles in the market. To make sound investment decisions, it is important to have good estimates of these premiums.

References

Diermeier, Jeffrey J., Roger G. Ibbotson, and Laurence B. Siegel. 1984. "The Supply of Capital Market Returns." *Financial Analysts Journal*, vol. 40, no. 2 (March/April):74–80.

Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns of Stocks and Bonds." *Journal of Financial Economics*, vol. 33, no. 1 (February):3–56.

Fernández, Pablo, Javier Aguirreamalloa, and Luis Corres. 2011. "Market Risk Premium Used in 56 Countries in 2011: A Survey with 6,014 Answers." Working Paper WP-920, IESE Business School (May).

Goetzmann, William N., and Roger G. Ibbotson, eds. 2006. *The Equity Risk Premium: Essays and Explorations*. New York: Oxford University Press.

Ibbotson, Roger G., and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):88–98.

Ibbotson, Roger G., and Laurence B. Siegel. 1988. "How to Forecast Long-Run Asset Returns." *Investment Management Review* (September/October).

Ibbotson, Roger G., Zhiwu Chen, and Wendy Y. Hu. 2011. "Liquidity as an Investment Style." Working paper, Yale University (April).

Ibbotson, Roger G., Jeffrey J. Diermeier, and Laurence B. Siegel. 1984. "The Demand for Capital Market Returns: A New Equilibrium Theory." *Financial Analysts Journal*, vol. 40, no. 1 (January/February):22–33.

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers for Stock Market Efficiency." *Journal of Finance*, vol. 48, no. 1 (March):65–91.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, vol. 15, no. 2 (March):145–161.

 ©2011 The Research Foundation of CFA Institute

Reflections After the 2011 Equity Risk Premium Colloquium

Clifford Asness
Founding and Managing Principal
AQR Capital Management, LLC

In 2001, and again in 2011, I participated in a forum about the equity risk premium. Presented here are some informal thoughts about the equity premium that I composed after the second forum. These thoughts are an eclectic collection inspired by, but not limited to, what we discussed together.

Sequels Are Rarely as Good as Originals

The 2011 forum reprised the earlier gathering with many of the same presenters from 2001. When we met in 2001, it was not long after the peak of the technology bubble (I call it a bubble, although that label is still in some dispute). At that time, equity prices were still well above historical norms, although they were lower than in March 2000. In 2011, many of us would say that equity prices are still high versus historical prices, but the divergence is nowhere near as dramatic as in 2001.

We Still Do Not All Agree about Long-Term Predictability

It is clear from the 2011 forum that a division remains among the participants that was clearly present in 2001. Some believe in long-term predictability; others do not. Thus, when equity prices are high versus fundamentals (I am assuming that we agree on how to measure this comparison), some believe conditional long-term expected real equity returns are low, and vice versa.

I am in this camp, but I have to admit the relationship is not as obvious as it may seem. Point estimates—the actual observed history—show that long-term (say, 10-year) historical rolling returns are indeed negatively related to starting prices. And the market's performance since the first forum, when high prices indeed led to very low realized equity returns, might make it seem that the case is closed.

It is incredibly hard, however, to say anything with precision and confidence about the relationship between long-term return and price because not that much independent data are available and in-sample regressions often contain biases. As was mentioned in the forum, it really comes down to what

©2011 The Research Foundation of CFA Institute

an investment manager believes about long-term returns beforehand. If a manager believes that expected returns are constant, then when prices are high, expected growth will be higher than normal (making expected returns come out the same despite the higher prices). The data in fact point in the other direction, but only weakly after accounting for all the problems. In other words, the data barely help to resolve this debate.

It has to be one way or the other; it is a mathematical identity. High prices forecast either low expected returns or high expected growth. For me, despite its low statistical power, the point estimate is still a reasonable guess. Rather than looking for a definitive relationship between high prices and subsequent low returns, I find it more useful to focus on the absolute lack of evidence that high prices forecast high future growth. The relationship is equivalent, but it is how I like to frame the problem.

This point estimate is only a small part of why I believe in predictability. It is more important to me that return predictability agrees with my intuition and prior experience, largely formed from other time-series and cross-sectional experiments. A vast body of literature shows that when prices of anything are high versus fundamentals, expected returns are low, and vice versa. For instance, in the cross-section, when a given set of stocks has high prices versus fundamentals (such as book value, earnings, or cash flow), the expected returns on these stocks are low relative to other (cheaper) stocks. This finding is nearly ubiquitous. Thus, although I find the point estimate for the equity risk premium (ERP) versus the price relationship comforting, I find it far more compelling in the context of the literature. I think the way finance works is that when prices are high, as measured against any reasonable form of fundamentals, expected returns are lower than normal, and vice versa. Admittedly, that is hard to prove, especially if the focus is only on ERP data, and clearly some are still not convinced.

I posed the following question to the 2011 group, particularly to those who were skeptical about the possibility of long-term predictability: When prices are at true extremes (e.g., the high in March 2000 or the other direction, the low in the early 1980s), would forecasters project any difference in forward-looking expected real returns? If the answer is yes, the issue then is a variation in the degree of our beliefs, not a difference in dogma. (I never quite got an answer!)

Some Still Believe Silly Ideas, but They Also Have Learned Important Truths

Ten years after the technology bubble, some unsubstantiated beliefs remain. The so-called Fed model, which is the idea that high stock prices are reasonable when nominal interest rates are low, is still very common (although no one at the forum advanced this view). My own research and others' have shown this

 ©2011 The Research Foundation of CFA Institute

proposition to be a form of money illusion with no power to predict (even noisily) long-term stock returns. But the Fed model still yields a far more bullish forecast than focusing just on equity prices (unadjusted for nominal interest rates), as it has for a long time. Its bullishness probably accounts for its continued popularity, particularly among strategists on Wall Street.

The Shiller P/E (the current price of the S&P 500 Index divided by the previous 10-year average real earnings) has become the *lingua franca* of those that discuss the ERP and how it relates to current equity prices. This choice is not because the Shiller P/E is perfect—no measure is—but simply because it is reasonable and historically consistent. It also helps to have a common standard. Recently, the Shiller P/E has been back in the news because some broker research has called it into question. The attacks are mostly ridiculous; they are based on bullish researchers using Wall Street's long-term preferred "operating" earnings, which are earnings before negative events are deducted, or throwing out historical periods that the researchers do not want in the data. If the price of the S&P 500 is compared only with other times when the price was high, then of course it will look lower.

One argument the critics advance, with some possible merit in my view, is that the most recent financial crisis was so severe that the past 10 years of earnings are too low to be a reasonable proxy for trend. Even that effect, however, is tiny and ultimately unconvincing.[1]

Finally, reflecting the controversy about predictability discussed earlier, those who have issues with the Shiller P/E assume that today's low dividend payouts are sensible because earnings will grow more in the future. Rob Arnott and I (Arnott and Asness 2003) established empirically that this notion is not only wrong but also backward for the past 140 years. Some notions die hard, and notions that are more bullish tend to die harder. Both the Fed model and the current critique of the Shiller P/E lean in the direction of liking stocks.

More optimistically, investment managers seem to have learned some important lessons since 2001. Again, many still argue about long-term mean reversion and predictability, but many also believe, as I do, that after long-term strong returns (if mirrored in higher valuations at the end), expected future returns will be lower.

[1]This argument at least is in the right direction. For instance, if instead of looking at average 10-year earnings, investors looked at median 10-year earnings (thus giving no weight to the magnitude of the crisis), the resulting Shiller P/E would be very high versus history but slightly less high compared with the conventional approach of taking the average. In my view, this minor adjustment, which still shows an overvalued stock market, is not what the bulls are looking for, but it is a reasonable adjustment to make.

In contrast, in 2001, reflecting the thinking of the technology bubble, many in the investment world seemed to believe that high past returns meant *higher* long-term future returns. This belief can creep into prices in various ways, but perhaps the simplest occurs when an investor uses a past average of realized returns to forecast the future. I cannot say this view is gone, but many investors, perhaps most, now seem to understand that it never made sense.

After a time of strong long-term returns, future long-term returns will be lower. Reasonable people may believe that future long-term returns will be unaffected. No rational investor will expect long-term returns to be higher than normal; there are far fewer of such irrational investors today than in 2001.

My Forecast and Some Thoughts on Dispersion

Even those who believe in long-term predictability should acknowledge that it is a noisy process. The standard deviation of average annual returns over 10 years around a forecast that moves with the Shiller P/E is about 4–5 percent. It is a bit tighter when the Shiller P/E is very high or low. This tightness could mean greater predictability at those times, but it could also be a bias from investors not seeing the true extremes possible in the distribution. Nonetheless, 4–5 percent is a lot for standard deviation, and it is big relative to the dispersion among all the forecasters at the forum. Bullish and bearish forecasters at the forum mostly did not differ from each other by more than one time-series standard deviation of 10-year returns. Thus, it will be very hard for anyone to claim a convincing victory!

The financial world, however, still demands a specific forecast, so I will oblige. Guesswork is always involved in making such a forecast, but the thought process around the guesswork can be interesting. I will forecast only the real (consumer price index–adjusted) return on the S&P 500, not the risk premium versus bonds. At the 2001 forum, we failed in deciding what benchmark to use in forecasting the equity risk premium, thus confusing the issue somewhat. In my view, our discussion was not meant to reflect differing bond forecasts; forecasting the real return on the S&P 500 is more to the point.

To do so, I like to start with the Shiller P/E, which was roughly 23.5 in early April 2011. I then reduce that number by 10 percent to get a measure of the current P/E using trend earnings (because earnings grow over time, the unmodified Shiller P/E is a lagging indicator of valuation). Doing so drops the Shiller P/E to about 21.5, which makes the earnings yield about 4.7 percent. To get a sustainable dividend yield, I cut the earnings yield figure in half to about 2.3 percent. Reducing the earnings yield reflects a historically reasonable payout ratio of about 50 percent, not the current payout ratio, which is lower. I am sneaking in some optimism by ignoring my own work with Arnott that

 ©2011 The Research Foundation of CFA Institute

shows growth is slower when payouts are low, as they are today. Next, I add about 1.5 percent for expected real growth in earnings. Using the Gordon growth model (Dividend/Price + Growth), the result is a long-term forecast real equity return of 3.8 percent.

Finally, I round to 4 percent (not to round is arrogantly overprecise!); that is my 10-year forecast, but with some more caveats. This rate assumes a steady state in the markets. That is, it assumes that the best forecast of the future Shiller P/E is the current Shiller P/E. A more pessimistic vision of the future would assume some regression to the long-run mean Shiller P/E, which is about 15. A very pessimistic vision of the future would assume a regression through the long-term mean, as some argue happens eventually after all bubbles. Aside from about three days in early 2009, and then only trivially, valuations have not been below historical means since well before 2000. But I am not that pessimistic.

I agree with others who have argued that valuations in the past were too low, partly because the returns that investors study are far more attainable today with diversified index funds. I think those at the forum in 2001 were just beginning to appreciate this argument, and it is one of the most important considerations when examining the historical ERP. Too often, investors take for granted that they can mimic the market's ERP by buying diversified index funds at very low fees. During much of the historical period, however, this option did not exist. Thus, investors today should require a lower total return, and pay a higher P/E, because they retain more of the return at lower risk. So, my forecast does not incorporate any mean reversion of P/Es. I will stick with a real 4 percent.

Although the journey to arrive at my forecast is messy, and as much art as science, I think the thought process is useful for investment managers.

References

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):70–87.

Equity Premiums around the World

Elroy Dimson

Leverhulme Emeritus Professor, London Business School
Visiting Professor of Finance, Cambridge Judge Business School

Paul Marsh

Emeritus Professor of Finance
London Business School

Mike Staunton

Director, London Share Price Database
London Business School

We update our global estimates of the historical equity risk premium that were first presented in *The Millennium Book: A Century of Investment Returns* (Dimson, Marsh, and Staunton 2000) and in *Triumph of the Optimists: 101 Years of Global Investment Returns* (Dimson, Marsh, and Staunton 2002). More detailed analysis is published in our annual volumes, the *Credit Suisse Global Investment Returns Yearbook* and the *Credit Suisse Global Investment Returns Sourcebook* (Dimson, Marsh, and Staunton 2011a and 2011b).

We provide estimates for 19 countries, including two North American markets (the United States and Canada), eight markets from what is now the euro currency area (Belgium, Finland, France, Germany, Ireland, Italy, the Netherlands, and Spain), five other European markets (Denmark, Norway, Sweden, Switzerland, and the United Kingdom), three Asia-Pacific markets (Japan, Australia, and New Zealand), and one African market (South Africa).

The Dimson–Marsh–Staunton (DMS) database, which is distributed by Morningstar, also includes six U.S. dollar–denominated regional indices (Dimson, Marsh, and Staunton 2011c). The indices are a 19-country World equity index, an 18-country World ex-U.S. equity index, a 13-country European equity index, and three corresponding government bond indices for the World, World ex-U.S., and Europe. For the equity indices, each country is weighted by market capitalization (or by GDP for the years before capitalizations were available). The bond indices are GDP weighted throughout.

Our dataset includes equities, long government bonds, bills, inflation, exchange rates, and GDP. More details about the data, the sources, and the index construction methods are presented in Dimson, Marsh, and Staunton (2008, 2011b).

©2011 Elroy Dimson, Paul Marsh, and Mike Staunton

Long-Run Global Returns

Investment returns can be extremely volatile. The 2000s were a period of disappointment for most equity investors, and few would extrapolate future returns from this recent experience. Including the 1990s adds a period of stock market exuberance that is also not indicative of expectations. To understand risk and return, long periods of history need to be examined. That is why we ensure that all our return series embrace 111 years of financial market history, from the start of 1900 to the end of 2010.

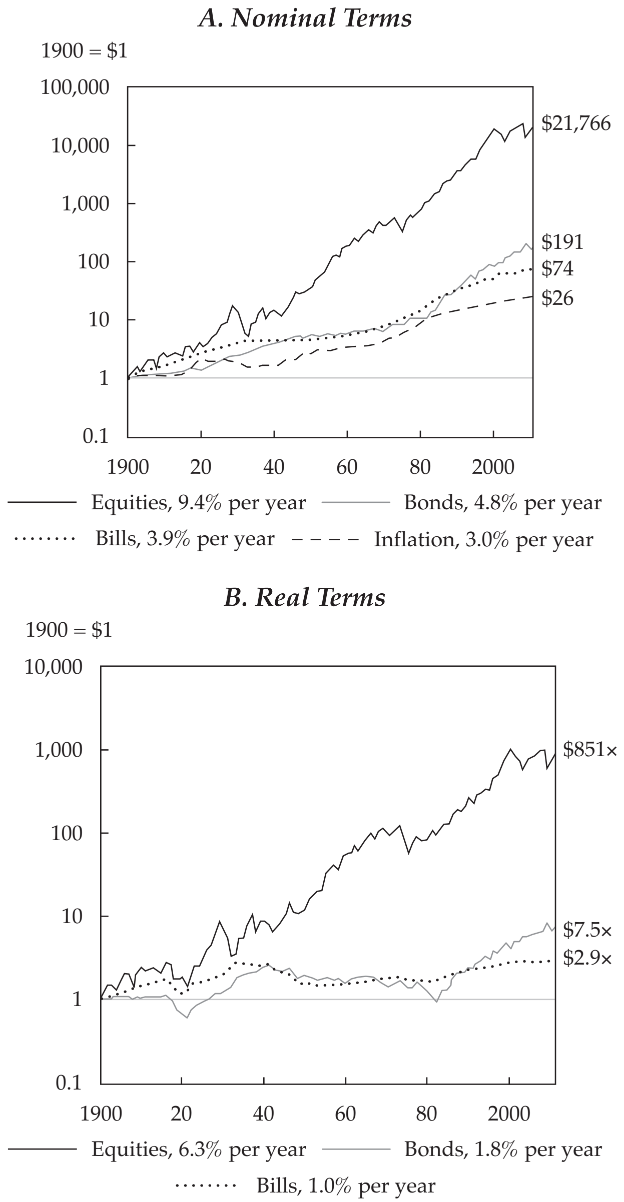
Panel A in **Figure 1** shows the cumulative total returns in nominal terms for U.S. equities, bonds, bills, and inflation for 1900–2010. Equities performed best, with an initial investment of $1 growing to $21,766 by year-end 2010. Long bonds and bills had lower returns, although they beat inflation. Their respective levels at the end of 2010 were $191 and $74, with the inflation index ending at $26. The legend shows the annualized returns were 9.4 percent for equities, 4.8 percent for bonds, and 3.9 percent for bills; inflation was 3.0 percent per year.

Because U.S. prices rose 26-fold over this period, it is helpful to compare returns in real terms. Panel B of Figure 1 shows the real returns on U.S. equities, bonds, and bills. Over the 111 years, an initial investment of $1 in equities, with dividends reinvested, would have grown in purchasing power by 851 times. The corresponding multiples for bonds and bills are 7.5 and 2.9 times the initial investment, respectively. As the legend shows, these terminal wealth figures correspond to annualized real returns of 6.3 percent for equities, 1.8 percent for bonds, and 1.0 percent for bills.

The United States is by far the world's best-documented capital market. Prior to the assembly of the DMS database, long-run evidence was invariably taken from U.S. markets and typically treated as being applicable universally. Few economies, if any, can rival the long-term growth of the United States, which makes it dangerous to generalize from U.S. historical returns. That is why we have put effort into documenting global investment returns.

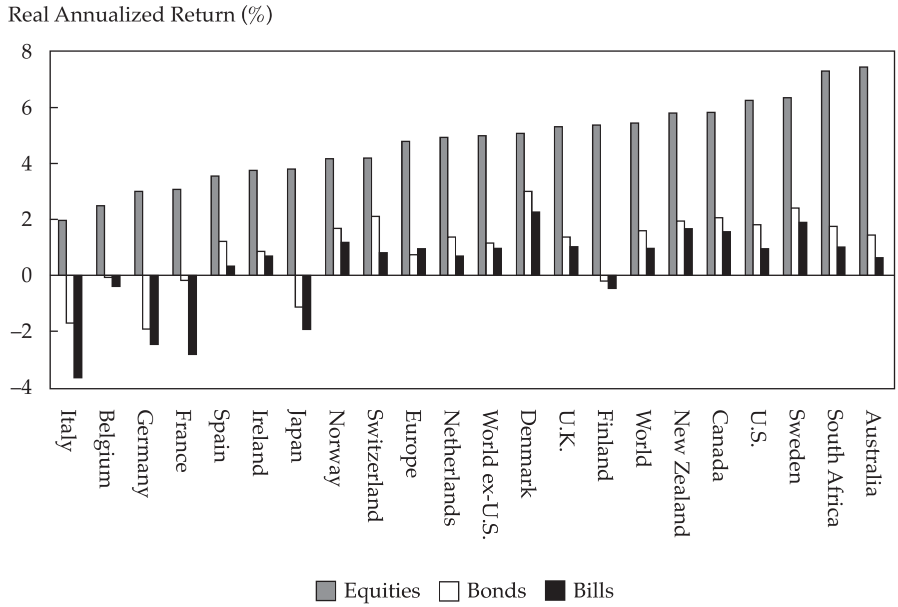
Figure 2 shows annualized real equity, bond, and bill returns for 19 countries as well as the World, the World ex-U.S., and Europe indices. The countries and regions are ranked in ascending order of equity market performance. The real equity return was positive in every location, typically 3–6 percent per year. Equities were the best-performing asset class within every market. Furthermore, bonds performed better than bills in all the countries. This pattern of equities outperforming bonds, and of bonds outperforming bills, is precisely what we would expect because equities are riskier than bonds, whereas bonds are riskier than cash.

Figure 1. Cumulative Returns on U.S. Equities, Bonds, Bills, and Inflation, 1900–2010



Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

Figure 2. Real Annualized Returns on Equities vs. Bonds and Bills Internationally, 1900–2010



Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

Figure 2 also shows that although most countries' bonds had a positive real return, six countries experienced negative returns. With the exception of Finland, the latter were also among the worst equity performers. Mostly, their poor performance dates back to the first half of the 20th century, when these countries suffered most from the ravages of war and civil strife as well as periods of high inflation or hyperinflation associated with the wars and their aftermath.

The chart confirms that the United States performed well, ranking fourth for equity performance (real 6.3 percent per year) and sixth for bonds (real 1.8 percent per year). This result confirms the conjectures that U.S. returns would be high because the U.S. economy has been such an obvious success story and that it is unwise for investors to base their future projections solely on U.S. evidence. Figure 2 helps set this debate in context, however, by showing that although U.S. stocks did well, the United States was not the top performer nor were its returns especially high relative to the world averages. The real return on U.S. equities of 6.3 percent is more than a percentage point higher than the real U.S. dollar–denominated return of 5.0 percent on the World ex-U.S. index. A common factor among the best-performing equity markets over the past 111 years is that they tended to be rich in resources and/or to be New World countries.

Table 1 provides statistics on real equity returns from 1900 to 2010. The geometric mean shows the 111-year annualized returns achieved by investors, and these are the figures that are plotted in Figure 2. The arithmetic mean shows the average of the 111 annual returns for each country or region. The arithmetic mean of a sequence of different returns is always larger than the geometric mean, and the more volatile the sequence of returns, the greater the gap between the arithmetic and geometric means. This fact is evident in the fifth column of Table 1, which shows the standard deviation of each equity market's annual returns.

The U.S. equity standard deviation of 20.3 percent places it at the lower end of the risk spectrum, ranking sixth after Canada (17.2 percent), Australia (18.2 percent), New Zealand (19.7 percent), Switzerland (19.8 percent), and the United Kingdom (20.0 percent). The World index has a standard deviation of just 17.7 percent, showing the risk reduction obtained from international diversification. The most volatile markets during this period are Germany (32.2 percent), Finland (30.3 percent), Japan (29.8 percent), and Italy (29.0 percent), which are the countries that were most affected by the world wars and inflation; Finland's case also reflects its heavy concentration in a single stock (Nokia) during recent periods. Additionally, Table 1 shows that, as one would expect, the countries with the highest standard deviations experienced the greatest range of returns—that is, the lowest minimum returns and the highest maximum returns.

Bear markets underline the risk of equities. Even in a less volatile market, such as the United States, losses can be huge. Table 1 shows that the worst calendar year for U.S. equities was 1931, with a real return of −38 percent. However, from peak to trough, U.S. equities fell by 79 percent in real terms during the 1929–31 Wall Street crash. The worst period for U.K. equities was the 1973–74 bear market, with stocks falling 71 percent in real terms and by 57 percent in a single year. More recently, 2008 had the dubious distinction of being the worst year on record for eight countries, the World index, the World ex-U.S., and Europe. The table shows that in several other countries, even more extreme returns have occurred, on both the downside and the upside.

Common-Currency Returns

So far, we have reported the real returns to a domestic equity investor based on local purchasing power in that investor's home country. For example, during 1900–2010, the annualized real return to a U.S. investor buying U.S. equities was 6.27 percent, whereas for a British investor buying U.K. equities, it was 5.33 percent. When considering cross-border investment, however, it is also

Table 1. Real (Inflation-Adjusted) Equity Returns around the World, 1900–2010

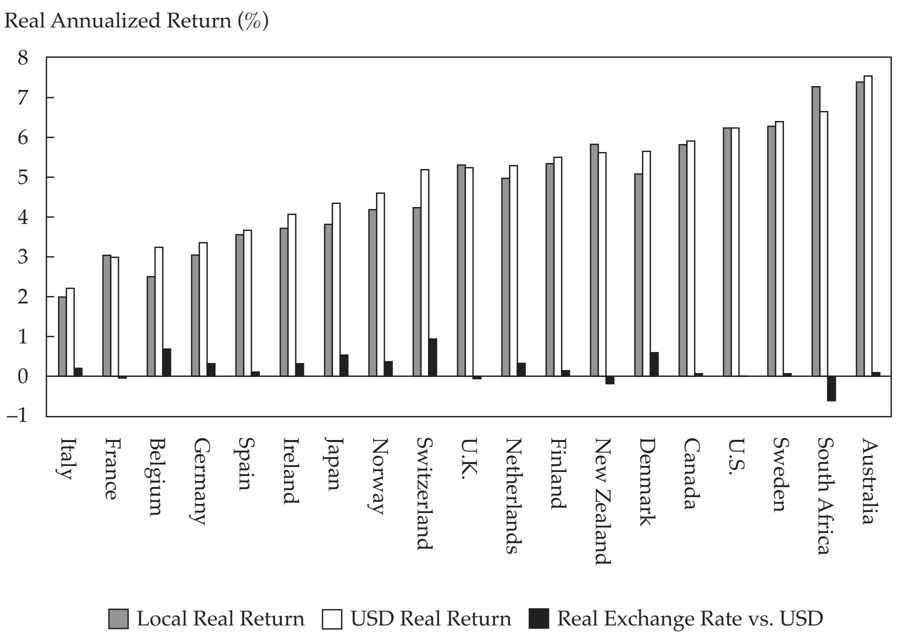
Country/Region	Geometric Mean (%)	Arithmetic Mean (%)	Standard Error (%)	Standard Deviation (%)	Minimum Return (%)	Year of Minimum	Maximum Return (%)	Year of Maximum
Australia	7.4	9.1	1.7	18.2	–42.5	2008	51.5	1983
Belgium	2.5	5.1	2.2	23.6	–57.1	2008	109.5	1940
Canada	5.9	7.3	1.6	17.2	–33.8	2008	55.2	1933
Denmark	5.1	6.9	2.0	20.9	–49.2	2008	107.8	1983
Finland	5.4	9.3	2.9	30.3	–60.8	1918	161.7	1999
France	3.1	5.7	2.2	23.5	–42.7	2008	66.1	1954
Germany	3.1	8.1	3.1	32.2	–90.8	1948	154.6	1949
Ireland	3.8	6.4	2.2	23.2	–65.4	2008	68.4	1977
Italy	2.0	6.1	2.8	29.0	–72.9	1945	120.7	1946
Japan	3.8	8.5	2.8	29.8	–85.5	1946	121.1	1952
Netherlands	5.0	7.1	2.1	21.8	–50.4	2008	101.6	1940
New Zealand	5.8	7.6	1.9	19.7	–54.7	1987	105.3	1983
Norway	4.2	7.2	2.6	27.4	–53.6	2008	166.9	1979
South Africa	7.3	9.5	2.1	22.6	–52.2	1920	102.9	1933
Spain	3.6	5.8	2.1	22.3	–43.3	1977	99.4	1986
Sweden	6.3	8.7	2.2	22.9	–43.6	1918	89.8	1905
Switzerland	4.2	6.1	1.9	19.8	–37.8	1974	59.4	1922
United Kingdom	5.3	7.2	1.9	20.0	–57.1	1974	96.7	1975
United States	6.3	8.3	1.9	20.3	–37.6	1931	56.3	1933
Europe	4.8	6.9	2.0	21.5	–46.6	2008	76.0	1933
World ex-U.S.	5.0	7.0	1.9	20.4	–43.3	2008	79.3	1933
World	5.5	7.0	1.7	17.7	–40.4	2008	69.9	1933

Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

necessary to account for exchange rate movements—for example, a U.S. investor buying U.K. equities or a U.K. investor buying U.S. equities. Each investor now has two exposures, one to foreign equities and the other to foreign currency, and each return needs to be converted into each investor's reference currency.

Rather than just comparing domestic returns, we translate all countries' local returns into a common currency. **Figure 3** shows the results of translating from the local currency to U.S. dollars. These dollar returns are expressed as real returns, adjusted for U.S. inflation. The gray bars show the annualized real domestic currency returns from 1900 to 2010, as presented earlier. The white bars are the common-currency returns, in real U.S. dollars, from the perspective of a U.S. investor. The black bars are the difference between the annualized real local-currency return and the annualized real dollar return. The black bars equate to the annualized inflation-adjusted exchange rate movement over the same period. The gap between the two return measures is less than 1 percent per annum for every country, indicating that purchasing power parity (PPP) held reasonably closely over the very long run (see Taylor 2002).

Figure 3. Real Annualized Equity Returns in Local Currency and U.S. Dollars, 1900–2010



Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

In Figure 3, countries are ranked in ascending order based on the white bars, which show the annualized real dollar returns to a U.S. investor. Because PPP tends to hold, equity markets have a similar ranking whether they are ranked by domestic real returns or by their real dollar returns. Note that although the magnitude of the returns varies according to the choice of common currency, the rankings of the countries are the same regardless of which reference currency is used.

Worldwide Premium

Investment in equities has proven rewarding over the long run, but as we noted in Table 1, it has been accompanied by significant variability of returns. Investors do not like volatility—at least on the downside—and will be prepared to invest in riskier assets only if there is some compensation for this risk (for more on this subject, see Dimson, Marsh, and Staunton 2004). The reward for equity risk that investors have achieved in the past can be measured by comparing the return on equities with the return from risk-free investments, such as Treasury bills. The difference between equity and bill returns is known as the "equity risk premium." For long-term government bonds, the difference between bond and bill returns is referred to as the "maturity premium." Although our focus in this article is on the equity risk premium, we provide up-to-date evidence on the maturity premium in Dimson, Marsh, and Staunton (2011b).

We measure the historical equity risk premium by taking the geometric difference between the equity return and the risk-free return. The formula is

$$(1 + \text{Equity rate of return}) / (1 + \text{Risk-free return}) - 1.$$

For example, if we were evaluating stocks with a one-year return of 21 percent relative to T-bills yielding 10 percent, the realized equity risk premium would be 10 percent because (1 + 21/100) / (1 + 10/100) is equal to 1 + 10/100 and deducting 1 gives a premium of 10/100, which is 10 percent. This measure of the risk premium is based on a ratio, and it thus has no numeraire. It is hence unaffected by whether returns are computed in dollars or pounds or euros or by whether returns are expressed in nominal or real terms.

Our preferred benchmark for the risk-free return is Treasury bills (i.e., very short-term, default-free, fixed-income government securities, or going back in history, the closest available equivalent in the years before T-bills became available). Many people, however, also measure the equity premium relative to long bonds, so we report both measures, even though bonds are clearly far from risk free in real terms. Detailed statistics on the equity risk premium relative to bills and bonds are given in **Table 2** and **Table 3**.

Table 2. Worldwide Equity Risk Premiums Relative to Bills, 1900–2010

Country/Region	Geometric Mean (%)	Arithmetic Mean (%)	Standard Error (%)	Standard Deviation (%)	Minimum Return (%)	Year of Minimum	Maximum Return (%)	Year of Maximum
Australia	6.7	8.3	1.7	17.6	–44.4	2008	49.2	1983
Belgium	2.9	5.5	2.3	24.7	–58.1	2008	130.4	1940
Canada	4.2	5.6	1.6	17.2	–34.7	2008	49.1	1933
Denmark	2.8	4.6	1.9	20.5	–50.6	2008	95.3	1983
Finland	5.9	9.5	2.9	30.2	–53.6	2008	159.2	1999
France	6.0	8.7	2.3	24.5	–44.8	2008	85.7	1941
Germany[a]	5.9	9.8	3.0	31.8	–45.3	2008	131.4	1949
Ireland	3.0	5.3	2.0	21.5	–66.7	2008	72.0	1977
Italy	5.8	9.8	3.0	32.0	–49.1	2008	150.3	1946
Japan	5.9	9.0	2.6	27.7	–48.3	1920	108.6	1952
Netherlands	4.2	6.5	2.2	22.8	–51.9	2008	126.7	1940
New Zealand	4.1	5.7	1.7	18.3	–58.3	1987	97.3	1983
Norway	3.0	5.9	2.5	26.5	–55.1	2008	157.1	1979
South Africa	6.2	8.3	2.1	22.1	–33.9	1920	106.2	1933
Spain	3.2	5.4	2.1	21.9	–39.9	2008	98.1	1986
Sweden	4.3	6.6	2.1	22.1	–41.3	2008	84.6	1905
Switzerland	3.4	5.1	1.8	18.9	–37.0	1974	54.8	1985
United Kingdom	4.3	6.0	1.9	19.9	–54.6	1974	121.8	1975
United States	5.3	7.2	1.9	19.8	–44.1	1931	56.6	1933
Europe	3.8	5.8	2.0	21.0	–47.4	2008	76.3	1933
World ex-U.S.	4.0	5.9	1.9	19.9	–44.2	2008	79.6	1933
World	4.5	5.9	1.6	17.1	–41.3	2008	70.3	1933

[a]All statistics for Germany are based on 109 years, excluding the hyperinflationary years of 1922–1923.

Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

Table 3. Worldwide Equity Risk Premiums Relative to Bonds, 1900–2010

Country/Region	Geometric Mean (%)	Arithmetic Mean (%)	Standard Error (%)	Standard Deviation (%)	Minimum Return (%)	Year of Minimum	Maximum Return (%)	Year of Maximum
Australia	5.9	7.8	1.9	19.8	–52.9	2008	66.3	1980
Belgium	2.6	4.9	2.0	21.4	–60.3	2008	84.4	1940
Canada	3.7	5.3	1.7	18.2	–40.7	2008	48.6	1950
Denmark	2.0	3.4	1.6	17.2	–54.3	2008	74.9	1972
Finland	5.6	9.2	2.9	30.3	–56.3	2008	173.1	1999
France	3.2	5.6	2.2	22.9	–50.3	2008	84.3	1946
Germany[a]	5.4	8.8	2.7	28.4	–50.8	2008	116.6	1949
Ireland	2.9	4.9	1.9	19.8	–66.6	2008	83.2	1972
Italy	3.7	7.2	2.8	29.6	–49.4	2008	152.2	1946
Japan	5.0	9.1	3.1	32.8	–45.2	2008	193.0	1948
Netherlands	3.5	5.8	2.1	22.2	–55.6	2008	107.6	1940
New Zealand	3.8	5.4	1.7	18.1	–59.7	1987	72.7	1983
Norway	2.5	5.5	2.7	28.0	–57.8	2008	192.1	1979
South Africa	5.5	7.2	1.9	19.6	–34.3	2008	70.9	1979
Spain	2.3	4.3	2.0	20.8	–42.7	2008	69.1	1986
Sweden	3.8	6.1	2.1	22.3	–48.1	2008	87.5	1905
Switzerland	2.1	3.6	1.7	17.6	–40.6	2008	52.2	1985
United Kingdom	3.9	5.2	1.6	17.0	–38.4	2008	80.8	1975
United States	4.4	6.4	1.9	20.5	–50.1	2008	57.2	1933
Europe	3.9	5.2	1.6	16.6	–47.6	2008	67.9	1923
World ex-U.S.	3.8	5.0	1.5	15.5	–47.1	2008	51.7	1923
World	3.8	5.0	1.5	15.5	–47.9	2008	38.3	1954

[a]All statistics for Germany are based on 109 years, excluding the hyperinflationary years of 1922–1923.

Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

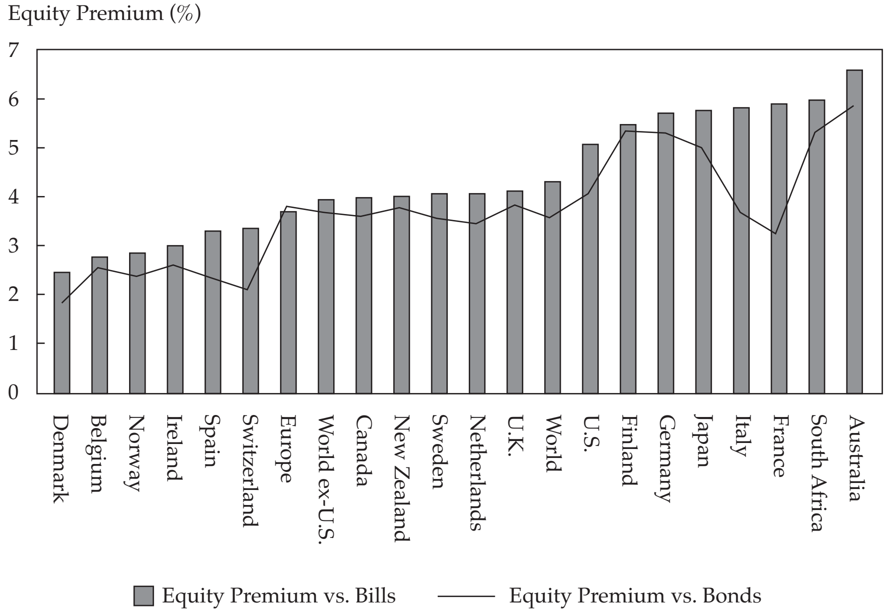
The estimates in Table 2 and Table 3 are lower than frequently quoted historical averages, such as the Ibbotson Yearbook (2011) figures for the United States and the earlier Barclays Capital (1999) studies for the United Kingdom. The differences arise from a bias (subsequently corrected) in the construction of the U.K. index used in Barclays' studies and, for both countries, our use of a long time frame (1900–2010) that incorporates the earlier part of the 20th century as well as the opening years of the 21st century, utilizing data described in Dimson, Marsh, and Staunton (2008). Our global focus also results in lower risk premiums than previously assumed. Prior views have been heavily influenced by the experience of the United States, whereas the view expressed here reflects an average of 19 countries, of which the United States is only one and in which the U.S. risk premium is somewhat higher than average.

The annualized equity premiums for the 19 countries and the World indices are summarized in **Figure 4**, in which countries are ranked according to the equity premium measured relative to bills, displayed as bars. The line plot presents each country's corresponding risk premium, measured relative to bonds. Over the entire 111 years, the annualized (geometric) equity risk premium, relative to bills, is 5.3 percent for the United States and 4.3 percent for the United Kingdom. Averaged across all 19 countries, the risk premium relative to bills is 4.6 percent, whereas the risk premium on the World equity index is 4.5 percent. Relative to long-term government bonds, the story is similar. The annualized U.S. equity risk premium relative to bonds is 4.4 percent and the corresponding figure for the United Kingdom is 3.9 percent. Across all 19 markets, the risk premium relative to bonds averages 3.8 percent; for the World index, it is also 3.8 percent.

Survivorship Bias

For the World index, our estimate of the annualized historical equity premium relative to bills is 4.5 percent. This estimate is based on the 19 countries in the DMS database, all of which survived from 1900 to 2011. These 19 countries accounted for an estimated 89 percent of the world equity market in 1900. The remaining 11 percent came from markets that existed in 1900 but for which we have been unable to obtain data. Some of these omitted markets failed to survive, and in cases like Russia in 1917 and China in 1949, investors lost all of their money. To quantify the maximum possible impact of omitted markets on the magnitude of the historical equity risk premium, we make an extreme assumption. We assume that all omitted markets became valueless and that this outcome occurred for every omitted country in a single disastrous year, rather than building up gradually. We then ask what risk premium investors would have earned if in 1900, they had purchased a holding in the entire World

Figure 4. Worldwide Annualized Equity Risk Premium Relative to Bills and Bonds, 1900–2010



Note: Statistics for Germany are based on 109 years, excluding the hyperinflationary years of 1922–1923.
Source: Based on Dimson, Marsh, and Staunton (2002) and as updated in Dimson, Marsh, and Staunton (2011b).

market, including countries omitted from the DMS database, and held this portfolio for 111 years. At the start of the period, their portfolio would have comprised an 89 percent holding in the DMS World index and an 11 percent holding in countries that we have assumed were all destined to become valueless.

Given these extreme assumptions, we demonstrate (see Dimson, Marsh, and Staunton 2008) that survivorship bias could, at most, give rise to an overstatement of the geometric mean risk premium on the World equity index by about one-tenth of a percentage point. If omitted markets did not all become valueless—and we know that very many did not—the magnitude of survivorship bias would be smaller still. Although debate continues about the precise impact of the bias because some, but not all, of these equity markets experienced a total loss of value, the net impact on the worldwide geometric mean equity premium is no more than 0.1 percent. The effect on the arithmetic mean is similar. The intuition involves the disappearance of 11 percent of the value of the market over 111 years, which represents a loss of value averaging 0.1 percent per year. We conclude that survivorship bias in world stock market returns is negligible.

Decomposing the Equity Risk Premium

Many people argue that the historical equity premium is a reasonable guide to what to expect in the future. Their reasoning is that over the long run, investors should expect good luck to balance out bad luck. If this view is correct, then the average premium investors receive should be close to the premium they required and "priced in" before the event. But even over a period as long as 111 years, this expectation may fail to be the case. It is possible that investors have enjoyed more than their share of good luck, making the past too good to last. If so, the historical premium would reflect "the triumph of the optimists" and would overstate expectations.

As an alternative approach, we seek to infer what investors may have been expecting, on average, in the past. To understand investors' expectations, we separate the historical equity premium into elements that correspond to investor expectations and elements of non-repeatable good or bad luck. In our article "The Worldwide Equity Premium: A Smaller Puzzle" (Dimson, Marsh, and Staunton 2008), we show that the equity premium can be decomposed into five components: the annualized mean dividend yield, plus the annualized growth rate of real dividends, plus the annualized expansion over time of the price/dividend ratio, plus the annualized change in the real exchange rate, minus the real risk-free rate.

Of these components, the dividend yield has been the dominant factor historically. At first sight, this may seem surprising because on a daily basis, investors' interest tends to focus mainly on the capital gains element of returns, such as stock price fluctuations and market movements. Indeed, over a single year, equities are so volatile that most of an investor's performance is attributable to capital gains or losses. Dividend income adds a relatively modest amount to each year's gain or loss. But although year-to-year performance is driven by capital appreciation, long-run returns are heavily influenced by reinvested dividends.

The difference in terminal wealth that results from reinvested dividend income is very large. As Figure 1 shows, the total real return from investing $1 in U.S. equities at the start of 1900—and reinvesting all dividend income—is an annualized 6.3 percent, such that by the start of 2011, the initial investment would have grown in purchasing power by 851 times. If dividends had not been reinvested, the initial $1 investment would have grown in purchasing power by just 8.5 times, equivalent to a real capital gain of 1.9 percent per year over the 111 years. A portfolio of U.S. equities with dividends reinvested would have grown to 100 times the value it would have attained if dividends had been spent. The longer the investment horizon, the more important dividend income becomes. For the seriously long-term investor, the value of a portfolio corresponds closely to the present value of dividends.

Components of the Equity Premium

To quantify the components of the equity premium, we examine the decomposition for all 19 countries and the World index over 1900–2010. The results are presented in **Table 4**, and we examine each component in turn. The second column of the table shows the annualized dividend yield for each market, reinforcing the point that the dividend yield has been the dominant factor historically. Across all 19 countries, the mean yield was 4.5 percent, although it was as large as 5.8 percent (South Africa) and as low as 3.5 percent (Switzerland). The annualized dividend yield for the United States (4.2 percent)

Table 4. Decomposition of the Historical Equity Risk Premium, 1900–2010

Country/Region	Geometric Mean Dividend Yield	*plus* Real Dividend Growth Rate	*plus* Expansion in the P/D Ratio	*plus* Change in Real Exchange Rate	*minus* U.S. Real Interest Rate	*equals* Equity Premium for U.S. Investors
Australia	5.76	1.10	0.48	0.10	0.96	6.53
Belgium	3.72	−1.48	0.36	0.70	0.96	2.28
Canada	4.39	0.84	0.56	0.09	0.96	4.94
Denmark	4.58	−1.13	1.64	0.57	0.96	4.69
Finland	4.76	0.49	0.09	0.15	0.96	4.53
France	3.81	−0.90	0.18	−0.04	0.96	2.05
Germany	3.66	−1.16	0.58	0.31	0.96	2.40
Ireland	4.57	−0.94	0.16	0.31	0.96	3.09
Italy	4.06	−1.52	−0.47	0.20	0.96	1.24
Japan	5.22	−2.39	1.08	0.54	0.96	3.39
Netherlands	4.94	−0.51	0.55	0.35	0.96	4.34
New Zealand	5.38	1.26	−0.84	−0.21	0.96	4.60
Norway	4.00	−0.13	0.33	0.38	0.96	3.62
South Africa	5.82	0.95	0.46	−0.61	0.96	5.65
Spain	4.18	−0.60	0.01	0.12	0.96	2.71
Sweden	4.02	1.77	0.43	0.09	0.96	5.41
Switzerland	3.48	0.46	0.28	0.94	0.96	4.22
United Kingdom	4.63	0.46	0.20	−0.06	0.96	4.27
United States	4.24	1.37	0.56	0.00	0.96	5.26
Average	4.49	−0.11	0.35	0.21	0.96	3.96
Standard dev.	0.69	1.18	0.51	0.35	0.00	1.39
World (USD)	4.11	0.83	0.48	0.00	0.96	4.49

Notes: Premiums are relative to bills. Summations and subtractions are geometric.

Source: Based on Dimson, Marsh, and Staunton (2008) and as updated in Dimson, Marsh, and Staunton (2011b).

was close to the cross-sectional average. For the World index, the annualized dividend yield was 4.1 percent, which is 3.1 percent higher than the real risk-free return from Treasury bills (see the penultimate column).

The real dividend growth rates in the third column of Table 4 reveal that in most markets, real dividend growth was lower than it was in the United States. In more than half of the countries, real dividends declined, and only four countries enjoyed real dividend growth of more than 1 percent per year. The equal-weighted average rate of real dividend growth across the 19 countries was slightly negative, although the World index's real dividend growth rate was 0.83 percent, bolstered by its heavy U.S. weighting. Dividends, and probably earnings, barely outpaced inflation. Over sufficiently long intervals, higher equity returns are generally associated with higher profits, which, in turn, generate larger dividends; comparing real equity returns (Table 1) with real dividend growth rates (Table 4) reveals a strong correlation (0.82) between the two.

The fourth column shows the expansion in the price-to-dividend ratio (P/D). Superior stock market performance and the magnitude of the historical equity risk premium are sometimes attributed to the expansion of valuation ratios, but the importance of this can be overstated. Table 4 shows that over the last 111 years, the P/D rose (dividend yields have fallen) in all but two countries, whereas the P/D of the World index grew by 0.48 percent per year. There are two possible explanations for this long-term decline in dividend yields: It may represent a repricing of equities (a downward shift in the capitalization rate or an upward shift in growth expectations), or the average payout ratio may have declined. In *Triumph of the Optimists* (Dimson, Marsh, and Staunton 2002), we note that equities enjoyed a rerating over this period but that in some countries, especially the United States, there were well-known changes in the cash distribution policies of corporations that made it necessary to take into account the impact of repurchases as well as cash dividends. The long-term multiple expansion of 0.48 percent per year is modest, however, given the improved opportunities for stock market diversification that took place over this period.

The fifth column shows the long-term change in the real (inflation-adjusted) exchange rate. As noted earlier, to examine the equity premium from the perspective of a global investor located in a specific home country, such as the United States, the real, local-currency returns need to be converted to real, common-currency returns. The annualized change in the 19 countries' real exchange rates averages only 0.21 percent per year, so this effect is small. As noted earlier, every country's real exchange rate change was within the range of ±1 percent.

The penultimate column is the historical real U.S. risk-free interest rate, and the final column computes the historical annualized equity premium for all the markets from the perspective of a U.S. investor. The realized equity premium relative to bills was, on average, 4.0 percent, with a cross-sectional standard deviation of 1.4 percent. For the U.S. dollar–denominated World index, the realized equity premium relative to bills was 4.5 percent (see the final entry in the bottom row of Table 4).

Investor Expectations

Over the long term, purchasing power parity has been a good indicator of long-run exchange rate changes (for more information, see Taylor 2002 and Dimson, Marsh, and Staunton 2011b, p. 19). The contribution to equity returns of real exchange rate changes is, therefore, an unanticipated windfall. It implies an upward bias of 0.21 percent in the cross-sectional average of the country equity premiums (there is no bias for the World index because it is denominated in the reference currency). Furthermore, as noted by Grinold, Kroner, and Siegel in their paper in this book, valuation ratios cannot be expected to expand indefinitely. Consequently, the contribution to equity returns of repricing is also likely to have been unanticipated; it implies an upward bias of 0.35 percent in the cross-sectional average of the country equity premiums and of 0.48 percent for the World index. Together, these two adjustments cause the equity premium to decline from 4.0 percent to 3.4 percent for the average country and from 4.5 percent to 4.0 percent for the World index.

In the sample of 19 countries, the average country had a long-term real dividend growth rate of slightly less than zero. In the World index, dividends outpaced inflation by an annual 0.8 percent, bolstered by the heavy weighting of the United States, where real dividends grew by 1.4 percent. But the 111-year annualized growth rate conceals a game of two halves. The 20th century opened with much promise, and only a pessimist would have believed that the next half-century would involve widespread civil and international wars, the Wall Street crash, the Great Depression, episodes of hyperinflation, the spread of communism, and the start of the Cold War. During 1900–1949, the annualized real return on the World equity index was 3.4 percent. By 1950, only a rampant optimist would have dreamed that during the following half-century, the annualized real return would be 9 percent. Yet, the second half of the 20th century was a period when many events turned out better than expected: There was no third world war, the Cuban missile crisis was defused, the Berlin Wall fell, the Cold War ended, productivity and efficiency accelerated, technology progressed, economic development spread from a few industrial countries to most of the world, and governance became stockholder driven.

The 9 percent annualized real return on world equities during 1950–1999 almost certainly exceeded expectations and more than compensated for the poor first half of the 20th century.

The question now is, What real dividend growth can be projected for the future? Pessimists may favor a figure of much less than the 0.8 percent historical average on the grounds that the "good luck" after 1950 more than outweighed the "bad luck" before 1950. Optimists may foresee indefinite real growth of 2 percent or more. Ilmanen (2011, p. 58) argues for a forward-looking approach. The yield on the World index as of year-end 2010 was 2.5 percent, well below the long-run historical average. If we assume future real dividend growth of 2 percent from this lower starting point, then the prospective premium on the World index declines to 3–3.5 percent, depending on the assumption made about the expected future real risk-free rate. The corresponding arithmetic mean risk premium would be around 4.5–5 percent, as we explained in Dimson, Marsh, and Staunton (2008). Our estimate of the expected long-run equity risk premium is less than the historical premium and much less than the premium in the second half of the 20th century. Many investment books still cite figures as high as 7 percent for the geometric mean and 9 percent for the arithmetic mean, but investors who rely on such numbers are likely to be disappointed.

Time-Varying Risk Premiums

The equity premium should be higher at times when the equity market is riskier and/or when investors are more risk averse. Yet, when markets are very volatile, extensive empirical evidence indicates that volatility tends to revert quite rapidly to the mean (for more information, see Dimson, Marsh, and Staunton 2011b, p. 34). We can, therefore, expect the period of extreme volatility to be short-lived, elevating the expected equity premium only over the relatively short run. But the premium may also vary with changes in investors' risk aversion. The latter will naturally vary among individuals and institutions and will be linked to life cycles as well as wealth levels.

The links between wealth levels and risk aversion suggest that there will be periods when risk aversion will be more or less than its long-run average. Particularly after sharp market declines, investors in aggregate will be poorer and more risk averse. At such times, markets are also typically more volatile and highly leveraged. Investors will thus demand a higher risk premium, which will drive markets even lower. Stocks are then priced to give a higher future expected return. So on average, achieved returns should be higher after market declines. The reverse logic applies following bull markets; when investors are richer, then risk aversion and, hence, the equity premium are expected to be lower.

Therefore, equity markets might be expected to exhibit mean reversion, with higher returns typically following market declines and lower returns, on average, following market rises. If there is appreciable mean reversion, then a market-timing strategy based on, for example, buying stocks after large price drops (or when market dividend yields are high or price-to-earnings ratios are low) and selling stocks after significant market rises should generate higher absolute returns. This rational economic explanation for mean reversion is based on time-varying equity premiums and discount rates. The more widely held view among investment practitioners, however, is that equity markets exhibit mean reversion for behavioral reasons—namely, that markets overreact. It is believed that in down markets, fear and over-pessimism drive prices too low, whereas in up markets, irrational exuberance and over-optimism cause markets to rise too high. In both cases, there will eventually be a correction so that equity markets mean revert.

A key difference between the rational economic view and the behavioral view is that if the former is correct, investors simply expect to earn a fair reward at all times for the risks involved. Thus, although market-timing strategies might seem to increase returns *ex post*, these higher *ex post* returns may simply reflect a realization of the higher *ex ante* returns required to compensate investors for additional risk. Put another way, the good news is that short-term expected returns are likely to be higher after market declines. The bad news is that volatility and risk aversion are correspondingly higher, and larger returns are needed to compensate for this increase. Loading up on equities at these risky times may take courage, but if subsequent returns prove to be higher, this outcome is a reward for risk, not for timing skill.

The problem with both the rational economic and behavioral views is that the evidence for mean reversion is weak. Mean reversion would imply that the equity premium is to some extent predictable, that risk over the long run is less than short-run volatility suggests, and that investors with a long horizon should favor equities compared with short-horizon investors. Yet, despite extensive research, this debate is far from settled. In a special issue of the *Review of Financial Studies*, leading scholars expressed opposing views, with Cochrane (2008) and Campbell and Thompson (2008) arguing for predictability, whereas Welch and Goyal (2008, p. 1455) find that "these models would not have helped an investor with access only to available information to profitably time the market." Cochrane's (2011) recent Presidential Address demonstrates the persistence of this controversy.

As we pointed out in our article (Dimson, Marsh, and Staunton 2004), and as articulated more formally by Pástor and Stambaugh (Forthcoming), mean reversion (if it exists) does not make equities safer in the long run. The reason is that there are three additional components of long-term risk that pull in the opposite direction. For example, an investor does not know what the average stock market return is going to be in the future, nor what the equity premium is today, nor what the other parameters of the return process are. These issues leave the investor with substantial estimation risk, and all three components of uncertainty get bigger as the investment horizon lengthens. As a result, Pástor and Stambaugh conclude that on a forward-looking basis, stocks are more risky over the long run. Diris (2011) elaborates on this view and points out that although stocks can be safer over long investment horizons, provided markets are fairly stable, they are riskier when held for the long term over periods that suffer from financial crises or other turmoil.

In summary, although some experts say that knowledge of current and recent market conditions can improve market timing, others conclude that investors cannot do better than to forecast that the future equity premium will resemble the (long-term) past. Moreover, although a lot of money could be earned if investors managed to invest at the bottom of the market, sadly the bottom can be identified only in hindsight. There are, of course, good reasons to expect the equity premium to vary over time. Market volatility clearly fluctuates, and investors' risk aversion also varies over time. But although sharply lower (or higher) stock prices may have an impact on immediate returns, the effect on long-term performance will be diluted. Moreover, volatility does not usually stay at abnormally high levels for long, and investor sentiment is also mean reverting. For practical purposes, therefore, and consistent with our discussion here, we conclude that when forecasting the long-run equity premium, it is hard to improve on evidence that reflects the longest worldwide history that is available at the time the forecast is being made.

Conclusion

Our approach is based on analyzing a comprehensive database of annual asset class returns from the beginning of 1900 to the end of 2010 and estimating realized returns and equity premiums for 19 national markets and three regions. Our estimates, including those for the United States and the United Kingdom, are lower than some frequently quoted historical averages. Yet, we find that the equity premium is positive and substantial in all markets and that survivorship bias has had only a very small effect on the estimate of the premium for the World index.

The historical equity premiums, presented here as annualized (i.e., geometric mean) estimates, are equal to investors' *ex ante* expectations plus the effect of luck. The worldwide historical premium was larger than investors are likely to have anticipated because of such factors as unforeseen exchange rate gains and unanticipated expansion in valuation multiples. In addition, past returns were also enhanced during the second half of the 20th century by business conditions that improved in many dimensions. We infer that investors expect a long-run equity premium (relative to bills) of around 3–3.5 percent on a geometric mean basis and, by implication, an arithmetic mean premium for the World index of approximately 4.5–5 percent. From a long-term historical and global perspective, the equity premium is smaller than was once thought. The equity premium survives as a puzzle, however, and we have no doubt that it will continue to intrigue finance scholars in the foreseeable future.

Elroy Dimson thanks the Leverhulme Trust, and all three authors thank the Credit Suisse Research Institute for its support.

References

Barclays Capital. 1999. *Equity Gilt Study*. London: Barclays Bank PLC.

Campbell, John Y., and Samuel B. Thompson. 2008. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, vol. 21, no. 4 (July):1509–1531.

Cochrane, John. 2008. "The Dog That Did Not Bark: A Defense of Return Predictability." *Review of Financial Studies*, vol. 21, no. 4 (July):1533–1575.

———. 2011. "Presidential Address: Discount Rates." *Journal of Finance*, vol. 66, no. 4 (August):1047–1108.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2000. *The Millennium Book: A Century of Investment Returns*. London: ABN-Amro and London Business School.

———. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

———. 2004. "Irrational Optimism." *Financial Analysts Journal*, vol. 60, no. 1 (January/February):15–25.

———. 2008. "The Worldwide Equity Premium: A Smaller Puzzle." In *The Handbook of the Equity Risk Premium*. Edited by Rajnish Mehra. Amsterdam: Elsevier.

———. 2011a. *Credit Suisse Global Investment Returns Yearbook 2011*. Zurich: Credit Suisse Research Institute.

———. 2011b. *Credit Suisse Global Investment Returns Sourcebook 2011*. Zurich: Credit Suisse Research Institute.

———. 2011c. *The Dimson–Marsh–Staunton Global Investment Returns Database*. New York: Morningstar, Inc.

Diris, Bart. 2011. "Model Uncertainty for Long-Term Investors." Erasmus University Rotterdam working paper presented at the 38th Annual Meetings of the European Finance Association, Stockholm (August).

Ibbotson/Morningstar. 2011. *Ibbotson SBBI Classic Yearbook*. Chicago: Ibbotson/Morningstar.

Ilmanen, Antti. 2011. *Expected Returns: An Investor's Guide to Harvesting Market Rewards*. Chichester, U.K.: John Wiley & Sons.

Pástor, Luboš, and Robert F. Stambaugh. Forthcoming. "Are Stocks Really Less Volatile in the Long Run?" *Journal of Finance*.

Taylor, Alan M. 2002. "A Century of Purchasing-Power Parity." *Review of Economics and Statistics*, vol. 84, no. 1 (February):139–150.

Welch, Ivo, and Amit Goyal. 2008. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, vol. 21, no. 4 (July):1455–1508.

A Supply Model of the Equity Premium

Richard C. Grinold
Former Managing Director
Barclays Global Investors

Kenneth F. Kroner
Managing Director
BlackRock

Laurence B. Siegel
Research Director, Research Foundation of CFA Institute
Senior Advisor, Ounavarra Capital LLC

The equity risk premium (ERP) is almost certainly the most important variable in finance. It tells you how much you need to save, how much you can spend, and how to allocate your assets between equities and bonds. Yet, recognized experts cannot agree on the ERP's value within an order of magnitude or even agree whether it is negative or positive. At a 2001 symposium, the predecessor of the one documented in this book, Robert Arnott and Ronald Ryan set forth an ERP estimate of –0.9 percent and Roger Ibbotson and Peng Chen proposed +6 percent.[1] The estimates in this book are much more tightly clustered, but considerable disagreement remains about how to estimate the premium as well as its size.

Grinold and Kroner (2002) proposed a model of the ERP that linked equity returns to gross domestic product (GDP) growth.[2] The key insight, which draws on earlier work by a number of authors, was that aggregate corporate profits cannot grow indefinitely much faster—or much slower—than GDP. (And as Herbert Stein was fond of reminding us, any economic trend that cannot continue forever will not.) If profits grow faster than GDP, they eventually take over the economy, leaving nothing for labor, government, natural resource owners, or other claimants. If profits grow more slowly than

[1]See Arnott and Ryan (2001); Ibbotson and Chen (2003). The Ibbotson and Chen estimate of 6 percent is an arithmetic mean expectation; their geometric mean expectation was 4 percent.
[2]A second printing of this article, from March 2004, is available online at www.cfapubs.org/userimages/ContentEditor/1141674677679/equity_risk_premium.pdf.

©2011 The Research Foundation of CFA Institute

GDP, they eventually disappear and businesses will have no profit motive to continue operating. Thus, in the very long run, the ratio of profits to GDP is roughly constant.

The title of this paper, a shortened and updated version of Grinold and Kroner (2002), refers to the "supply model" of Diermeier, Ibbotson, and Siegel (1984), who differentiated between the demand for capital market returns (what investors need to compensate them for risk) and the supply of returns (what the macroeconomy makes available). The original supply model likewise made use of a link between profits and GDP. Grinold and Kroner (2002) was titled "The Equity Risk Premium: Analyzing the Long-Run Prospects for the Stock Market," but the similarity with the title of this book forced us to rename the current paper. Although our method is designed to produce an ERP estimate that reflects both supply and demand, the link to macroeconomic performance gives it a supply-side flavor.[3]

When we revisited the estimates from Grinold and Kroner (2002), we found that not all the components could be updated with equal accuracy, so the ERP estimate provided here is subject to some important caveats regarding data adequacy. The method that we recommend, however, remains largely unchanged from Grinold and Kroner (2002).

The Equity Risk Premium Model

We define the equity risk premium as the expected total return differential between the S&P 500 Index and a 10-year par U.S. government bond over the next 10 years. Our forecast of the return to the 10-year government bond over the next 10 years is simply the yield on that bond. Therefore, the ERP becomes

$$E\left(R_S - R_B\right) = \text{Expected S\&P 500 return} - \text{10-year bond yield.} \tag{1}$$

A purer and more "modern" approach is to conduct the whole analysis in real terms and to use the yield on a 10-year par Treasury Inflation-Protected Securities (TIPS) bond or, alternatively, a 10-year TIPS strip as the relevant bond yield. The authors of some of the other papers in this book do just that. We estimate the ERP over 10-year nominal bonds, however, because that is what Grinold and Kroner (2002) did. The numerical difference between the results of the two methods, real and nominal, is not large.

Forecasting the return on the S&P 500 over the next 10 years is more difficult and, therefore, gets most of the attention in this paper. The framework we use is to decompose equity returns into several understandable pieces and then examine each piece separately.

[3] A more detailed history of the estimation of the ERP can be found in the foreword (by Laurence B. Siegel) in Kaplan (2011).

 ©2011 The Research Foundation of CFA Institute

The return to equities over a single period can always be broken down as

$$R_S = \text{Income return} + \text{Nominal earnings growth} + \text{Repricing}. \tag{2}$$

The income return is the percentage of market value that is distributed to shareholders as cash. If dividends are the only source of income, then the income return is equivalent to the dividend yield. Today, share repurchase programs (buybacks) are another common means of distributing cash to shareholders. Cash takeovers (by one company of another) should also be counted in the income return of an index that includes the stock of the acquired company.

The next two terms in Equation 2 represent the capital gain. Capital gains come from a combination of earnings growth and P/E expansion or contraction, which we call "repricing."

For expository purposes, we decompose the components further and use more precise notation. The return over a single period is

$$R = \underbrace{\frac{D}{P} - \Delta S}_{\text{Income}} + \underbrace{i + g}_{\text{Earnings growth}} + \underbrace{\Delta PE}_{\text{Repricing}}. \tag{3}$$

The first term, D/P, is simply the dividend yield. The second term, $-\Delta S$, is the percentage change in the number of shares outstanding. The percentage change in the number of shares outstanding equals the "repurchase yield" (which theoretically also includes cash takeovers) minus new shares issued (dilution); it has a negative sign because a decrease in the number of shares outstanding adds to return and an increase subtracts from return.[4] Together, the terms D/P and $-\Delta S$ measure the fraction of market capitalization that the companies in an index, in aggregate, return to shareholders in cash. Therefore, we refer to the sum of these two terms as the "income return."

The remaining terms, $i + g + \Delta PE$, make up the capital gain. The term i represents the inflation rate. The term g is the real earnings (not earnings per share) growth rate over the period of measurement. The final term, ΔPE, is the percentage change in the P/E multiple over the period. We refer to this last piece as the "repricing" part of the return.

[4] Share buybacks may be viewed as either a component of income return or a component of capital gain. An owner of a single share who holds on to the share through the share buyback program experiences the buyback as a component of capital gain because the same earnings are divided among fewer shares, which causes EPS to rise although earnings (not per share) have not changed. If the stock's P/E and all other factors are held equal, then the stock price rises. An index fund investor, however, experiences the share buyback as cash income because the index fund manager—who tenders some of the shares to the issuer to keep the stock's (now decreased) weight in the fund proportionate to its weight in the index—receives cash, which is then distributed to, or held by, fund shareholders like any other cash (tax considerations aside). We choose to view share buybacks as a component of income return.

It is important to realize that this decomposition of returns is essentially an identity, not an assumption, *so any view on the equity risk premium can be mapped into these components*. To illustrate, if the current 10-year bond yield is 3 percent, anyone who believes that the ERP is currently 4 percent must believe that the income return, nominal earnings growth, and repricing sum to 7 percent.

Historical Returns

Let us briefly consider what risk premium markets have provided historically. Over the last 85 years (1926–2010), the U.S. stock market and the intermediate-term U.S. Treasury bond market have delivered compound annual nominal returns of 9.9 percent and 5.4 percent, respectively.[5] Thus, the realized premium that stocks delivered over bonds was 4.5 percent.[6] The historical return decomposition in **Table 1** can be used to better understand this 9.9 percent annual equity return.

The income return (through dividends only, not share buybacks) on the S&P 500 was 4.1 percent annualized over this 85-year period. In this decomposition, we adjusted earnings growth for increases in the number of shares to arrive at *earnings per share (EPS) growth*. EPS grew at a rate of about 4.9 percent per year (1.9 percent real growth and 3.0 percent inflation) over the period.

Table 1. Decomposition of Total Returns on the S&P 500,[a] 1926–2010

Income return	4.10%
Real EPS growth	1.91
Inflation	2.99
P/E repricing	0.58
Within-year reinvestment return[b]	0.28
Total return	9.87%

[a]S&P 90 from January 1926 to February 1957; S&P 500 from March 1957 to 2010.
[b]Reinvestment of dividends paid during the year in the capital gain index (which consists of real EPS growth plus inflation plus P/E repricing).
Source: Morningstar/Ibbotson (used by permission).

[5]See the data for large-company stocks (i.e., the S&P 90 from January 1926 through February 1957 and the S&P 500 thereafter) in Table 2.1 in Ibbotson SBBI (2011, p. 32). Returns are before fees, transaction costs, taxes, and other costs.

[6]This amount is the arithmetic difference of geometric means. The geometric difference of geometric means, or the compound annual rate at which stocks outperformed bonds, is given by $(1 + 0.099)/(1 + 0.054) - 1 = 4.27$ percent.

 ©2011 The Research Foundation of CFA Institute

The remainder of the total return on equities was due to repricing. The P/E of the market, measured as the end-of-year price divided by trailing 12-month earnings, grew from 11.3 at year-end 1925 to 18.5 at year-end 2010.[7] This repricing works out to an additional return, or P/E expansion, of 0.58 percent per year. A common view is that this P/E expansion was understandable and reasonable in light of the technological and financial innovations over this long period. For example, accounting standards became more transparent (recent "fraud stocks" notwithstanding). Such innovations as the index fund made it easier for investors to diversify security-specific risk and to save on costs. Mutual fund complexes provided easier access to institutional-quality active management. Finally, many market observers perceive the business cycle to have been under better control in recent decades than it was in the 1920s and 1930s, which made expected earnings smoother; the recent near depression and quick recovery, at least in corporate profits and the stock market, support this view somewhat. All these factors have made equity investing less risky and contributed to the repricing over this 85-year period.

But the presence of these factors in the past does not mean that we should build continued upward repricing into our forecasts. We consider this issue later in this paper.

Chart 1 of Grinold and Kroner (2002) further dissects the return decomposition into annual return contributions. Their graph demonstrates that the noisiest component of returns is clearly P/E repricing, followed by real earnings growth. Inflation and income returns are relatively stable through time. This observation implies that our real earnings growth and repricing forecasts are likely to be the least accurate and our inflation and income return forecasts are likely to be more accurate.

Mehra and Prescott (1985), and many others, argued that the equity premium of 4.5 percent was a multiple of the amount that should have been necessary to entice investors to hold on to the risky cash flows offered by equities instead of the certain cash flows offered by bonds. This contention spawned a huge literature on the "equity risk premium puzzle."[8] We have always been perplexed by a debate that suggests that investors were wrong while a specific macroeconomic theory is right, but Rajnish Mehra sheds additional light on this question elsewhere in this book.

---

[7]Because earnings were growing very quickly at the end of 2010, the more familiar P/E calculated as the current price divided by 12-month *forward* (forecast) earnings was lower than the P/E shown here.

[8]For surveys of this literature, see Kocherlakota (1996); Mehra (2003).

Looking to the Future

Next, we will examine each term in Equation 3 to determine which data are needed to forecast these terms over the moderately long run (10 years). Later in the paper, we will combine the elements to estimate, or forecast, the total return on the S&P 500 over that time frame. Finally, we will subtract the 10-year Treasury bond yield to arrive at the expected equity risk premium.

Income Return. The income return is the percentage of market capitalization that is distributed to shareholders in cash. Currently, companies have two principal means of distributing cash to shareholders: dividend payments and share repurchases. A third method, buying other companies for cash, "works" at the index level because index investors hold the acquired company and the acquiring company if the index is broad enough.

Until the mid-1980s, dividends were essentially the only means of distributing earnings. Since then, repurchases have skyrocketed in popularity, in part because they are a more tax-efficient means of distributing earnings and in part because companies with cash to distribute may not want to induce investors to expect a distribution every quarter (and cutting dividends is painful and often causes the stock price to decline). In addition, dividend-paying companies may suffer from a stigma of not being "growth" companies.

In fact, according to Grullon and Michaely (2000), the nominal growth rate of repurchases between 1980 and 1998 was 28.3 percent. Numerous other studies have shown that share repurchases have surpassed dividends as the preferred means of distributing earnings.[9] According to Fama and French (2001), only about one-fifth of publicly traded (nonfinancial and nonutility) companies paid any dividends at the time of their study, compared with about two-thirds as recently as 1978. So the "repurchase yield" now exceeds the dividend yield.

Currently (as of 18 March 2011), the dividend yield is 1.78 percent.[10] Like a bond yield, the current (not historical average) dividend yield is likely the best estimate of the income return over the near to intermediate future, so we use 1.78 percent as our estimate of D/P in Equation 3.

To estimate the repurchase yield, we used historical data over the longest period for which data were available from Standard & Poor's, the 12 years from 1998 through 2009. We calculated the annual repurchase yield as the sum of a given year's share repurchases divided by the end-of-year capitalization of the market. **Table 2** shows these data. The average of the 12 annual repurchase yields is 2.2 percent, which we use in our ERP estimate.

[9] See, for example, Fama and French (2001); Grullon and Michaely (2000); Fenn and Liang (2000).
[10] We obtained this number at www.multpl.com/s-p-500-dividend-yield on 18 March 2011.

 ©2011 The Research Foundation of CFA Institute

Table 2. Repurchase Return of the S&P 500, 1998–2009

Year	Year-End Market Capitalization ($ billions)	Share Repurchases during Year ($ billions)	Share Repurchase Return (%)
1998	9,942.37	125	1.26
1999	12,314.99	142	1.15
2000	11,714.55	151	1.29
2001	10,463.39	132.21	1.26
2002	8,107.41	127.25	1.57
2003	10,285.83	131.05	1.27
2004	11,288.60	197.48	1.75
2005	11,254.54	349.22	3.10
2006	12,728.86	431.83	3.39
2007	12,867.85	589.12	4.58
2008	7,851.81	339.61	4.33
2009	9,927.56	137.60	1.39
Average			2.20

Source: Standard & Poor's.

It is possible to make the case for a much higher repurchase yield forecast by giving greater weight to more recent information (which is basically what we did with the dividend yield). According to Standard & Poor's (2008), "Over the past fourteen quarters, since the buyback boom began during the fourth quarter of 2004, S&P 500 issues have spent approximately $1.55 trillion on stock buybacks compared to . . . $783 billion on dividends." Although buybacks collapsed in 2009, they rebounded in 2010 and 2011. If the two-to-one ratio of buybacks to dividend payments observed by Standard & Poor's over 2004–2008 persists in the future, the repurchase yield will be as high as 3.5–3.6 percent. Aiming for a "fair and balanced" estimate, we use the lower number, 2.2 percent, which we obtained by weighting all 12 years of historical share repurchase data equally.[11]

We have not included cash buyouts in our estimate of the repurchase yield. From the perspective of an investor who holds an index containing companies A, B, C, and so forth, a cash buyout or takeover—a payment by company A to

[11]The use of this lower number is neutral, not conservative in the sense of numerically minimizing the ERP estimate. The reason is that there are offsetting biases. Our buyback estimate of 2.2 percent is too high because we do not subtract the historical contribution of buybacks to the dilution estimate (discussed later). And it is too low because very recent buyback rates have been much higher than 2.2 percent, not to mention the fact that we fully ignore the cash takeover yield.

an investor holding shares of company B in exchange for a tender of those shares—is no different from a share buyback, which is a payment by company A to an investor holding shares of A in exchange for a tender of *those* shares. Thus, the "cash buyout yield" needs to be added to the repurchase yield when summing all the pieces of $-\Delta S$. However, we do not have data for cash buyouts. If we did, they would increase our forecast of the equity risk premium (because cash buyouts must be a positive number and no other component of the ERP would change).

■ *Effect of Dilution on Income Return.* Dilution is the effect of new issuance of shares by existing companies and takes place through secondary offerings and the exercise of stock options. Dilution may be regarded as reflecting capital that needs to be injected from the labor market (or from elsewhere) into the stock market so investors can participate fully in the real economic growth described in the next section. Formally, dilution (expressed as an annual rate or a decrement to the total expected equity return) is the difference between the growth rate of dividends and the growth rate of dividends per share. If the payout ratio is assumed to be constant, dilution is also equal to the difference between the earnings growth rate and the EPS growth rate.

Grinold and Kroner (2002) estimated dilution from secondary offerings using historical data and dealt with stock options separately. Here, because we do not have the data to properly update the dilution estimates in Grinold and Kroner (2002), we use a shortcut: We directly adopt the 2 percent per year dilution estimate from Bernstein and Arnott (2003).

Bernstein and Arnott (2003) studied U.S. stocks from 1871 to 2000 and stocks from other countries over shorter periods. Instead of measuring the difference between the growth rate of earnings and that of EPS, they used a proxy: They measured the difference between the growth rate of total market capitalization and the capital appreciation return (price return) on existing shares. Dilution thus measured is net of share buybacks and cash buyouts (which are forms of negative dilution because giving cash back to shareholders is the opposite of raising capital by selling shares). The 2 percent dilution estimate for U.S. stocks is supported by evidence from other countries.[12]

[12]For a fuller discussion of dilution and an excellent description of the Bernstein and Arnott (2003) method, see Cornell (2010), who wrote, "Bernstein and Arnott (2003) suggested an ingenious procedure for estimating the combined impact of both effects [the need of existing corporations to issue new shares and the effect of start-ups] on the rate of growth of earnings to which current investors have a claim. They noted that total dilution on a marketwide basis can be measured by the ratio of the proportionate increase in market capitalization to the value-weighted proportionate increase in stock price. More precisely, net dilution for each period is given by the equation Net dilution = $(1 + c)/(1 + k) - 1$, where c is the percentage capitalization increase and k is the percentage increase in the value-weighted price index. Note that this dilution measure holds exactly only for the aggregate market portfolio" (p. 60).

 ©2011 The Research Foundation of CFA Institute

We should subtract from the 2 percent dilution estimate that part of historical dilution that was due to buybacks and cash takeovers (but *not* the part of dilution that was due to stock option issuance because these cash flows went to employees, not shareholders). We do not have the data to perform these adjustments, however, so we do not attempt them. We simply use the 2 percent estimate. (Note that the number of buybacks was tiny until the mid-1980s—that is, over approximately the first 115 years of the 130-year sample—so historical buybacks probably had a minimal impact on the average rate of dilution for the entire period.)

■ *Numerical Estimate of Income Return.* The income return forecast consists of the expected dividend yield, D/P, minus the expected rate of change in the number of shares outstanding, ΔS. The expected dividend yield is 1.78 percent. The number of new shares is expected to decline at a −0.2 percent annual rate, consisting of 2 percent dilution minus a 2.2 percent repurchase yield. After adding up all the pieces, the income return forecast is 1.98 percent.

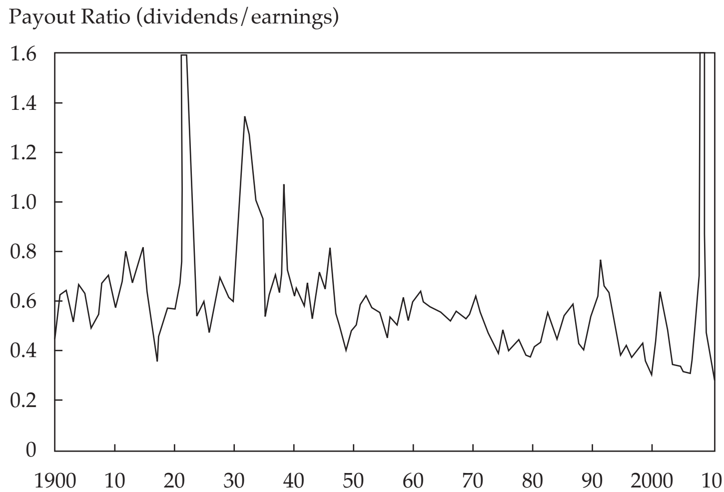
Expected Real Earnings Growth. We expect real dividend growth, real earnings growth, and real GDP growth—all expressed in aggregate, not in per share or per capita, terms—to be equal to each other.

We expect dividend and earnings growth to be equal because we assume a constant payout ratio. Although the payout ratio has fluctuated widely in the past, it has trended downward over time, presumably because of tax and corporate liquidity considerations. But the decline has effectively stopped. **Figure 1** shows the dividend payout ratio for the U.S. stock market for 1900–2010; this curious series looks as though it has been bouncing between a declining lower bound (which has now leveled off near 30 percent) and an almost unlimited upper bound. The highest values of the payout ratio occurred when there was an earnings collapse (as in 2008–2009), but companies are loath to cut dividends more than they have to.[13] The lower bound reflects payout policy during normally prosperous times.

The current lower bound of about 30 percent would be a reasonable forecast of the payout ratio, but we do not need an explicit forecast because we have already assumed that it will be constant over the 10-year term of our ERP estimate. It is helpful to have empirical support for our assumption of a constant payout ratio, however, and the recent relative stability of the lower bound in Figure 1 provides this support.

[13]The all-time high level of the payout ratio, 397 percent, occurred in March 2009, when annualized monthly dividends per "share" of the S&P 500 were $27.25 and annualized monthly earnings per "share" were $6.86.

Figure 1. Payout Ratio of the U.S. Equity Market, 1900–2010



Source: Raw data are from Robert Shiller (www.econ.yale.edu/~shiller/data/ie_data.xls, as of 4 November 2011); calculations are by the authors.

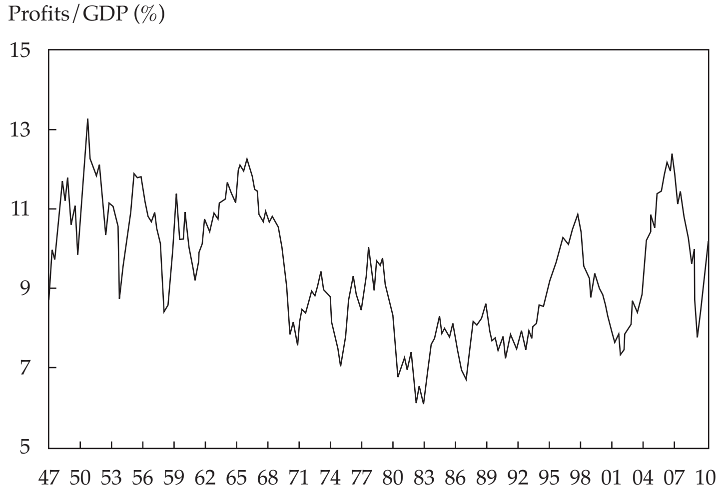
We expect real earnings growth to equal real GDP growth for the macro-consistency reason stated earlier: Any other result would, in the very long run, lead to an absurdity—corporate profits either taking over national income entirely or disappearing. **Figure 2** shows the (trendless) fluctuations in the corporate profit share of GDP since 1947.

These observations leave us with the puzzle of forecasting real GDP growth. Grinold and Kroner (2002) engaged in a fairly typical macroeconomic analysis that involved productivity growth, labor force growth, and the expected difference between S&P 500 earnings and overall corporate profits. They did not use historical averages or trends directly as forecasts; rather, they argued that the data plus other factors justified the conclusion that real GDP would most likely grow at 3 percent over the relevant forecast period and that real S&P 500 earnings would grow at 3.5 percent.

Real economic growth, by definition, equals real productivity growth plus labor force growth. Although we can update the historical productivity and labor force growth numbers, doing so would not produce an especially useful forecast any more than it did for Grinold and Kroner (2002), who distanced themselves somewhat from the productivity and labor force growth approach. The reason is that extrapolating recent trends in these components of economic growth can produce unrealistically high or low expectations, and using

 ©2011 The Research Foundation of CFA Institute

Figure 2. Quarterly U.S. Corporate Profits as a Percentage of GDP, 1947–2010



Note: Profits are pre-tax.

Source: Haver Analytics, citing U.S. National Income and Product Accounts data.

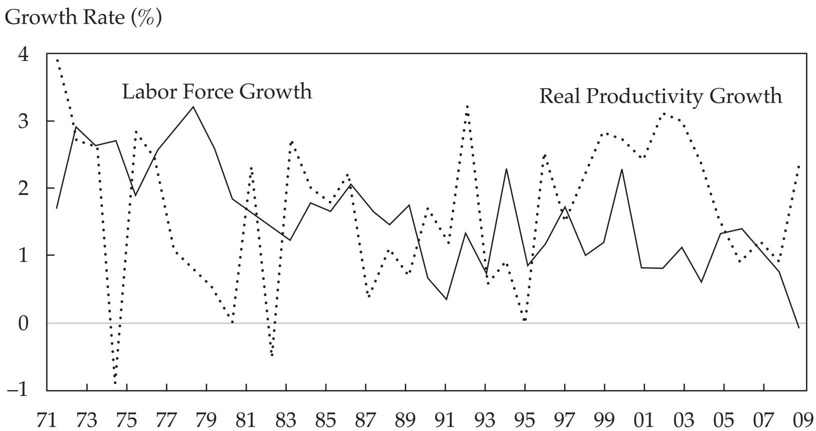
historical averages provides no insight into possible future changes in the components, which are important. Nevertheless, updates of these components are provided for informational purposes in **Figure 3**.

We can, however, use a different decomposition of real economic growth, which is also definitional: Expected GDP growth equals expected *per capita* GDP growth plus expected population growth. We believe that population growth is easier to forecast than labor force growth because the latter is partly endogenous (e.g., people work longer if they need the money because of a weak economy).[14]

Figure 4 shows that since 1789, real per capita U.S. GDP has grown at a fairly constant 1.8 percent compound annual rate. Cornell (2010) arrived at a global estimate from the high-growth postwar period (1960–2006) that is higher, but not dramatically so: 2.42 percent for mature economies and 2.79 percent for emerging economies. A cautious forecast is that the 1.8 percent growth rate will continue. If this forecast entails substantial risk, it is to the upside because an investment in the S&P 500 is not a pure bet on the U.S. economy; many, if not most, of the companies in the index are global companies that sell to markets that are growing more rapidly than the U.S. market.

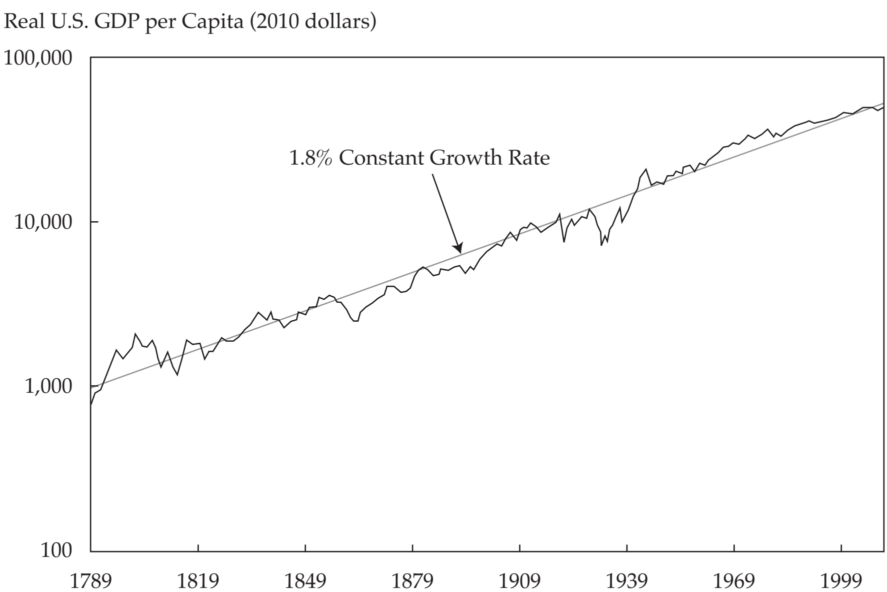
[14]Population growth is also partly endogenous (because the decisions of how many children to have, whether to emigrate, and so forth, may depend on economic performance). These effects, however, operate with long lags and tend to move the population growth rate slowly.

Figure 3. U.S. Real Productivity and Labor Force Growth Rates, 1971–2009



Source: Organisation for Economic Co-Operation and Development, OECD StatExtracts (http://stats.oecd.org/Index.aspx, as of 14 November 2011: total labour force, U.S., and labour productivity annual growth rate, U.S.).

Figure 4. Real U.S. GDP per Capita, 1789–2008



Source: Data are from Robert D. Arnott.

 ©2011 The Research Foundation of CFA Institute

We add to the 1.8 percent real per capita GDP growth estimate the Economist Intelligence Unit 10-year U.S. population growth estimate of 0.85 percent,[15] which gives a total real GDP growth forecast of 2.65 percent. This number is slightly below current consensus estimates.

This simplified method presents some difficulty because if the rate of dilution is 2 percent at all population growth rates, then population growth has a one-for-one effect on the estimate of the expected return on equities and, therefore, on the ERP. This suggests an easy beat-the-market strategy: Invest only in countries with the fastest population growth. This strategy has not worked well in the past, and even if it did over some sample period, easy beat-the-market strategies are usually illusory. Thus, the dilution estimate should probably be higher for countries with high population growth rates or for a country during periods of above-normal population growth. Although the logic of using a link to real GDP growth to forecast the stock market has great intuitive appeal, putting it into practice with any precision will take more work and more thought regarding dilution.[16]

Expected Inflation. Because we are deriving the ERP relative to Treasury bonds, we do not need our own inflation forecast as much as we need an estimate of the inflation rate that is priced into the 10-year Treasury bond market. Historical inflation rates have no bearing on this number, so we do not present them. Fortunately, the yield spread between 10-year nominal Treasury bonds and 10-year TIPS is a direct, although volatile, measure of the inflation rate that is expected by bondholders. (The spread also includes an inflation risk premium, present in nominal bond yields but not in TIPS yields, for which we need to adjust.)

---

[15]This number was obtained at http://7marketspot.com/archives/2276 on 2 May 2011 under the heading "USA economy: Ten-year growth outlook" in the column "2011–20." If we instead used real productivity growth plus labor force growth to estimate real GDP growth, we would get a slightly higher number for real productivity growth and a slightly lower number for labor force growth, which would provide a very similar overall real GDP forecast.

[16]Our simplified method has some other characteristics worth noting. It does not specifically account for the wedge between population growth and labor force growth if the proportion of retirees (or children) in the population is expected to change. A growing unproductive retiree population should be considered bearish. Many would-be retirees, however, are not financially prepared for retirement and, willingly or not, will work longer than they originally anticipated, which contributes to GDP. In addition, in an advanced technological society, an aging population distribution within the workforce is not all bad! We are accustomed to thinking of young workers as productive and older workers as unproductive, but this is the case only in a fairly primitive economy where the primary job description is something like "lift this and put it over there." In a technological society, young workers are unproductive—often startlingly so, earning only the minimum wage—and older workers produce most of the added value and make the lion's share of the money. Nevertheless, young workers' productivity grows quickly and older workers' productivity grows slowly or shrinks, so the impact of an aging workforce on *rates of change* in productivity may be less salutary than the impact on the *level* of productivity.

On 22 April 2011, the breakeven inflation rate (the yield spread described above) was 2.60 percent.[17] This rate is high by recent standards—it was as low as 1.5 percent in September 2010—but it is typical of the longer history of the series. Recent concerns about very high and rapidly growing levels of public indebtedness (of the U.S. government, of local governments in the United States, and of non-U.S. governments) have contributed to the increase in inflation expectations. We subtract 0.2 percent for the inflation risk premium to arrive at a 2.4 percent compound annual inflation forecast over the next 10 years.[18]

Expected Repricing. Grinold and Kroner (2002, p. 15, Chart 8) conducted an analysis of the market's P/E that led them to include a nonzero (–0.75 percent per year) value for the repricing term, ΔPE, in Equation 3. At the time the analysis was conducted (November 2001), the market's conventional trailing P/E (price divided by one-year trailing earnings) was a lofty 29.7 and the "Shiller P/E" (price divided by 10-year trailing real earnings) was 30.0, which prompted the authors to conclude that the P/E was likely to decline.[19] (The Shiller P/E is designed to smooth out fluctuations caused by yearly changes in earnings.) And decline it did.

Today, the situation is different. **Figure 5** shows the conventional P/E and the Shiller P/E of the U.S. market. Today's conventional P/E of 18.5 is only modestly higher than the very long-run (1900–2010) average P/E of 15.7, and it is lower than the more recent long-run (1970–2010) average P/E of 18.9. The Shiller P/E tells a slightly less favorable story: The current value is 22.4, compared with an average of 16.3 over 1900–2010 and 19.2 over 1970–2010.[20] Because it averages 10 years of trailing earnings, however, the current Shiller P/E includes an earnings collapse in 2008–2009 that is almost literally unprecedented; even the Great Depression did not see as sharp a contraction in S&P composite index earnings, although overall corporate profits in 1932 were negative. (Huge losses in a few large companies, such as those that occurred in 2008–2009, go a long way toward erasing the profits of other companies when summed across an index.) Only the depression of 1920–1921 is comparable.

Thus, we see no justification for using a nonzero value for the repricing term in Equation 3. The market's current level is already reflected in the (low) dividend yield. To include a repricing term even though the dividend yield already incorporates the market's valuation is, theoretically, not double-counting because the influence of the dividend yield is amortized over an infinite horizon,

---

[17] See www.bloomberg.com/apps/quote?ticker=USGGBE10:IND.

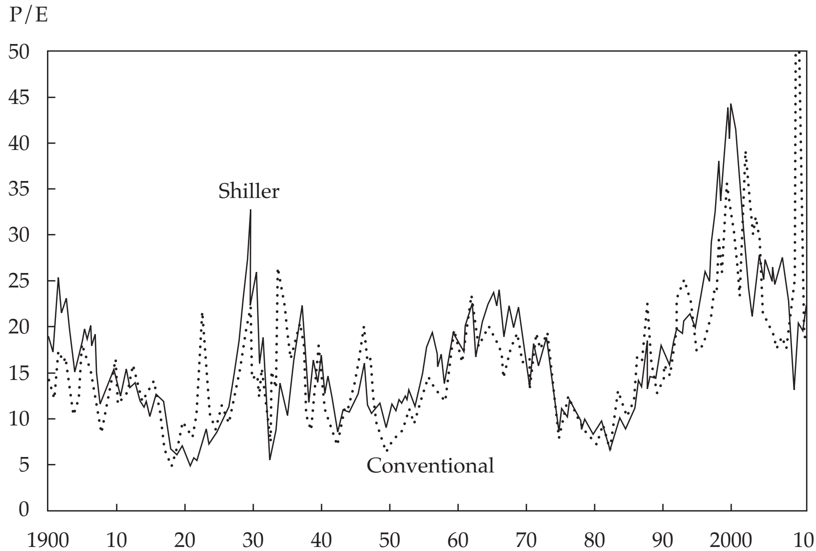
[18] This estimate of the inflation risk premium comes from Hördahl (2008, p. 31, Graph 2).

[19] Shiller (2000) describes the Shiller P/E.

[20] In this section, "current" values are as of December 2010.

 ©2011 The Research Foundation of CFA Institute

Figure 5. Conventional and Shiller P/Es for the U.S. Equity Market, 1900–2010



Note: The October 2009 conventional P/E equals 86.

Source: Spreadsheet available at Robert Shiller's website (www.econ.yale.edu/~shiller/data/ie_data.xls).

whereas our forecast is for only the next 10 years. Thus, if we believe that the market is mispriced in such a way that it will be fully corrected within 10 years, a nonzero repricing term is warranted. Although Grinold and Kroner (2002) argued that the market P/E was too high at that time and would decline at an expected rate of 0.75 percent per year over the forecast horizon, we think the market is currently not too high (or too low), and our repricing forecast is zero.

Bringing It All Together

In this section, we estimate the expected total nominal return on equities, as expressed in Equation 3, using the inputs we derived in the foregoing sections. We then subtract the 10-year nominal Treasury bond yield to arrive at our estimate of the ERP over the next 10 years.

Income return ($D/P - \Delta S$) = 1.78 percent dividend yield
– (–0.2 percent repurchase yield net of dilution)
= 1.98 percent.

Capital gain ($i + g + \Delta PE$) = 2.4 percent inflation
+ 1.8 percent real per capita GDP growth
+ 0.85 percent population growth
= 5.05 percent.

Total expected equity return = 1.98 percent + 5.05 percent
= 7.03 percent (rounded to 7 percent)
– 3.40 percent 10-year Treasury bond on 22 April 2011[21]
= 3.6 percent expected ERP over 10-year Treasuries.

Arithmetic vs. Geometric Mean Forecasts

Our forecasts thus far have been geometric means (r_G). To estimate the equivalent arithmetic mean return expectation (r_A) for use as an optimizer input, we rely on the following approximation:

$$1+r_G \approx \left(1+r_A\right)-\frac{\sigma^2}{2}. \tag{4}$$

We use standard deviations drawn from 1970 to 2010 because we do not necessarily expect bond returns to be as placid as they have been recently. Thus, for the purpose of estimating standard deviations, we include this long period because it includes the bond bear market of 1970–1980 and the dramatic subsequent recovery.[22] We obtain the following:

Expected arithmetic mean equity total return = 8.59 percent.

Expected arithmetic mean 10-year Treasury bond total return = 3.96 percent.

Difference (expected arithmetic mean ERP) = 4.63 percent.

A limitation of this study is that we use U.S., not global, macroeconomic data in our estimate of the expected return on the S&P 500. The S&P 500 is a global index, in that it contains many companies that earn most, or a substantial share, of their profits outside the United States. Perhaps global economic growth rates are more relevant to the expected return on the S&P 500 than U.S. growth rates. Future research should examine this possibility.

Assessing the Previous Grinold and Kroner Forecast

Grinold and Kroner (2002) identified three camps of ERP forecasters: "risk premium is dead," "rational exuberance," and "risk is rewarded." They called the first two views "extreme" and wished to be counted among the moderate "risk is rewarded" camp, in keeping with the belief that markets are generally efficient and that prices, therefore, do not stray far from genuine values for very long.

---

[21] This number was obtained from Yahoo! Finance on 22 April 2011.

[22] Stocks = 17.68 percent; bonds = 9.73 percent (these data are from Aswath Damodaran's website, http://pages.stern.nyu.edu/~adamodar, as of 3 June 2011).

©2011 The Research Foundation of CFA Institute

Grinold and Kroner's (2002) forecast, evaluated over 2002–2011, was too high. The main problem was the volatile repricing term. They seriously underestimated the speed with which the unusually high P/Es that then prevailed would revert toward their historical mean. In this paper, we forecast a repricing of zero, consistent with our view that the market is finally, after two bear markets and two recoveries, roughly fairly priced. Because the repricing term is noisy, we know that our current forecast is more likely to be too high or too low than just right when evaluated over the next 10 years. We believe, however, that we have identified the middle of the range of likely outcomes. Although black swans, fat tails, and tsunamis are the talk of the day, such large unexpected events tend to fade in importance as they are averaged in with less dramatic events over extended periods and the underlying long-term trends reveal themselves once more.[23] We expect moderate growth in the stock market.

The authors thank Antti Ilmanen for his very generous contribution of a number of different data sources and for his wise counsel. Paul Kaplan also provided helpful advice and contributed invaluable data.

References

Arnott, Robert D., and Ronald J. Ryan. 2001. "The Death of the Risk Premium." *Journal of Portfolio Management*, vol. 27, no. 3 (Spring):61–74.

Bernstein, William J., and Robert D. Arnott. 2003. "Earnings Growth: The Two Percent Dilution." *Financial Analysts Journal*, vol. 59, no. 5 (September/October):47–55.

Cornell, Bradford. 2010. "Economic Growth and Equity Investing." *Financial Analysts Journal*, vol. 66, no. 1 (January/February):54–64.

Diermeier, Jeffrey J., Roger G. Ibbotson, and Laurence B. Siegel. 1984. "The Supply of Capital Market Returns." *Financial Analysts Journal*, vol. 40, no. 2 (March/April):74–80.

Fama, Eugene F., and Kenneth R. French. 2001. "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" *Journal of Financial Economics*, vol. 60, no. 1 (April):3–43.

Fenn, George W., and Nellie Liang. 2000. "Corporate Payout Policy and Managerial Stock Incentives." Working paper, Federal Reserve Board (March).

Grinold, Richard C., and Kenneth F. Kroner. 2002. "The Equity Risk Premium: Analyzing the Long-Run Prospects for the Stock Market." *Investment Insights*, vol. 5, no. 3 (July):7–33.

[23] Siegel (2010) provided a skeptical look at the phenomenon of black swans.

©2011 The Research Foundation of CFA Institute

Grullon, Gustavo, and Roni Michaely. 2000. "Dividends, Share Repurchases and the Substitution Hypothesis." Unpublished manuscript, Johnson Graduate School of Management, Cornell University (April).

Hördahl, Peter. 2008. "The Inflation Risk Premium in the Term Structure of Interest Rates." *BIS Quarterly Review* (September):23–38.

Ibbotson, Roger G., and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):88–98.

Ibbotson SBBI. 2011. *2011 Classic Yearbook: Market Results for Stocks, Bonds, Bills, and Inflation, 1926–2010* (Chicago: Morningstar).

Kaplan, Paul D. 2011. *Frontiers of Modern Asset Allocation*. Hoboken, NJ: John Wiley & Sons.

Kocherlakota, Narayana R. 1996. "The Equity Premium: It's Still a Puzzle." *Journal of Economic Literature*, vol. 34, no. 1 (March):42–71.

Mehra, Rajnish. 2003. "The Equity Premium: Why Is It a Puzzle?" *Financial Analysts Journal*, vol. 59, no. 1 (January/February):54–69.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, vol. 15, no. 2 (March):145–161.

Shiller, Robert J. 2000. *Irrational Exuberance*. 2nd ed. Princeton, NJ: Princeton University Press.

Siegel, Laurence B. 2010. "Black Swan or Black Turkey? The State of Economic Knowledge and the Crash of 2007–2009." *Financial Analysts Journal*, vol. 66, no. 4 (July/August):6–10.

Standard & Poor's. 2008. "S&P 500 Stock Buybacks Retreat in Q1 but Remain Strong." Press release (18 June).

 ©2011 The Research Foundation of CFA Institute

Equity Risk Premium Myths

Robert D. Arnott
Chair and Founder, Research Affiliates, LLC

For the capital markets to "work," stocks should produce higher returns than bonds. Otherwise, stockholders would not be paid for the additional risk they take for being lower down in the capital structure. This relationship should be particularly true when stocks are compared with government bonds that (ostensibly) cannot default. It comes as no surprise, therefore, that stockholders have enjoyed outsized returns from their investments. When investors collectively expect an outsized return, as they should relative to bonds or cash, we call this expectation the "equity risk premium."

Many of the controversies surrounding the equity risk premium (ERP) are rooted in semantics: The same term is used for multiple purposes. The ERP may be based on the difference between two backward-looking rates of return—which is *not* a risk premium because it reflects past returns rather than return expectations—or on forward-looking return expectations. It may be based on single-year arithmetic return differences or compounded multiyear geometric return differences. It may be based on comparisons with cash or with bonds or with U.S. Treasury Inflation-Protected Securities (TIPS).

In any dialogue on the topic, these semantic differences mean that we may, unfortunately, be talking past one another. A 1 percent ERP (calculated as an expected multiyear geometric return difference between stocks and bonds) can be consistent with a 7 percent ERP (calculated as an expected single-year arithmetic return difference between stocks and cash at a time when the yield curve is steep, as it is at this writing), and both can be wholly consistent with a 6.5 percent observed historical excess return (the arithmetic average single-year difference between stock and cash returns over the past 60 years, which many observers erroneously label the "equity risk premium").[1]

So, perhaps this discussion should begin with definitions—the distinction between excess returns and the ERP. Because cash yields are inherently short term and hugely variable whereas forward-looking stock market returns are inherently long term and rather more stable (the sum of the yield and long-term expected growth in income is not likely to move more than 1–2 percentage points in a single year), I prefer to compare expected stock market returns with the return expectations for forward-looking government bonds or TIPS.

[1]By convention, I express the equity risk premium as a "percentage" rather than the more accurate "percentage points" or in basis points.

©2011 The Research Foundation of CFA Institute

Backward-looking excess returns are hugely variable. Over rolling 20-year spans, the gap between stock and bond market returns—the excess return for stocks—ranges from +20.7 percent to −10.1 percent per year. Wow! Most of us would consider 20 years to be a long time span. Yet, few observers would consider a 20 percent annual risk premium to be reasonable; none would consider a −10 percent risk premium reasonable.

These historical excess returns also exhibit large negative serial correlation with subsequent excess returns. Over the past 210 years, the correlation between consecutive 10-year stock market excess returns over 10-year government bonds has been a whopping −38 percent. When stocks beat bonds by a wide margin in one decade, they reversed with reasonable reliability over the next decade. This correlation is both statistically significant and economically meaningful.

Forecasting the future ERP by extrapolating past excess returns is, therefore, fraught with peril. Yet, extrapolating the past is so tempting that much of the finance community sets return expectations in exactly this fashion. No wonder our industry got it so wrong at the peak of the technology bubble in 2000: The average corporate pension fund was using an all-time-high 9.5 percent "pension return assumption" for conventional balanced 60 percent equity/40 percent bond portfolios at a time when bond yields were 6 percent and the stock market offered an all-time-low 1.1 percent dividend yield! There may also be a Machiavellian aspect to this "expectation," in that some pension plan sponsors may have known the forecasts were too high but used them anyway to avoid having to increase contributions to their pension plans.

Except when I specifically indicate to the contrary, I use the term "excess returns" to refer to realized differences between stock market returns and long-term government bond returns and the term "the ERP" to refer to expected (forward-looking) long-term differences between stock returns and long bond market expected returns (geometric or compounded annual rates). Occasionally, I use cash or long-term TIPS rather than long-term government bonds, but when I do, I acknowledge that I am doing so.

Myths

Over the years, a number of myths related to the ERP have emerged. One of the most widely "cited" myths is that the ERP is 5 percent. Before discussing the natural limits for the risk premium, I will explore an array of these ERP myths and reflect on why we so eagerly embrace myths rather than test them to objectively gauge their legitimacy.

Take, for example, the myth that the ERP is a static 5 percent. According to Ibbotson Associates (now Morningstar) data, equity investors earned a real return of 8 percent and stocks outpaced bonds by more than 5 percent from

 ©2011 The Research Foundation of CFA Institute

1926 until the early 2000s.[2] More recently, these figures have sagged to 6.5 percent and 4.5 percent, respectively. Intuition suggests that investors should not require such outsized returns in order to bear equity market risk. If we examine the historical record, neither the 8 percent real return nor the 5 percent risk premium for stocks relative to government bonds has ever been a realistic *expectation*, except at major market bottoms or at times of crisis, such as wartime.

Should investors have expected these returns in the past, and why shouldn't they continue to do so? We can break this question into two parts. First, can we derive an objective estimate of what investors had good reasons to expect in the past? Second, should we expect less in the future than we have earned in the past, and if so, why?

The answers to these questions lie in the difference between the observed excess return and the prospective risk premium. When we distinguish between past excess returns and future expected risk premiums, the idea that future risk premiums should be different from past excess returns is entirely reasonable.

Most of the ERP myths take on the character of a classic urban legend—so seductively plausible that they linger despite overwhelming evidence to the contrary. Note that most of these myths can be used to rationalize a higher, not a lower, ERP. No one seems to construct a myth or a fable to explain why we should expect lower returns!

The myths I examine include the following:

- *The risk premium is 5 percent and changes little, except perhaps in proportion to a stock's beta*. Nothing in finance theory requires any such assumption, but the notion of a large risk premium has been used to justify some truly heroic growth assumptions when yields or payout ratios have been low.
- *The ERP is static over time, across markets, and across companies*. Higher or lower yields, yield spreads, valuation multiples, and so forth have no bearing on the ERP. The proponents of this myth argue that constantly changing yields, spreads, and valuation multiples reflect changing investor expectations for future growth—in a fashion that offsets the yield, spread, or valuation changes—leaving the ERP unaltered. Nothing in neoclassical finance theory, however, suggests that the ERP must be static. Moreover, behavioral finance observers would emphatically contradict the notion of a static ERP because risk, risk expectations, and risk tolerance are all nonstatic.
- *The "ERP Puzzle": Stocks beat bonds by more than they should*. If we adhere to the view that the excess return for stocks should be measured in 10ths of a percent (10s of basis points), as most utility functions suggest for the long-term investor, this observation is true. But the ERP Puzzle seems to

[2]This section is excerpted and amended from Arnott and Bernstein (2002).

be posed as though 5 percent is the excess return that needs to be explained. Such a high excess return has not been earned in "normal" markets. In the absence of gains in valuation multiples, an excess return of 2–3 percent is more normal, and even that margin seems to be more consistent with high yields than with the low yields we observe today.

- *Stocks will beat bonds for anyone willing to think long term, which is typically taken to mean 20–30 years or longer*. This myth lingers in spite of a 41-year span (early 1968 to early 2009) in which the returns of ordinary long U.S. T-bonds eclipsed the S&P 500 Index return. Non-U.S. examples counter to this myth also abound.
- *When yields and payout ratios are low, stock buybacks can replace the dividend in a tax-advantaged fashion*. However, true buybacks—that is, buybacks that truly reduce shares outstanding rather than merely recapture shares issued in a context of management stock option redemption—are much more the exception than the rule.
- *Stock market earnings grow with GDP*. If this myth were true, the expected return on stocks would match yield plus expected GDP growth. Unfortunately, this enduring myth ignores the fact that the share of corporate profits in GDP growth consists of the growth in existing enterprises *plus* the creation of new enterprises. The "new enterprises" portion is often the larger component of real GDP growth. Therefore, the ERP is much smaller than adherents to this misconception expect.
- *Dividends do not really matter*. This myth is twofold. First, it involves the belief that *lower yields are entirely consistent with continued high return and a high ERP*. In an efficient market, investors will accept a lower yield whenever they are confident that future real growth in earnings will make up the difference. But overwhelming global evidence suggests a strong positive link between the dividend yield and both the subsequent real return for stocks and the subsequent excess return of stocks over bonds.

  The second part of this myth is that *lower payout ratios lead to faster earnings growth*. The Modigliani and Miller indifference theorem is often used to justify this view. But M&M is a theory based on a large array of simplifying assumptions and, therefore, an approximation of reality.

Both of these instances show that, in reality, dividends *do* matter.

The 5 Percent Risk Premium

Ibbotson Associates—whose annual data compendium covers U.S. stocks, T-bonds, and T-bills since January 1926—shows the S&P 500 compounding through February 2011 at an annual rate of 9.8 percent, versus 5.5 percent for

 ©2011 The Research Foundation of CFA Institute

long-term government bonds, which is an excess return of 4.3 percent. This return compounds exponentially with time. Albert Einstein whimsically declared that compound interest is "the most powerful force in the universe." Disregarding inflation, taxes, transaction costs, and fees, a $1,000 U.S. stock investment in 1926 would have ballooned to $3 million by February 2011, versus $94,000 for an investment in long-term bonds—a 32-fold difference.

In the 1980s and 1990s, stocks—bolstered by soaring valuation multiples—compounded at, respectively, 17.6 percent and 18.2 percent per year. As a result, "Stocks for the Long Run" became the mantra for long-term investing, as well as the title of a best-selling book by Siegel (2007). This view is now embedded into the psyche of an entire generation of professional and casual investors, who ignore the fact that much of that outsized return in the 1980s and 1990s was a consequence of soaring valuation multiples and tumbling yields. Because most investors anchor their decisions on personal experience, we have a population that largely assumes that this long-term 5 percent excess return of stocks over bonds is their birthright. This view constitutes the "cult of equities."

Let's Talk Really Long Term. For those willing to do the homework, very long-term stock and bond data exist for the United States. The picture of the difference between stocks and bonds if we start at 1802 is not quite as rosy as it is from 1926 to 2010; therefore, this view does not receive as much attention from the relentlessly optimistic stock sellers of Wall Street. From 1802 to 2010, U.S. stocks generated a 7.9 percent annual return, versus 5.1 percent for long-term government bonds. So, the realized excess return was cut to 2.8 percent—a one-third reduction—by including an additional 125 years of capital market history.

Of course, many observers declare 19th century data irrelevant. A lot has changed. The survival of the United States as we know it was in doubt during the first part of the century (the War of 1812), and in the middle stages, we waged a debilitating civil war. Government bonds were thus not riskless. And by modern standards, the United States was an emerging market. Citizens lived shorter lives than now, and the economy was notably short on global trade and long on subsistence agriculture. Furthermore, three major wars and four depressions—two roughly comparable to the Great Depression—occurred between 1800 and 1870, a span during which the data on market returns are notably meager.

One could as easily make the case, however, that the 20th century is not representative either. The 20th century brought great and unexpected fortune to the United States and its equity markets. The country was not invaded and occupied by a foreign power, and it did not suffer a government overthrow. For contrast, consider the return on capital for Russian investors after the Bolshevik Revolution—a 100 percent loss. Benjamin Graham cautioned on

©2011 The Research Foundation of CFA Institute

the difference between the loss *on* capital (a drop in price, from which the investor can recover) and a loss *of* capital (100 percent loss, from which the investor cannot recover). Russia's stock market was not alone in devastating losses of capital in the 20th century; 2 additional markets of the top 15 in 1900, Egypt and China, suffered a 100 percent loss of capital; Argentina, Germany (twice), and Japan (once) came close.

Markets tend to be unkind to those who ignore history, and the severity of the penalty is highly correlated with our reliance on viewing a span of history that is too short. The long history of the markets should not be ignored even when we are dealing with the shorter time horizons of most investment programs. Even for such "perpetual" institutions as university endowments, the relevant horizon is only 10–30 years. As Bernstein (1997) commented about 80–100 years of data, ". . . this kind of long run will exceed the life expectancies of most people mature enough to be invited to join such boards of trustees" (p. 22).

Nonetheless, the relevant investment span should be long enough that equity investors will be rewarded for bearing risk, right? Not always! As displayed in **Table 1**, trailing returns for stocks have not come close to the excess returns over bonds that we have all come to expect, even after stocks worldwide doubled from the lows reached during the global financial crisis that began in early March 2009. They have not come close in the United States, in the rest of the developed world, and most assuredly not in the emerging markets.

Where is the wealth creation implied by the long-term Ibbotson data? Stock market investors took the risk. They rode out every bubble, every crash, every spectacular bankruptcy and bear market during a 30-year stretch that finished with a 100 percent gain in two years. How much was their cumulative excess return for the blood, sweat, and tears spilled with all this volatility? Through 2010—a splendid span for bonds as yields tumbled for 30 years while

Table 1. Annualized Returns for Stocks over the "Long Run," for 10, 20, and 30 Years Ended 2010: Where Is the Reward?

	10-Year Return	20-Year Return	30-Year Return
S&P 500	1.41%	9.14%	10.71%
Ibbotson U.S. long-term government bonds	6.64	8.44	10.18
U.S. equity risk premium	−5.23	0.70	0.53
MSCI Europe/Australasia/Far East Index (net)	3.50	5.85	
JPM Government Bond Index: Global ex U.S. TR USD[a]	7.64	7.07	
International equity risk premium	−4.14	−1.22	

[a]TR stands for "total return."

Source: Based on data from Morningstar EnCorr.

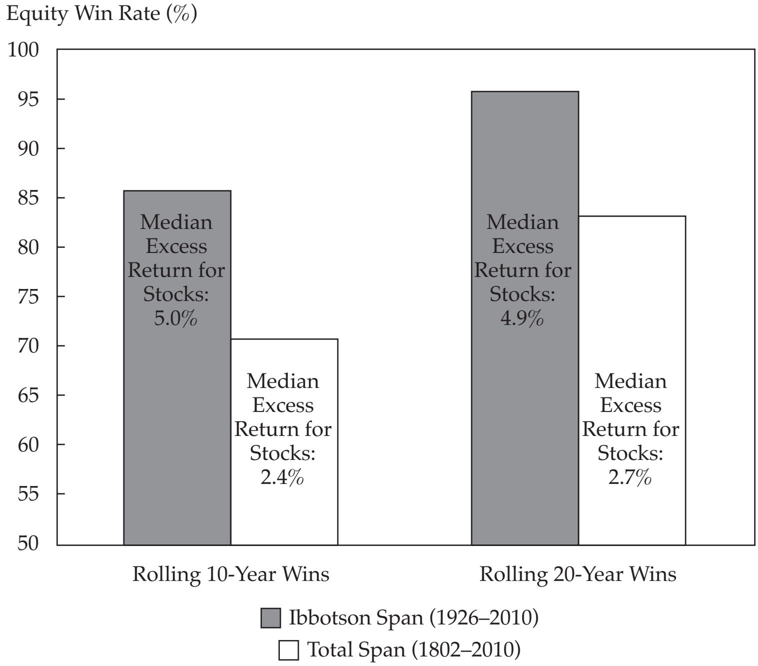
 ©2011 The Research Foundation of CFA Institute

stock market yields followed a less relentless downward course—the cumulative excess return was only 0.66 percent per year. Indeed, investors who incurred the ups and downs over the past 10 years have lost money compared with what they could have earned from long-term government bonds. They have paid for the privilege of incurring stomach-churning risk. Not only did T-bond investors sleep better and more over the past 10 years than stock investors, but they also ate better.

Although recent years have been far from normal, a 30-year stock market excess return of approximately zero is a slap in the face for the legions of "stocks at any price" long-term investors. Yet, it is not the first extended drought. From 1803 to 1857, U.S. equities struggled; the stock investor would have received a third of the ending wealth of the bond investor. For the 1803 investor in U.S. stocks, the shortfall against the bond investor was only recovered in 1871. These early U.S. stock market return data are of dubious quality, but the better U.K. data show a similar trajectory. Most observers would be shocked to learn a 68-year stretch of stock market underperformance occurred in either country. After a 72-year run from 1857 through 1929, when stocks outperformed handily in both the United States and the United Kingdom, another dry spell ensued. From 1929 through 1949, U.S. stocks failed to match bonds. It is the only long-term shortfall in the Ibbotson time sample until the 40-year period ending in March 2009. Perhaps the spectacular 1950–99 aftermath of the extraordinary period of history comprising the Great Depression and World War II lulled recent investors into a false sense of security regarding extended equity performance.

The Odds. Fortunately for the capital markets and equity investors, an examination of history shows that stocks have a high tendency to outperform government bonds over 10- and 20-year periods. **Figure 1** illustrates rolling 10- and 20-year "win rates" for equities versus government bonds for Ibbotson data and data for the whole 1802–2010 period. The Ibbotson time frame confirms investor behavior in the 30 years since Ibbotson and Sinquefield published their groundbreaking study (1977). For the vast majority of periods—92 percent for 10 years and almost 98 percent for 20 years—equities outperformed bonds. The solid consistency goes hand-in-hand with a large average excess return; stocks beat long government bonds by 4.6 percent per year over this span. But the longer-term data are much less convincing than the Ibbotson data. Equities outperformed in 70 percent of the 10-year periods and 84 percent of the 20-year spans, which is wholly consistent with the smaller 2.7 percent risk premium earned by stocks over long bonds during this much longer two-century span. Similar data for other countries indicate that the advantage of equities is even less reliable there than in the United States.

Figure 1. Percentage of Time U.S. Stocks Have Outperformed Long-Term U.S. Government Bonds over Monthly Rolling Periods



Odds are still with the equity investor. Odds of 70 percent or 80 percent are pretty good. In professional basketball, those odds would be average to above-average free throw percentages. But the relatively small probability of failure masks the magnitude of a miss. Just as a single missed free throw can cost a basketball championship, so too can an equity "miss" lead to drastic consequences, as the past 10 years have shown. Superior equity returns are not guaranteed, so why does our industry act as if they are? More importantly, why do investors take all that risk for a skinny equity premium?

We at Research Affiliates do not expect bonds to beat stocks over the next 10 or 20 years. I offer this brief history lesson to illuminate the fact that the much vaunted 4–5 percent risk premium for holding stocks is unreliable and a dangerous assumption to rely on for future plans. In our view, a more reasonable assumption would be 2–3 percent, which reflects history excluding the rise in valuation multiples of the past 30 years. A consideration of today's low starting yields, the prospective challenges from our addiction to debt-financed consumption, and headwinds from demographics would put the ERP closer to 1 percent.

©2011 The Research Foundation of CFA Institute

To act as if the past 200 years were fully representative of the future would be foolish. For one thing, the United States was an emerging market for much of that period, with only a handful of industries and an unstable currency. In the past century, we dodged challenges and difficulties that laid waste to the plans of investors in many countries. Taleb (2007) has pointed out that black swans—unwelcome outliers that spring up well beyond the bounds of normalcy—are a recurring phenomenon; the abnormal is, indeed, normal. U.S. stock market history is but a single sample of a large and unknowable population of potential outcomes.

Peter Bernstein relentlessly reminded us that there are things we can never know, that prosperity and investing success are inherently "risky" and can disappear in a flash. Uncertainty is always with us; the old adage puts it succinctly: If you want God to laugh, tell him your plans. Concentrating the majority of one's investment portfolio in one investment category on the basis of an unknowable and fickle long-term equity premium is a dangerous game of "probability chicken."

The Unchanging ERP

An enduring myth is the notion that the ERP should be static across time and across assets. Why, however, should British Petroleum, struggling to recover from the largest oil spill in history, command the same risk premium as Apple, enjoying acclaim for a product line that serves the appetites of the consumer market with remarkable prescience? BP seems to be riskier than Apple. Should it not command a higher risk premium (and, therefore, a lower price)? Why should the broad stock market command the same risk premium when it is gripped by fear of the apocalypse in the financial services community (as in early 2009) as when optimism is being fueled by a booming economy and a startling surge in technological innovation (as in early 2000)? The year 2009 felt riskier than 2000. So, should stocks have broadly commanded a higher risk premium (and, therefore, a lower price) in 2009 than in 2000? Intuitively, the ERP should obviously vary both across time and across assets.

Many in academia like the simplicity of a fixed risk premium. Simplicity is a good thing, but recall that Einstein was fond of saying, "Make everything as simple as possible, but not simpler." A fixed risk premium is a hypothesis, not a fact; indeed, it is one of the least defensible hypotheses in the finance world today. There is no reason to assume a static risk premium. Nothing in neoclassical finance theory requires a static risk premium, and behavioral finance essentially insists on a risk premium that varies over time and across

©2011 The Research Foundation of CFA Institute

assets. Indeed, recent developments in neoclassical finance theory have focused on time-varying and cross-sectional differences in risk premiums.[3]

A question that emerges from these recent developments in neoclassical finance is: What's the difference between an inefficient market and a market in which the risk premium varies both cross-sectionally and across time? Would it not be easier to simply dispense with the efficient market hypothesis and recognize that price equals an invisible fair value plus or minus a mean-reverting error? Siegel (2006) and Hirshleifer, Glazer, and Hirshleifer (2005) have both likened the debate about this question to the slow acceptance of Copernican cosmology in preference to the bizarre epicycles that were needed to defend Aristotle for more than 1,500 years. Without Copernicus, people could explain the movement of the planets with considerable precision, but because the basic pre-Copernican theory was wrong, no one could figure out why. With Copernicus, Newton was able to answer "why."

The notion that fair value equals price deprives fair value of any independent meaning. Moreover, this notion deprives the academic, empirical, and practitioner communities of a rich opportunity to consider the mathematics and the practical implications of a world in which price and value differ.

The ERP Puzzle: Less Puzzling Than We Might Think

Academia has been abuzz for most of three decades about the ERP Puzzle: Stocks have delivered premium returns relative to bonds or cash that are outsized relative to the return premium that would, in theory, suffice to justify the incremental risk. Although much of macroeconomics points toward a *rational* ERP (for stocks relative to bonds) measured in 10ths of a percent, observed excess returns over long spans have often been 5 percent or more. Until recently.

An observed excess return of 5 percent is not the same thing as an *ex ante* expectation for a 5 percent ERP. For example, if stock market valuation multiples soar, adding a large unexpected increment to returns, excess returns can soundly exceed the *ex ante* ERP. But the opposite can happen just as readily. Indeed, the opposite was the nature of the past decade: Stock market yields nearly doubled as bond yields tumbled, fueling both the bleak stock market returns and the robust real returns for bonds. Yet, despite stocks delivering 700 bps less than long-term

[3]The capital asset pricing model allowed for cross-sectional differences in expected returns, but these returns were driven solely by beta. Many extra dimensions seem to be necessary to fit the data; Fama and French (1992, 1993) explored the joint influence of size and valuation, but a myriad of other dimensions have appeared in recent years. Campbell and Shiller (1988) opened the door in the 1980s for time-varying stock market returns; this approach was subsequently extended by Fama and French (1988). Theoretical explanations were explored by Campbell and Cochrane (1999). Finally, Cochrane's (2011) presidential address to the American Finance Association focuses specifically on the whole issue of time-varying and cross-sectional variation in risk premiums.

 ©2011 The Research Foundation of CFA Institute

T-bonds, no reframing of the ERP Puzzle has occurred; there has been no questioning of why the recent risk premium is far *lower* than finance theory would suggest. Evidently, for many observers a history supported by soaring valuation multiples (and plunging dividend yields) is fair game for bolstering the forward-looking ERP, while a plunge in valuation multiples (and a huge jump in dividend yields) should be ignored in setting that same forward-looking ERP.

If the historical norm for the *expectational* ERP has been roughly half as large as the observed excess return from that rather special span of 1926–2000, the ERP Puzzle remains unsolved, but it is a bit less puzzling. If 100 people are polled on their appetite for equity market risk (I have done this informally many times), almost everyone will be found to eagerly embrace equity market risk if they truly believe that they will earn a 5 percent excess return over bonds, on a long-term compounded basis. That appetite diminishes with a shrinking ERP. The breakeven point, where half of the 100 people will choose *not* to hold an equity-centric portfolio, tends to center on roughly a 2 percent gap or a little more. That percentage point difference is the same ERP that Bernstein and I identified as the historical "normal" ERP in our 2002 article. Hardly anyone will want an equity-centric portfolio if they truly believe that they will garner only 1 percentage point more than long bonds or TIPS.

In our polling experiments, I venture to state that we would find almost no "votes" for accepting equity risk for the few 10ths of a percent incremental return for stocks that finance theory would justify. No one wants 15 percent annual volatility (compounding to about 50 percent total volatility over a 10-year span) if the expected annual return for all the risk is only about 0.5 percent more than the return for bonds.[4]

If market inefficiencies are firmly rooted in behavioral finance, it is easier to close a 2 percent gap than a 4 percent or 5 percent gap. The ERP Puzzle is considerably less puzzling.

Stocks for the Long Run? Yes, but How Long?

For most people, "slender" is an attractive goal.[5] For investors, however, a slender return or a slender risk premium is not at all attractive. For those seeking investments that are priced to offer material benefits to compensate for risk—a solid risk premium—bigger is better.

Few serious observers of the capital markets would argue that the future risk premium for stocks relative to bonds can rival the lofty excess return that stocks have delivered in the past. In the 85 years covered by the Ibbotson data, stocks delivered a real return of 6.6 percent, against 2.1 percent for bonds.

[4]By "total volatility," I mean 10-year (not annualized) lognormal volatility.

[5]This section is excerpted and amended from Arnott (2004).

Terrific! But a big part of this return is attributable to the past increase in the value that the market attaches to each dollar of earnings or dividends. Most observers would think subtracting expansion in the valuation multiple would be reasonable when framing future return expectations.

Using the growth of $100 over time, **Figure 2** breaks the total return on equities into its constituent parts.[6] Panel A does so for the 209 years from 1802 to 2010, and Panel B does so for the 85-year span covered by the Ibbotson data.

For the 209-year time span, the total return is 7.9 percent and the breakdown is as follows:

- *4.9 percent from dividends*. Suppose an investor received only the dividend yield, with no price appreciation, no growth in dividends, and no inflation contributing to price and dividend growth. Then, the investor's $100 would be worth $2.1 million in 2010. Pretty good.
- *1.5 percent from inflation*. Suppose an investor participated only in the part of the capital gain that came from inflation—no income, no growth in income, and no rising valuation multiples. This investor's $100 would have grown to $2,200 by 2010: The cost of living has risen 22-fold, according to U.S. Consumer Price Index statistics. Of course, the $2,200 would buy only what $100 would have bought in 1802 (by definition of "inflation").
- *0.8 percent from real growth in dividends*. Suppose an investor gave away his or her income, experienced no inflation, and did not participate in rising valuation levels but did participate in the real growth in the dividends from stocks. This investor would now have $552—after many more than 200 years. That amount is far less than most people would have expected.
- *0.5 percent from rising valuation multiples (hence, falling yields)*. Suppose an investor received no income, saw no growth, and suffered no inflation but did have assets rise with the rise in equity valuation levels. This investor would have had $100 grow to $265 because dividend yields fell to 35 percent of their 1802 levels [or, viewed in terms of valuation multiples, price-to-dividend ratios (P/Ds) rose to nearly three times the 1802 levels]. P/Es saw a similar increase.
- *0.2 percent from compounding of the multiple sources of return*.

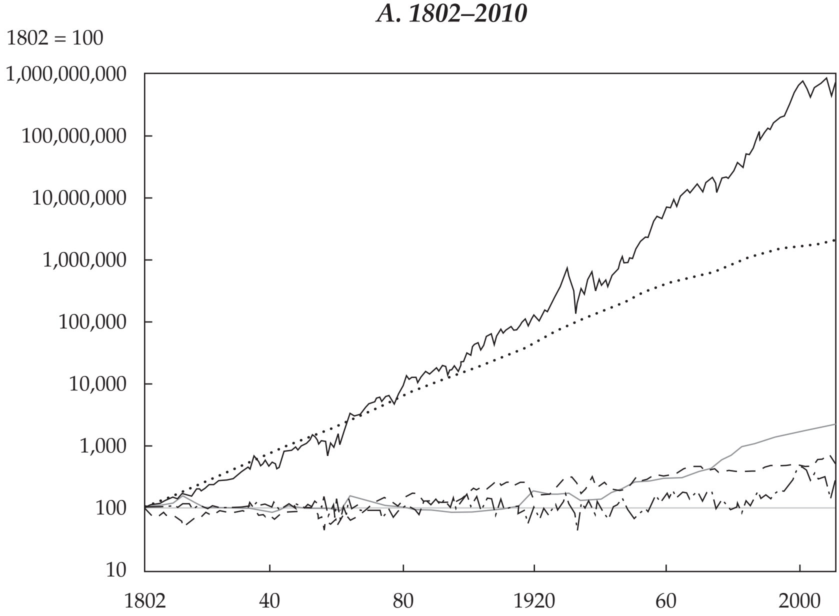
The total return from equities for 1926–2010 is 9.9 percent, and the breakdown is similar to that in Panel A:

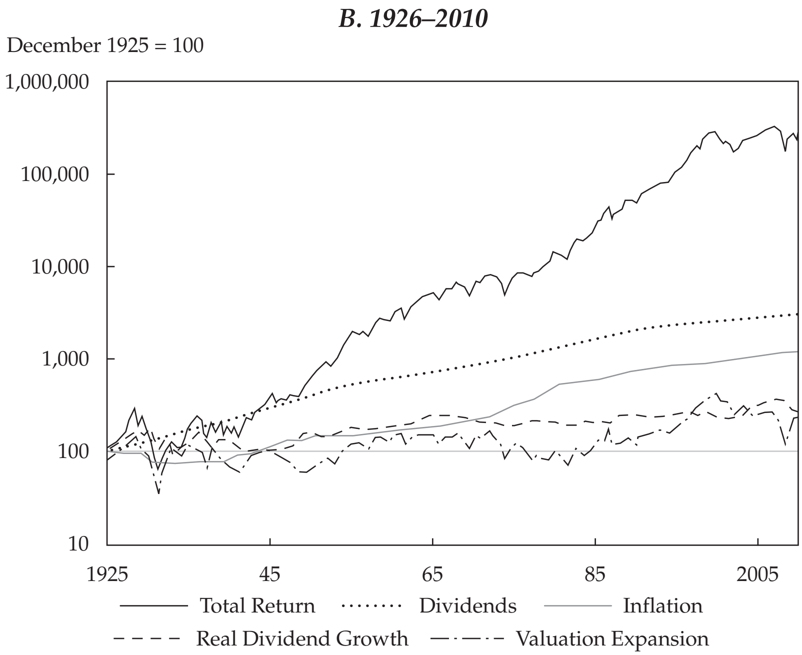
- *4.1 percent from dividends*.
- *3.0 percent from inflation*.

[6]Figure 2 updates Arnott (2003).

©2011 The Research Foundation of CFA Institute

Figure 2. Attribution of Stock Market Returns
(lognormal scale)





Source: Based on data from CRSP, Morningstar (Ibbotson), Robert Shiller, and William Schwert.

- *1.3 percent from real growth in dividends.*
- *1.1 percent from rising valuation levels.*
- *0.4 percent from compounding.*

For the full 209-year span starting in 1802, the 7.9 percent total return for stocks compares with 5.1 percent for long-term government bonds, giving us a 209-year excess return of 2.7 percent (net of compounding). Over the 85-year Ibbotson span, the long-term bond return is 5.2 percent and stock market excess return is 4.4 percent (again, net of compounding). If we take out the historical rise in valuation level—0.5 percent and 1.1 percent, respectively—these excess returns shrink to 2.2 percent for the longer period and 3.3 percent for the 85-year span.

Details of the impact of a "new normal" (in which GDP growth is impeded by the triple threat of deficits, debt, and demographics) on the ERP are beyond the scope of this paper. I would like to observe, however, that as people live longer and work longer, they have more time to accumulate wealth in anticipation of retirement. This phenomenon should lead investors to accept lower forward-looking stock and bond market returns and a lower risk premium for stocks. This phenomenon may be the cause of Japan's low current yield for both stocks and bonds and the steady erosion in these yields in the United States. It may also help explain investors' tolerance of low sovereign yields—even in the face of steadily escalating debt burdens and escalating fears of eventual defaults. Apparently, the risk premium should be lower than the historical 2–3 percent excess return, and a lower risk premium is wholly consonant with lower long-term return expectations for both stocks and bonds.

Let's explore the consequences of a slender risk premium. If stocks always offered a 5 percent risk premium relative to bonds, then no long-term investor would diversify away from stocks. The arithmetic is compelling. If stocks normally delivered better returns than bonds by 5 percent per year compounded over time, the long-term investor would have almost a 95 percent chance of winning with stocks by the end of a 20-year span. The cult of equities and the notion of stocks for the long run are predicated on such a lofty risk premium. If the risk premium is smaller, then the arithmetic quickly becomes drastically less interesting: If the risk premium falls by half, the time required to have high confidence of winning with stocks quadruples. The arithmetic is simple but powerful.[7]

Consider a disaster scenario for an investor—the 5th percentile outcome. **Figure 3** shows the 5th percentile relative wealth outcome for various risk premiums over time. In Panel A, if the difference in returns between stocks and

[7]I am indebted to André Perold for pointing out that if the risk premium falls by half, the time required to have high confidence of winning with stocks quadruples.

 ©2011 The Research Foundation of CFA Institute

bonds is 5 percent and has a volatility of 15 percent, then the 5th percentile outcome is a 19 percent shortfall of stocks relative to bonds after one year.[8] That is, the investor would have a 5 percent chance of stocks underperforming bonds by 19 percent or more in a year. But over two years, the 5th percentile outcome is *not* another loss of 19 percent after the initial loss of 19 percent. Because risk expands with the square root of time, the 5th percentile outcome over two years is 34 percent below the mean. But the mean return has now grown another 5 percent, to a 10 percent gain. Thus, the 5th percentile outcome is a loss of only 24 percent over the two years, barely 5 percent worse than the one-year case.

In fact, if stocks can reasonably be expected to deliver 5 percent more than bonds, the "worst-reasonable" (or 5th percentile) outcome is that the equity investor is underwater relative to bonds by 26 percent after five years and never falls any lower. After five years, the picture becomes brighter. And, after 25 years, the investor has a better than 95 percent chance of winning with stocks, relative to bonds. In a nutshell, this kind of analysis is the basis for recommending stocks for the long run.

Unfortunately, some time periods, including the past decade, delivered far worse outcomes than a mere 26 percent peak-to-trough relative performance drawdown. If long-term bonds yield 4 percent, an investor needs to get a long-term return of 9 percent from stocks to get a 5 percent risk premium. If stocks are yielding 2 percent and if stocks have to return 9 percent, then stocks must deliver long-term earnings and dividend growth of 7 percent above the dividend yield. Such performance is a lot to ask. Annual per share earnings growth in the 20th century (no slacker for growth as centuries go) averaged slightly more than 4 percent, of which fully 3 percent was inflation.

Suppose earnings growth is only 4 percent, or 3 percent, or 2 percent. These growth rates, added to a 2 percent dividend yield, will correspond to a (respective) 6 percent, 5 percent, and 4 percent total return and, therefore, a (respective) 2 percent, 1 percent, and zero risk premium. After 25 years, the 5th percentile bleak outcome has the equity investor, respectively, 50 percent, 60 percent, and 70 percent behind the bond investor and still headed south. This bad news is the 5th percentile outcome, but it is well within the realm of possibility.

With smaller risk premiums, the shortfalls can be larger and it takes longer to recover. For example, Panel B shows that the worst-reasonable outcome for a 2 percent risk premium reaches about a 50 percent shortfall, and the equity investor finally has 95 percent confidence that stocks will beat bonds in 150

[8]The 5th percentile is 1.6 standard deviations below the mean. The standard deviation of 15 percent times 1.6 results in a 5 percent chance of having stocks perform 24 percent below this 5 percent mean outperformance, for a shortfall of 19 percent relative to bonds.

years. This point is also about the time that the worst-reasonable outcome with a 1 percent risk premium hits its low point, at 77 percent less wealth than the bond investor has. At this risk premium, the equity investor is still way behind bonds after 200 years in the 5 percent outcome.

In short, stocks work for the long run if the risk premium is large. But the "normal" risk premium over the past two centuries has been shown to be about 2.4 percent (Arnott and Bernstein 2002) and, if the same technology is used as in the 2002 paper, would be about 1.4 percent today. If the long-term average of 2.4 percent is right, then 100-year investors can expect their stocks to beat their bonds with 95 percent confidence. If the current risk premium is lower than 2.4 percent, the investor will need a longer horizon to have this much confidence in the superiority of the stock holdings.

Naturally, if the investor is willing to settle for a 60 percent likelihood of success, the span needed to wait for success is considerably shorter. But the myth is that a reasonable span for patient investors is all that is needed for stocks to assuredly outpace bonds. This myth is simply untrue unless stocks are priced to deliver a large risk premium relative to bonds.

The Myth of Buybacks

The bull market of the 1990s was built largely on a foundation of two immense misconceptions.[9] Investors were told the following:

1. With the coming of the technology revolution and a "new paradigm" of low payout ratios and internal reinvestment, earnings will grow faster than ever before. Real growth of 5 percent will be easy to achieve.[10]
2. When earnings are not distributed as dividends and not reinvested into stellar growth opportunities, they are distributed back to shareholders in the form of stock buybacks, which are a vastly preferable way of distributing company resources to the shareholders from a tax perspective.[11]

The vast majority of the institutional investing community has believed these untruths and has acted accordingly. Whether these myths are lies or merely errors, they are serious and demand scrutiny. Let's examine reinvestment first.

[9]This section is excerpted and amended from Bernstein and Arnott (2003).

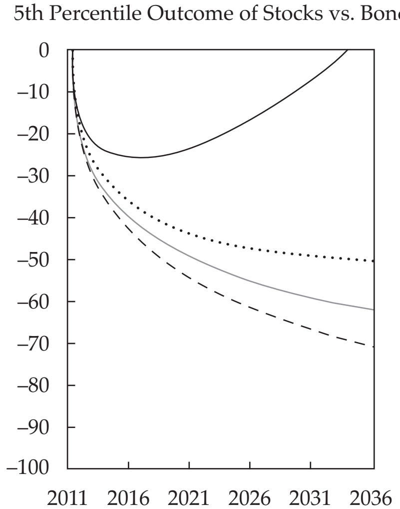
[10]Like the myth of Santa Claus, this story is highly agreeable but is supported by neither observable current evidence nor history. Asness and I debunked this idea in a 2003 article (Arnott and Asness 2003). The work of Miller and Modigliani (1961) is often used as theoretical justification for this claim, although their capital equivalence theorem makes a typical array of simplifying assumptions (market efficiency, no taxes, free trading, etc.) not found in the real world. Furthermore, their work applies cross-sectionally.

[11]Bernstein and I demonstrated that stock repurchases rarely exceed new share issuance. The norm appears to be a "Two Percent Dilution" (Bernstein and Arnott 2003).

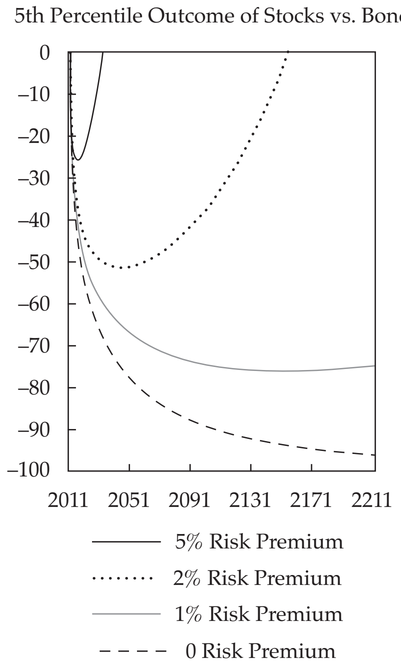
 ©2011 The Research Foundation of CFA Institute

Figure 3. The Arithmetic of Long-Term Returns in the United States: 5th Percentile Relative Wealth Outcomes vs. Equity Risk Premiums

A. Twenty-Five Years



B. Two Hundred Years

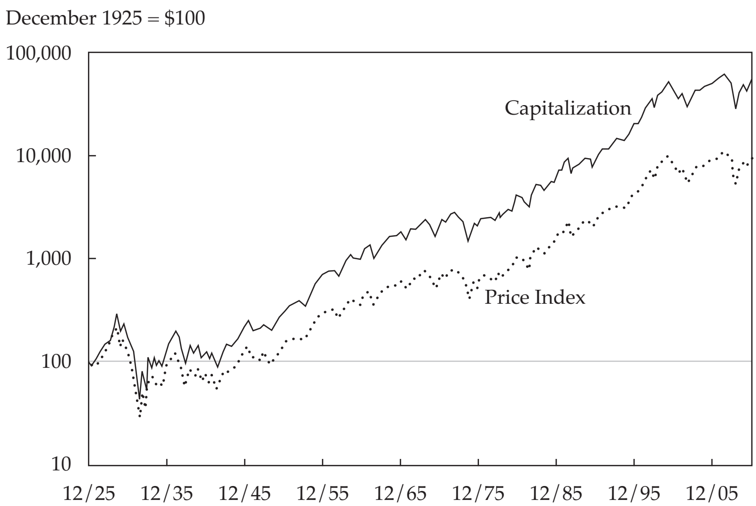


I would not dispute the attractions of stock buybacks. They are a tax-advantaged way to provide a return on shareholder capital, particularly when compared with dividends, which are taxed twice. Buybacks have enormous appeal. Contrary to popular belief, however, apart from brief spans in the 1980s and the latest decade, they have not occurred to any meaningful degree in the past 85 years.

I suggest a simple measure of net new issuance—namely, the ratio of the proportionate increase in market capitalization to the proportionate increase in price. For example, if over a given period the market cap increased by a factor of 10 and the cap-weighted price index increased by a factor of 5, then 100 percent net share issuance has taken place in the interim.

This relationship has the advantage of factoring out valuation changes and splits because they are embedded in both the numerator and denominator. Furthermore, it holds only for universal market indices, such as the CRSP Cap-Based Portfolio indices 1–10, because less inclusive indices can vary the above ratio simply by adding or dropping securities. **Figure 4** shows the growth of \$100.00 in total market cap and in the price of the CRSP 1–10. Note that even the CRSP data can involve adding securities: CRSP added the American Stock Exchange in 1962 and NASDAQ stocks in 1972.

Figure 4. Growth of U.S. Stock Prices and Capitalization, 1926–2010
(lognormal scale)



Source: Based on CRSP data.

©2011 The Research Foundation of CFA Institute

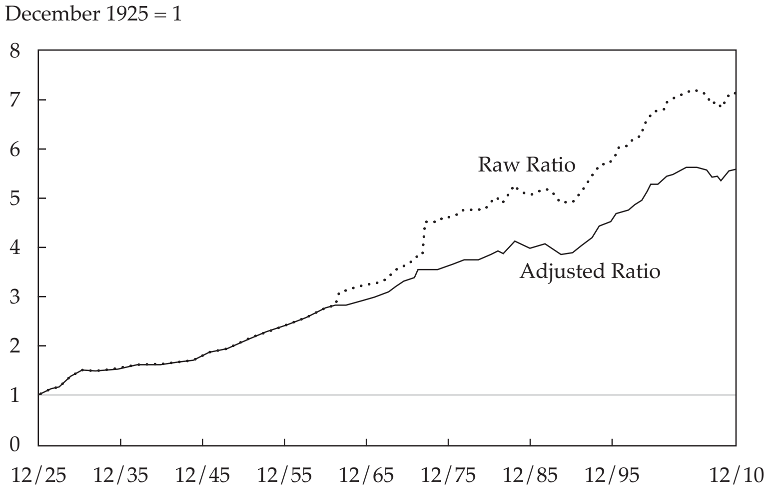
An initial public offering (IPO) or a secondary equity offering (SEO) dilutes investors in the broad index. A buyback that reduces a company's outstanding float increases existing shareholders' ownership of the company. A buyback that merely offsets management stock option redemption—a common so-called buyback—is a wash; it does not change the float, so it is not a true buyback.

Note in Figure 4 how market cap slowly and gradually pulls away from market price. The gap does not look large in this figure, but by the end of 2010, the U.S. market cap index had grown 567-fold whereas the price index had grown only 101-fold. The reason for this discrepancy is simple: 82 percent of today's stock market consists of businesses that did not exist in 1925. For every share of stock extant in 1926, there are now 5.65 shares. These data imply net new share issuance at an annualized rate of slightly more than 2 percent per year.

To give a better idea of how this phenomenon has proceeded over the past 85 years, **Figure 5** shows a plot of a dilution index, defined as the ratio of capitalization growth to price index growth. (The adjustment for the stock additions of 1962 and 1972 is evident in Figure 5, where the dilution ratio was held constant for the two months during which the shifts took place.) Figure 5 traces the growth in the ratio of (1) the total capitalization of the CRSP 1–10 to (2) the market value–weighted price appreciation of these same stocks. The fact that this line rises nearly monotonically shows clearly that new share issuance almost always sharply exceeds stock buybacks. The notable exceptions are in the late 1980s, when buybacks outstripped new share issuance, and in the mid-2000s, when a flurry of demand from shareholders for buybacks occurred. That stock buybacks were an important force in the 1990s is simply a myth. The *belief* that stock buybacks were happening at an unprecedented pace may have been an important force, however, in the bull market of the 1990s.

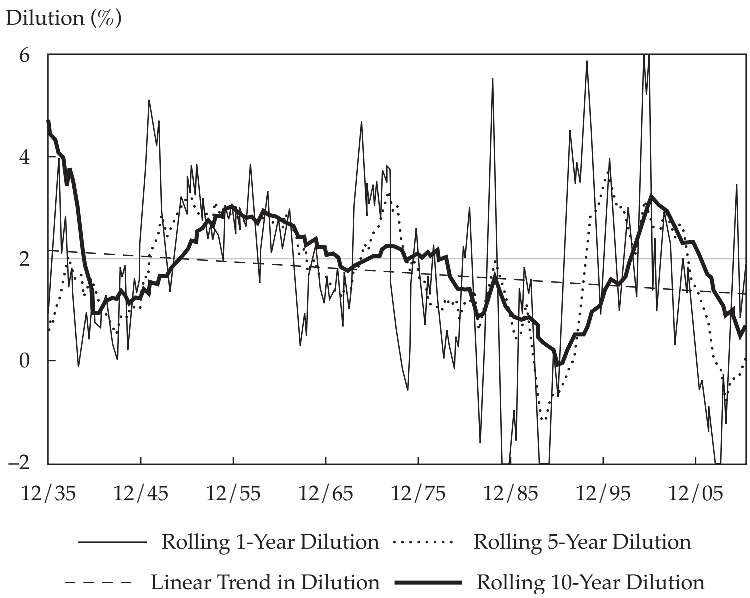
Figure 6 shows the rolling 1-, 5-, and 10-year growth in the aggregate supply of equity capital; hence, dilution of an index affects investors' ownership of the market portfolio. Keep in mind that every 1 percent rise in equity capital is a 1 percent rise in market capitalization in which existing shareholders did not (and could not) participate. Except for the 1980s, the supply growth was essentially never negative even on a 1-year basis. How the myth of stock buybacks gained traction after the 1980s is clear; it was such a pervasive pattern in those years that even the 10-year average rate of dilution briefly dipped negative. But then, during the late 1990s, stock buybacks were outstripped by new-share issuance at a pace that was exceeded only in the IPO binge of 1926–1930. This surge in the supply of new stock is evident whether we are looking at net new-share issuance on a 1-, 5-, or 10-year basis. A recent, 2005–2007, spate of buybacks brought back the illusion that stock buybacks are a normal means by which management rewards shareholders in a tax-advantaged fashion.

Figure 5. CRSP U.S. Market Capitalization/Price, 1926–2010



Source: Based on CRSP data.

Figure 6. Annualized Rate of Shareholder Dilution in the United States, 1935–2010



Source: Based on CRSP data.

©2011 The Research Foundation of CFA Institute

Those who argue that stock buybacks will allow future earnings growth to exceed GDP growth can draw scant support from history. Could buybacks be large enough to be an important complement to dividends as a means of rewarding shareholders? Of course. Enormous earnings growth, far faster than real economic growth, did occur from 1990 to 2000. But much of this earnings growth was dissipated through shareholder dilution in the form of IPOs and SEOs.

Expected stock returns would be highly agreeable if dividend growth, and thus price growth, proceeded at the same rate as aggregate economic growth, or better. Unfortunately, this growth does not occur: Comparing the Dimson, Marsh, and Staunton (2002) 20th century dividend growth series with aggregate U.S. GDP growth, we find that even in nations that were not savaged by the century's tragedies, dividends grew, on average, 2.3 percent more slowly than the GDP. Similarly, by measuring the gap between the growth of market capitalization and share prices in the CRSP database, we find that between 1926 and the present, a 2.3 percent net annual dilution occurred in the outstanding number of shares in the United States.

Thus, two independent analytical methods point to the same conclusion: In stable nations, net annual creation of new shares is roughly 2 percent, which is the "2 percent dilution" that separates long-term economic growth from long-term per share dividend, earnings, and share price growth.

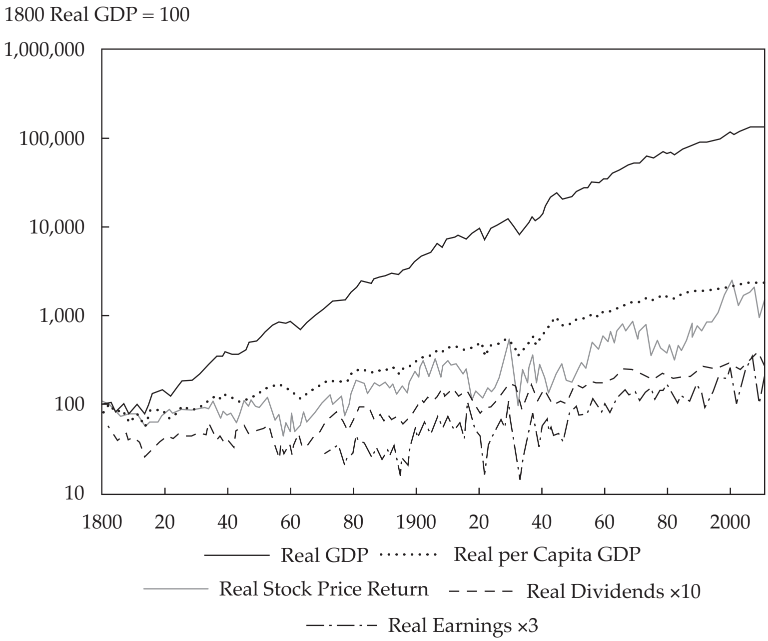
The Mythical Link of GDP Growth and Earnings Growth

Over the past two centuries, common stocks have provided a sizable excess return to U.S. investors: For the 200 years from 1802 through 2001, the returns for stocks, bonds, and bills were, respectively, 7.9 percent, 5.2 percent, and 4.2 percent. In the simplest terms, the reason is obvious: A bill or a bond is simply a promise to pay interest and principal, and as such, its upside is sharply limited. Shares of common stock, however, are a claim on the future dividend stream of the nation's businesses. The ever increasing fruits of innovation-driven economic growth accrue only to the shareholder, not the bondholder.

Viewed over the decades, this powerful economic engine produces remarkably even growth. **Figure 7** plots the real GDP of the United States since 1800. The economy, as measured by real GDP, has grown 1,300-fold since 1800, averaging about 3.5 percent per year. The long-term uniformity of economic growth is both a blessing and a curse. It is reassuring to know that real U.S. GDP has doubled every 20-odd years, partly on the basis of a rapidly growing population. But the data are also a dire warning to those predicting rapid acceleration of economic growth from the computer and internet revolutions. Such

©2011 The Research Foundation of CFA Institute

Figure 7. Growth in U.S. Real GDP, Real per Capita GDP, Real Stock Price Return, Real Earnings, and Real Dividends (lognormal scale)



Source: Based on data from CRSP, Morningstar (Ibbotson), Robert Shiller, and William Schwert.

extrapolations of technology-driven increased growth are painfully oblivious to the broad sweep of scientific and financial history in which innovation and change are constant; they are neither new to the current generation nor unique. The technological advances of the 1990s register barely a blip on the long-term history shown in Figure 7; the travails of the past decade are far more noticeable.

The impact of recent advances in computer science pales in comparison with the technological explosion that occurred between 1820 and 1855. This earlier era contained the deepest and most far reaching technology-driven changes in everyday existence in human history. These changes profoundly affected the lives of those from the top to the bottom of society in ways that can scarcely be imagined today.

At a stroke, the speed of transportation increased tenfold and communications became almost instantaneous. Until 1820, people, goods, and information could not move faster than the speed of a horse. Within a generation, journeys achieved an order-of-magnitude less time, expense, danger, and

 ©2011 The Research Foundation of CFA Institute

discomfort because of steam, canals, and the railroad; important information that had previously required the same long journeys—taking weeks or months—could be transmitted instantaneously by telegraph.

Put another way, the average inhabitant of 1815 would have found the world of 40 years later incomprehensible, whereas a person transported from 1971 to 2011 would be duly impressed by our technological advances but would have little trouble understanding the intervening changes in everyday life (and would be shocked that we have not revisited the moon in 40 years!). From 1815 to 1855, the U.S. economy grew eightfold, whereas in the past 40 years, it has grown barely 150 percent.[12]

The relatively uniform increase in GDP is matched by a similar uniformity in the growth of corporate profits. A direct relationship has existed between aggregate corporate profits and GDP since 1871, the earliest market earnings data that anyone has assembled for U.S. stocks. Therefore, shouldn't stock prices have grown at the same rate? The problem is that per share earnings and dividends keep up with GDP only if no new shares are created. Unfortunately, entrepreneurial capitalism has a dilution effect; it creates new enterprises and new stock in existing enterprises so that *per share* earnings and dividends grow considerably more slowly than the economy, as Figure 7 shows.

In fact, as Figure 7 shows, since 1871, real stock prices have grown at 1.8 percent per year, versus 3.4 percent for real GDP. Furthermore, the true degree of "slippage" is much higher because one-third of the rise in real stock prices after 1871 was the result of a substantial upward revaluation (increase in the P/E or P/D). The highly illiquid industrial stocks of the post–Civil War period rarely sold at much more than 10 times earnings and often sold for multiples of only 3 to 4 times earnings. Those stocks gave way to the instantly and cheaply tradable common shares, priced many times more dearly, that we see today.

Note also in Figure 7 that real per share prices, earnings, and dividends grew at a pace similar to that of per capita GDP (with some slippage associated with the "entrepreneurial" stock rewards to management). Indeed, since 1871, these growth rates have been 1.8 percent for real per share prices, 1.4 percent for earnings, 1.1 percent for dividends, and 1.9 percent for GDP. Why should these rates be so tightly linked? Per capita GDP is a measure of productivity (with slight differences for changes in the workforce, hours worked, and so forth). And aggregate GDP per capita must grow in reasonably close alignment with productivity growth. Productivity growth is also the key driver for per capita income growth and for per share earnings and dividends. Accordingly,

[12]Of course, much of the growth in earlier GDP was driven by population growth, especially in the 1815–55 span. Still, per capita real GDP doubled in 1815–1855 but rose only by slightly more than 60 percent in the past 40 years.

any difference in the growth rates of GDP and the other three measures will mean that capital is deriving outsized benefits from productivity growth relative to labor (and vice versa). If share prices, earnings, and dividends grow faster than productivity, return on labor migrates to return on capital; if slower by a margin larger than the value of stock awards to management, then the economy is migrating from rewarding capital to rewarding labor. Either way, such a change in the orientation of the economy cannot continue indefinitely. The migration of returns to capital is corrected by a labor backlash; the migration of returns to labor by a flight of capital.

This observation has sobering implications at a time when corporate profits are near an all-time record high share of GDP and wages are near an all-time low share, as was the case in 2007 and again in 2011. Any student of market history will see that mean reversion is a powerful force in the interplay between these measures.

Is the United States unique? In their book *Triumph of the Optimists,* Dimson, Marsh, and Staunton (2002) tracked stock, bond, and cash returns over the previous century in 16 countries. I compared dividend growth, price growth, and total return with data on GDP growth and per capita GDP growth for the 16 countries covered by Dimson et al. (2002) in the 20th century. The GDP data come from Maddison's (2001) world GDP survey for 1900–1998 and the International Finance Corporation for 1998–2000. For the average country, there is a startling gap of 3.3 percent between dividend growth and the growth rate of aggregate GDP. For per capita GDP growth, there is still a 2.4 percent annual shortfall between dividend growth and per capita GDP growth. In the 2010 update of the Dimson et al. study, the results changed little.

The 20th century was not without turmoil. In our 2003 study, Bernstein and I divided 16 nations (see Bernstein and Arnott 2003) into two categories according to the degree of devastation visited upon them by the era's calamities. One group included countries that suffered substantial destruction of their productive physical capital at least once during the century; the other group did not. The nine nations in the first group were devastated in one or both of the world wars or by civil war. The remaining seven suffered relatively little direct damage.

For the nations that were devastated during the world wars or revolutions, the good news is that their economies repaired the devastations by the end of the 20th century. They enjoyed overall GDP growth and per capita GDP growth that rivaled the growth of the less scarred nations. The bad news is that the same cannot be said for per share equity performance. A slippage of 4.1 percentage points occurred between the annual growth rates of their economies and per share corporate payouts.

 ©2011 The Research Foundation of CFA Institute

In the fortunate group—those untroubled by war, political instability, and government confiscation of wealth—we nevertheless found, on average, dividend growth 2.3 percentage points less than GDP growth and 1.1 percentage points less than per capita GDP growth. These results are similar to the 2.7 percent and 1.4 percent figures observed in the United States during the 20th century.

Why Does the Finance Industry Think Dividends Don't Matter?

Two misconceptions about the ERP that I stated in the opening are linked to the prevailing view that dividends aren't especially important. Respected academics have suggested the following:

1. If dividend yields are below historical norms, the market is clearly expecting faster future growth. (With this circular logic, we might as well buy at any valuation multiple because our buying creates still higher multiples and the resulting lower yields will imply even faster future growth.)
2. If payout ratios are below historical norms, the retained earnings will be reinvested in projects that will lead to faster future growth. (M&M are thus invoked. If that shortcut is sound, why not encourage management to retain all of the earnings? After all, the massive technological investments between 1998 and 2001, which were funded out of retained earnings, certainly must have led to a major step-up in subsequent earnings growth rates.)

A careful examination of the data provides no support for this intertemporal interpretation of M&M. Miller and Modigliani (1961) developed a brilliant thesis proving that dividend policy and structural debt/equity decisions do not matter so long as investors are rational, markets are efficient, there are no taxes, management operates in the best interests of the shareholders, bankruptcy costs are ignored, and so forth. These arguments seem to be tacitly based on the notion that because our "best" finance models (those that most accurately explain and predict phenomena) rely on certain assumptions, the assumptions must also be right. Even the best finance theories and models, however, rely on assumptions that are deliberate simplifications of the real world. Accordingly, even M&M's assumptions must be considered approximations of the real world.[13]

[13] Paul Samuelson said much the same: "Only the smallest fraction of economic writings, theoretical and applied, has been concerned with the derivation of *operationally meaningful* theorems. In part at least, this has been the result of the bad preconception that economic laws deduced from *a priori* assumptions possessed rigor and validity independently of any empirical human behavior. But only a very few economists have gone so far as this. The majority would have been glad to enunciate meaningful theorems if any had occurred to them." (Samuelson 1947, p. 3) [Italics in the original.]

When we approach the models, we can rely on common sense. Because the models are based on certain assumptions, we can examine the validity of those assumptions before we accept the dictates of the models as "truth."

Bond yields are accepted as the dominant factor in setting bond return expectations, but dividend yields (and, often, even earnings yields) are seen as secondary to growth in setting equity return expectations. Yet, overwhelming global evidence suggests a strong positive link between the dividend yield and both the subsequent real return for stocks and the subsequent excess return of stocks over bonds. It is a myth that in an efficient market investors will accept a lower yield whenever they are confident that future real growth in earnings will make up the difference. It is a myth that in an efficient market investors will not care about payout ratios because retained earnings make up for the deferred income in the form of more rapid growth; that is, lower dividends now mean higher ones later. These enduring myths lead to complacency about the ERP.

Conclusion: Why These Enduring Myths?

Why do we so readily accept forecasts based on extrapolating the past? If bond yields fall from 8 percent to 4 percent, and the bonds thereby deliver a 12 percent annualized return (including capital gains), should we assume 12 percent as a future bond return? Of course not! The capital gains that pushed the 8 percent yield up to a 12 percent return are nonrecurring. Should we "conservatively" assume a bit less than the historical 12 percent return—say, 10 percent—in recognition that yields are down? Of course not; the yield is 4 percent! So, the expected return is also 4 percent. Yet, much of our industry, with an assist from assorted academic luminaries, is wedded to forecasting equity returns by extrapolating past returns.

Returns are, for the most part, a function of simple arithmetic. For almost any investment, the total return consists of yield, growth, and multiple expansion or yield change. For bonds, the growth is simple: Fixed income implies zero growth. For high-yield or emerging market debt, growth is negative because of the occasional defaults. For stocks, based on a long history, growth tends to be around 1 percentage point above inflation.

The 7 percent real stock market returns of the past 78 years consist of roughly 4.3 percent from dividend yield, slightly more than 1 percent from real dividend growth, and 1.5 percent from multiple expansions. We cannot expect 7 percent in the future because we cannot rely on expansion of the multiple. Most observers would, at a minimum, subtract multiple expansions from future return expectations. Now, the return is down to about 5.5 percent. The current dividend yield, however, is only 1.6 percent, not 4.3 percent, which takes the real return down to around 2.5 percent to 3 percent. And that is without any

 ©2011 The Research Foundation of CFA Institute

"mean reversion" toward historical valuation levels. Much of our industry seems to prefer forecasting the future by extrapolating the past, however, because doing so produces a higher number.

Why is a low (even negative) risk premium considered shocking? Nothing assures a positive risk premium. Only finance theory (with numerous assumptions) suggests that this situation is not possible. But finance theory also posits that rational investors shun lotteries and casinos. Outside of finance theory, a temporary negative risk premium should be possible.

Should equity provide a positive risk premium relative to bonds? Of course. Is it written into contract law for any assets we buy? Of course not. In the long run, the market must adjust to provide a positive expected risk premium. But the adjustment to a positive rationally expected risk premium may be painful. A 5 percent risk premium is often taken as fact, but it is only a hypothesis and, many times, an ill-reasoned one.

Even the most aggressive, intellectually honest forecasts of long-term earnings or dividend growth assume GDP growth as an upper bound. Growth in the portion of GDP represented by corporate profits comes from the growth of existing enterprises and the creation of new enterprises. Stock market investments allow investors to participate in the former but not the latter. Because more than half of real GDP growth comes from entrepreneurial capitalism, real earnings and dividends should collectively grow a bit under half the rate of economic growth.

Nevertheless, consensus long-term earnings growth estimates routinely exceed sustainable GDP growth. The current consensus growth rate for earnings on the S&P 500, according to the Zacks Investment Research survey, is 10 percent, which, if we assume a consensus inflation expectation of 2–3 percent, corresponds to 7–8 percent real growth. Real earnings growth of 8 percent is six times the real earnings growth of the past century, however, and three times the consensus long-term GDP growth rate. This growth is not possible.

GDP growth, less the economic dilution associated with entrepreneurial capitalism, basically defines sustainable growth in per share earnings and dividends. Accordingly, it is hard to imagine that stocks offer a positive risk premium when they are yielding far less than TIPS. Yet, in December 1999 and January 2000, stock market yields were a scant 1.1 percent whereas the TIPS yield was 4.4 percent. Earnings and dividends on stocks would have needed to grow at 3.3 percent per year (triple the real growth rate of the prior century) for stocks to merely match the total return of TIPS. I believe a negative risk premium (at least for the broad stock market averages relative to TIPS) existed at the beginning of 2000.

©2011 The Research Foundation of CFA Institute

Many market observers would agree that the cult of equities and reliance on a 5 percent ERP were the most damaging errors in the institutional sponsor community in the past quarter century. Shouldn't our industry, as a matter of course, question aggressive, unsustainable growth forecasts before acting on them?

Why do we accept rising return expectations in a rising market? In 1982, at a time when stock yields were 5 percent and both earnings yields and bond yields were in the low teens, the average pension return assumption was barely 6 percent. In 2000, the average pension return assumption had risen to approximately 9.5 percent, even though stock dividend yields and bond yields were down by, respectively, 4 percentage points and 8 percentage points. When markets fell in 2007–2009, we began to see pension return assumptions drifting downward again!

Siegel (2007) recognized that this mean reversion reduces the risk of equities for the long-term investor. A puzzle that he does not acknowledge is that, following the largest equity revaluation in history in 1982–2000, mean reversion might exact consequences in the form of reduction of future returns.

Too often, analysts rely on finance theory as a shortcut to easy answers. We point to M&M to reassure ourselves that 70 percent or even 100 percent earnings retention is fine because the retained earnings are surely being used to fund innovations that will lead to unprecedented future growth. We point to the capital asset pricing model (CAPM) to compute expected rates of return and to assess the alphas of our strategies. But none of these remarkable models and theories fully capture reality. Behavioral finance, the principal rival to the models of neoclassical finance theory, helps us understand how human frailties can create the very market behaviors that classical finance theory seeks to explain away, but behavioral finance does not help us decide how to profitably invest.

Our industry, in both the academic and the practitioner communities, is too complacent. Too many people say, "Assuming this, then we can decide that." Too few are willing to question their basic assumptions. As fiduciaries, we owe it to our clients to be less accepting of dogma and more willing to explore the implications of errors in the root assumptions of finance theory. These basic assumptions often fail when they are tested. Failing assumptions are not bad; indeed, that is where the profit opportunities can be found.

If finance theory assumes that markets are efficient and behavioral finance suggests that markets are not efficient, do we discard the less convenient theory? Isn't it better to recognize elements of truth in seemingly incompatible theories? Economics is not physics. Classical finance and behavioral finance *can* both be partially correct. If we recognize this possibility, we gain a rich understanding of the markets in which we seek our clients' profits and our livelihood.

 ©2011 The Research Foundation of CFA Institute

References

Arnott, Robert D. 2003. "Dividends and the Three Dwarfs." *Financial Analysts Journal*, vol. 59, no. 2 (March/April):4–6.

———. 2004. "The Meaning of a Slender Risk Premium." *Financial Analysts Journal*, vol. 60, no. 2 (March/April):6–8.

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):70–87.

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal*, vol. 58, no. 2 (March/April):64–85.

Bernstein, Peter. 1997. "What Rate of Return Can You Reasonably Expect… or What Can the Long Run Tell Us about the Short Run?" *Financial Analysts Journal*, vol. 53, no. 2 (March/April):20–28.

Bernstein, William J., and Robert D. Arnott. 2003. "Earnings Growth: The Two Percent Dilution." *Financial Analysts Journal*, vol. 59, no. 5 (September/October):47–55.

Campbell, John Y., and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy*, vol. 107, no. 2 (April):205–251.

Campbell, John Y., and Robert J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, vol. 1, no. 3 (Autumn):195–228.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance*, vol. 66, no. 4 (August):1047–1108.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

———. 2010. *Global Investment Returns Yearbook 2010.* Zurich: Credit Suisse Research Institute.

Fama, Eugene F., and Kenneth R. French. 1988. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, vol. 22, no. 1 (October):3–25.

———. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance*, vol. 47, no. 2 (June):427–465.

———. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, vol. 33, no. 1 (February):3–56.

Hirshleifer, Jack, Amihai Glazer, and David Hirshleifer. 2005. *Price Theory and Applications: Decisions, Markets, and Information*. Cambridge, U.K.: Cambridge University Press.

Ibbotson, Roger G., and Rex A. Sinquefield. 1977. *Stocks, Bonds, Bills, and Inflation: The Past (1926–1976) and the Future (1977–2000)*. Charlottesville, VA: The Financial Analysts Research Foundation.

Maddison, Angus. 2001. *The World Economy: A Millennial Perspective.* Washington, DC: Organisation for Economic Co-Operation and Development.

Miller, Merton, and Franco Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *Journal of Business*, vol. 34, no. 4 (October):411–433.

Samuelson, P. 1947. *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.

Siegel, Jeremy J. 2006. "The 'Noisy Market' Hypothesis." *Wall Street Journal* (14 June):A14.

———. 2007. *Stocks for the Long Run: The Definitive Guide to Financial Market Returns & Long Term Investment Strategies*. 4th ed. New York: McGraw-Hill.

Taleb, Nassim. 2007. *The Black Swan: The Impact of the Highly Improbable*. 1st ed. New York: Random House.

 ©2011 The Research Foundation of CFA Institute

Time Variation in the Equity Risk Premium

Antti Ilmanen
Managing Director
AQR Capital Management (Europe) LLP

The equity risk premium (ERP) refers to the (expected; sometimes, realized) return of a broad equity index in excess of some fixed-income alternative. In the past decade, a dramatic shift has occurred in what is considered to be the best source of information about the future ERP: Is it historical average returns or forward-looking valuation indicators?

- Academics and practitioners alike used to think that the ERP is constant over time, in which case the future premium would best be estimated from the long-run average of the realized excess return. If the historical realized outperformance of stocks over bonds was 6 percent, for example, 6 percent would also be the best forecast for the future. Such a rearview-mirror perspective makes the ERP seem especially high at the end of each long bull market, just when market valuation ratios are abnormally high.
- The recent roller-coaster experiences in markets, as well as theoretical and empirical lessons, have converted many observers to the belief that expected returns and premiums vary over time. If so, then past average returns are a highly misleading indicator of future returns. Forward-looking valuation indicators are better and may provide useful timing signals. Low dividend yields or low earnings yields (or their inverse, high price-to-earnings ratios) are now seen as a sign of low prospective stock market returns in just the same way that low bond yields and narrow yield spreads are interpreted as a forecast of low returns in fixed-income markets. This forward-looking logic would have guided investors well during the low equity market yields of 2000 and high market yields of early 2009.

This shift in opinion can also be described as a change in the perceived information in market yields (valuation ratios). Does a low dividend yield in the equity market predict low future returns (reflecting low required risk premiums or investor irrationality) or high future cash flow growth (reflecting growth optimism)? The answer must be one or the other—or some combination of the two. Empirical research has shown that low dividend yields tend to precede subpar market returns rather than above-average growth. In January 2011 in Denver, John Cochrane of the University of Chicago, in the American

©2011 The Research Foundation of CFA Institute

Finance Association's presidential address (see Cochrane 2011), argued that a 100 percent reversal had occurred in academic thinking on this question in the past 20–30 years. Cochrane explained the following:

- The ERP is no longer thought to be constant *over time*. All time variation in market valuation ratios was once thought to reflect changing growth expectations (with an unchanging *ex ante* required risk premium), but now all such variation is thought to reflect changing required returns.
- All expected return variation *across stocks* was thought to reflect stocks' differing betas. Now, the beta is thought to explain none of the cross-sectional variation in expected returns.

Not all academics agree. Some harbor doubts about return predictability and argue that the evidence against a constant risk premium is limited. For example, variation in the ERP could be sample specific or reflect subtle econometric problems in predictability regressions.[1] And those who agree that expected returns vary over time have a follow-up debate over whether this time variation reflects rational drivers (such as wealth-dependent risk aversion), varying amounts of risk in the market, or investor irrationality.

Practitioner thinking has experienced similar shifts. Many investors have become open to the idea of market timing since the decade of boom-to-bust cycles, when forward-looking valuation indicators turned out to give decent forecasts. Yet, even if a time-varying ERP reflects a general tendency for investor risk aversion to rise in bad times, the typical investor should not necessarily become a contrarian market timer. As many investors found out in 2008, their risk appetites fell at least as fast as their wealth, so they did not feel inclined to jump at the bargains (low market valuations, high expected returns). Investors with a longer horizon or relatively stable risk preferences may well be the more natural buyers when such contrarian opportunities arise. Even for them, however, exploiting high expected returns is not easy because no one knows when the market will hit bottom—until after the fact.

Before we turn to forward-looking market analysis, consider the historical equity market performance over the past 111 years shown in **Table 1**. The geometric average excess return of stocks over long-term government bonds has been more than 4 percent in the United States but a bit lower in the rest of the world. (The excess returns would be higher if stocks were compared with short-dated U.S. T-bills or if arithmetic averages were used.) Equities have outperformed bonds in all of the markets Dimson, Marsh, and Staunton (2011) studied. The 20th century may have been especially favorable, however, for stocks versus bonds; the return gap for the 19th century was less than 1 percent in the United States.

---

[1]Typical is the debate between Welch and Goyal (2008) and Campbell and Thompson (2008).

 ©2011 The Research Foundation of CFA Institute

Table 1. Compound Annual (Geometric) Equity Returns and ERPs, 1900–2010

Market	Real Equity Return	ERP over Long-Term U.S. Government Bonds
United States	6.3%	4.4%
World ex-U.S. (in $)	5.0	3.8
World (in $)	5.5	3.8
Range among 19 markets	2.0–7.4%	2.0–5.9%

Source: Dimson, Marsh, and Staunton (2011).

My favorite valuation ratio for the equity market is the inverse of the "Shiller P/E10," which Yale Professor Robert Shiller conveniently updates each month on his website.[2] Because one-year earnings may be too volatile and cyclical for accurate comparisons, Shiller compares today's market prices with smoothed (10-year averages of real) earnings. **Figure 1** compares this ratio, which I'll henceforth call the "real E10/P" or just "E10/P," with the real long-term Treasury yield from January 1900 to February 2011.[3] The solid line correctly predicted high prospective returns for equities in the early 1920s, the 1930s, the 1980s, and more recently in late 2008–2009. Similarly, it captured the low prospective returns in 1929 and 2000, both in stand-alone equity investments and relative to bonds.

Framework to Anchor the Debates

The gap between the two lines in Figure 1 is roughly the forward-looking ERP. Yet, strictly speaking, the Shiller earnings yield equals the *ex ante* real return for equities only under fairly stringent conditions. The dividend discount model (DDM) provides a cleaner conceptual framework than the Shiller earnings yield for assessing the difference between the long-term expected returns of stocks and bonds. Analysts will, of course, debate the inputs of the model and the resulting ERP estimates, but this framework at least gives the debaters a common language.

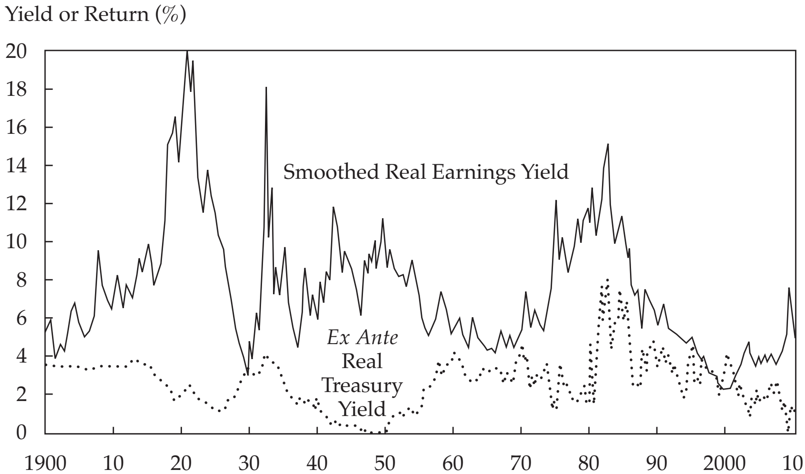
In the basic version of the DDM, cash flows to equity investors (which can be considered, narrowly, to be dividends) are assumed to grow at a constant annual rate G. A feasible long-run return on equities is then the sum of the cash flow yield (here, dividend yield, or D/P) and the trend of cash flow growth rate,

[2]The P/E10 is the price or index value of the S&P 500 Index divided by the average of the last 10 years of earnings. Shiller's website is www.econ.yale.edu/~shiller/data.htm.

[3]In the real long-term Treasury yield, the nominal Treasury yield is deflated by the consensus forecast inflation for the next decade (for the period before survey forecasts became available in the 1970s, statistical estimates were used). For details, see Ilmanen (2011).

©2011 The Research Foundation of CFA Institute

Figure 1. Smoothed Real Earnings Yields of U.S. Equities and *Ex Ante* Real Yields on 10-Year Treasuries, 1900–2011



Sources: Bloomberg; Shiller website (www.econ.yale.edu/~shiller/data.htm); U.S. Federal Reserve; *Blue Chip Economic Indicators*; *Consensus Economics.*

G. The required return on equities, or the discount rate, can be viewed as the sum of the riskless long-term Treasury yield, *Y*, and the required equity-over-bond risk premium, the ERP. Intuitively, markets are in equilibrium when the equity market return that investors require, $Y + ERP$, equals the return that markets are able to provide, $D/P + G$. These expressions can be reshuffled to state the *ex ante* ERP in terms of three building blocks:

$$ERP = D/P + G - Y.$$

The DDM can be expressed in nominal terms (with G_{nom} and Y_{nom}) or in real terms (with G_{real} and Y_{real}) if both expected cash flow growth and the bond yield for expected inflation are adjusted. The model can also be expressed as an earnings discount model if a constant dividend payout rate is assumed. With a constant payout rate, the growth rates of dividends and earnings are equal.

The DDM framework can be easily extended to include a variety of short-term and long-term growth rates, but the use of the DDM to analyze time-varying ERPs can only be informal because it is a steady-state model that assumes constant expected returns and valuation ratios. In a dynamic variant of the DDM, one that allows time-varying expected returns, D/P is a combination of the market's expectations of future (required) stock returns and dividend growth (see Campbell and Shiller 1988).

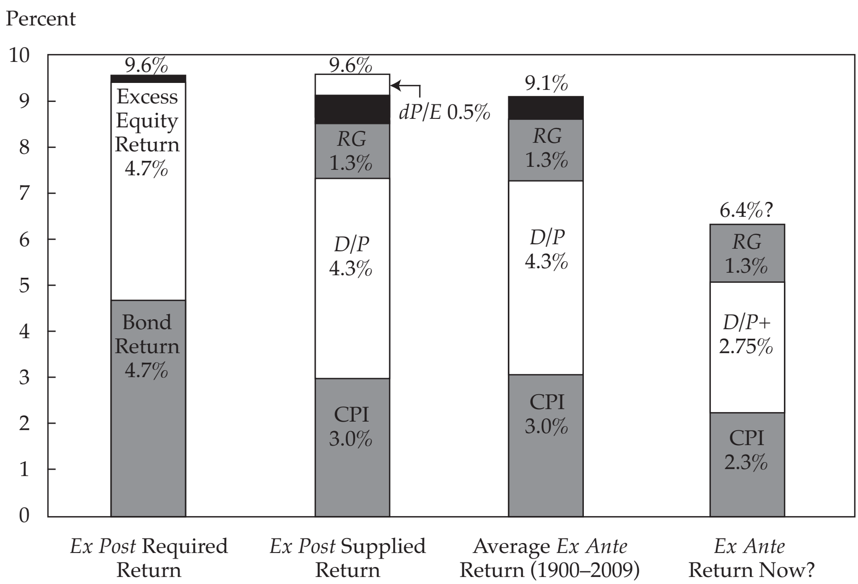
©2011 The Research Foundation of CFA Institute

The DDM framework is simple and flexible, but what inputs to use in calculating the ERP is a topic of wide disagreement. Even the observable inputs—dividend yield and bond yield—are ambiguous because broader payout yields (including, for example, share buybacks) may be appropriate for equities and the maturity and nature (nominal versus real) of the Treasury yield may be debated. The main source of contention, however, is the assumed trend of the growth rate of profits, or earnings per share (EPS), *G*.

Nevertheless, this framework can be used to analyze the building blocks of realized and prospective equity market returns (see Ibbotson and Chen 2003). **Figure 2** decomposes the realized 110-year (1900–2009) compound annual U.S. stock market return of 9.6 percent into its elemental parts with separate decompositions for the "demand" and "supply" of returns. The nomenclature follows Diermeier, Ibbotson, and Siegel (1984). The total return is split into either

- the sum of returns demanded by the investor (the first column in Figure 2), on the assumption that sample averages capture required returns well: 4.7 percent nominal T-bond return + 4.7 percent *ex post* ERP + small interaction terms, represented by the black bands or

Figure 2. Decomposed Historical Equity Market Returns, 1900–2009



Note: RG = real earnings growth, *dP/E* = repricing gains, and CPI is the U.S. Consumer Price Index.
Sources: Arnott and Bernstein (2002); Bloomberg; Shiller website (www.econ.yale.edu/~shiller/data.htm).

- the sum of returns supplied by the economy (the second column in Figure 2): 3.0 percent average inflation + 4.3 percent average dividend yield + 1.3 percent average real EPS growth rate + 0.5 percent repricing effect (which represents the annualized impact of the expansion of the P/E by 75 percent—from 12.5 to 21.9—during the sample period) + small interaction terms.

The third column shows the result when, following Ibbotson and Chen, I deemed the 0.5 percent repricing gain to be an unexpected windfall and subtracted it from the supplied returns.[4] This column suggests, then, that investors required an *ex ante* nominal equity market return of 9.1 percent between 1900 and 2009, on average. If expected returns vary over time and current values differ from the average levels over the sample, this analysis can be misleading for assessing current expected returns. The current inflation rate and equity and bond yields are clearly below historical averages. Using a 2.3 percent rate of CPI growth (the consensus forecast for long-term inflation) and a 2.0 percent *D/P* produces a forward-looking measure predicting only 5.6 percent nominal equity returns. Admittedly, the *D/P* value could be higher if a broader carry measure that included net share buybacks were used, so for the last column in Figure 2, I added 0.75 percent to the estimate (and called it "*D/P*+"). Return forecasts more bullish than the 6.4 percent nominal return in the fourth column would have to rely on growth optimism (beyond the historical 1.3 percent rate of real EPS growth, to be discussed later) or further P/E expansion in the future (my analysis assumes none). More bearish forecasts consider my buyback adjustment excessive and/or my growth or valuation forecasts overly optimistic.

Figure 2 is based on data at the end of 2009. Conveniently, market changes over the subsequent 15 months have been modest. Equity markets have rallied somewhat, with dividend yields dropping from 2 percent to 1.8 percent (and the Shiller E10/P falling from 5 percent to 4.3 percent), whereas Treasury yields and consensus inflation forecasts are virtually unchanged.

So, when asked what I expect the realized outperformance of U.S. equities over Treasuries to be for the decade from the first quarter (Q1) of 2011 to Q1:2021, I pretty much stay with the same numbers. In **Exhibit 1**, I predict 4 percent real (compound annual) return for the equity market and 1 percent real return for Treasuries—close to the current 10-year yield of Treasury Inflation-Protected Securities (TIPS)—thus, a 3 percent ERP. Because inflation terms wash out across stocks and bonds, I do not need to forecast inflation, which is currently an especially hard call. I would assign a ±0.25 percent band around each component estimate.

[4]To be a stickler, I'll note that the yield and growth estimates are consistent only if the payout ratio is constant over time. I could use the real dividend growth rate (averaging 1.2 percent) and the repricing effect based on dividend yield changes (which has a slightly higher annualized impact, 0.7 percent) instead of earnings data, and I would obtain, broadly, the same results.

 ©2011 The Research Foundation of CFA Institute

Exhibit 1. Components of the ERP

Component	Estimate for Next Decade
Equity cash flow yield	2.7% (*D/P* + addition for net buybacks)
+ Real cash flow growth	1.3 (historical average EPS growth)
+ Valuation change	0 (assume unchanged valuations)
− Real Treasury yield	−1
ERP	3%

For the global markets, my ERP forecast is similar. In most countries, I can see somewhat better growth prospects than in the United States, but these prospects are offset by higher real yields. Japan is the one exception; growth prospects are worse there than in the United States.

Debates about the Values of the Main Components

As I have stressed, these building blocks give us a useful framework for debating the values of key components of future ERPs. What are these debates?

Equity Market Yield. Dividend yield is the classic proxy for equity market yield. Having ranged between 3 percent and 6 percent for 40 years, the *D/P* of the S&P 500 Index fell below 3 percent in 1993 for the first time ever and then fell below 2 percent in 1997, remaining there for the next decade. The decline in the *D/P* in the 1980s and 1990s partly reflects a structural change: Many companies replaced dividends with repurchases (i.e., stock buybacks), which were more tax efficient and more flexible and which had a more positive impact on share price (and thereby executive compensation) than did dividends. One reason share buybacks increased is the 1982 change in U.S. SEC rules that provide a safe harbor from price manipulation charges for companies conducting share buybacks.

The obvious improvement in the measurement of the equity market yield would be to include share buybacks. The buyback yield never exceeded 1 percent before 1985 but did in most years thereafter. Even though the buyback yield has in some years exceeded the dividend yield, the buyback yield arguably should not get as high a weight as the dividend yield in any long-run yield measure because it is not as persistent. It is much easier for a corporation to reduce repurchase activities than to cut dividends.

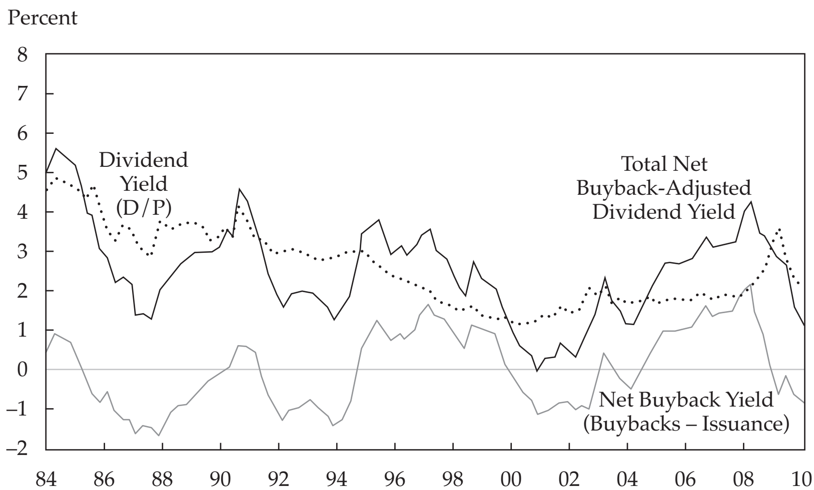
Only adding share buybacks (i.e., not subtracting share issuance), as is sometimes done, would overstate the effective yield. Companies may repurchase shares or pay dividends when they have excess cash, whereas they issue "seasoned" equity when they need more capital from investors.

Cash-financed merger and acquisition deals are another component of cash flows to the investor that could be included in a broad yield measure. The literature on this issue is diverse, however, and hardly conclusive. In computing the net buyback-adjusted yield, net payout yield, and change in Treasury stock, somewhat different data are used to adjust dividend yields, but the intent of all of them is the same: to estimate total cash flow from the company to the investor (see Allen and Michaely 2003; Boudoukh, Michaely, Richardson, and Roberts 2007; Fama and French 2001).

Figure 3 plots one estimate of broader cash flow yield, the dividend yield, and the buyback yield over a quarter century. This broad yield estimate has not been systematically higher than the dividend yield; buybacks and issuance have roughly canceled out over time. Other estimates imply higher cash flow yields, especially since the mid-1990s, so I stay with the 0.75 percent addition over *D/P*. Some may deem this adjustment too high; others, too low. More empirical research is clearly needed.

Equity Cash Flow Growth. Some studies use growth estimates based on analyst expectations for earnings growth or on P/Es, for which they use analyst forecasts of next-year operating earnings. Both approaches embed analyst overoptimism and result in upwardly biased estimates of the ERP.

Figure 3. Equity Market Yield Measures, 1984–2009



Sources: Haver Analytics; Nomura.

©2011 The Research Foundation of CFA Institute

A more conservative approach is to use the trend of the rate of growth in real GDP or corporate profits.[5] Even this approach turns out to be overoptimistic. Although many practitioners think that the GDP growth rate is a *floor* for earnings and dividend growth, the rate has historically been a *ceiling* that has been broken only during benign decades. Arnott and Bernstein (2002), Bernstein and Arnott (2003), and Cornell (2010) showed that growth rates of per share earnings and dividends have, over long histories, lagged the pace of GDP growth and sometimes even per capita GDP growth. As **Table 2** shows, between 1950 and 2009, growth rates of earnings and dividends per share almost matched the 1.9 percent real growth rate of GDP per capita but clearly lagged real GDP growth (3.1 percent).

Table 2. Average Real Long-Term Growth Rates (Geometric Means), 1900–2009

Period	Real GDP	Real GDP per Capita	Real EPS	Real Dividends per Share
1900–1949	3.2%	1.8%	1.0%	1.0%
1950–2009	3.1	1.9	1.5	1.3

Sources: Arnott and Bernstein (2002); Haver Analytics.

Taking even longer histories does not help. The first half of the 20th century looked even worse for earnings and dividend growth. When I looked at shorter histories, I saw a prettier picture for a while. Between 1988 and 2007, U.S. real EPS growth averaged 3.7 percent a year—clearly larger than the real GDP growth rate (2.4 percent). This period was an exceptionally benign one, however, for capital markets; for example, the share of GDP represented by corporate profits rose from 8 percent to 11 percent. After 2008, the trailing 20-year real EPS growth rate was negative; after the 2009 recovery, it was still only 1.3 percent.

Studying the global evidence also does not help to raise the growth estimate. Dimson, Marsh, and Staunton (2002) showed that between 1900 and 2000, growth in real dividends per share lagged growth in real GDP per capita in 15 of the 16 countries they examined. Across countries, real dividend growth averaged nearly zero and lagged growth in real GDP per capita by 2.4 percentage points. U.S. dividend growth was somewhat better but still lagged growth in real GDP per capita by 1.4 percentage points.

[5] Some analysts use the trend in the growth of nominal earnings (say, 7 percent). By doing so, they conveniently forget that such nominal growth occurred over a period when inflation averaged 4 percent, whereas the current expected inflation is closer to 2 percent.

©2011 The Research Foundation of CFA Institute

MSCI Barra (2010) has contrasted (real) EPS growth and GDP growth between 1969 and 2009 in 16 countries. The researchers found that, averaged across all the countries, annual GDP growth was 2.4 percent—compared with 0.1 percent EPS growth. (Comparable figures in the United States are 2.8 percent and 1.3 percent.) The gap in growth rates between GDP and EPS was positive (0.5–5.0 percent) in all the countries studied except Sweden.

Why? These patterns seem puzzling. In the long run, GDP and profits should have similar trends in growth rates; otherwise, the corporate sector would eventually dominate the economy. (Admittedly, this argument is only relevant over extremely long periods.) An important distinction must be made, however, between aggregate earnings growth and EPS growth. Aggregate earnings growth has matched GDP growth quite closely during the post–World War II era; EPS growth has not.

Investors in existing listed stocks capture only part of aggregate profit growth because a portion of this growth is financed with newly issued equity. Arnott and Bernstein (2002) stressed that new entrepreneurs and labor (including top management) capture a large share of economic growth at the expense of shareholders in existing companies. Stock market indices (made up of listed stocks) miss the most dynamic growth in the economy, which comes from unlisted start-up ventures, other small businesses, and sole proprietorships—all of which count toward total business profits.

Total corporate profit growth is, therefore, effectively diluted by net equity issuance. Cornell (2010) showed that the annual dilution rate (mainly through new business creation but also through net issuance by existing companies) between 1926 and 2008 was 2 percent and reasonably stable over time. Subtracting the 2 percent dilution effect from 3 percent real aggregate earnings growth makes 1 percent real EPS growth a realistic long-run prospect. Some evidence indicates, however, that the dilution effect has flattened during the past decade, perhaps reflecting the increasing use of buybacks.

Although several studies confirm these patterns, the crucial distinction between aggregate earnings growth and EPS earnings growth is not widely appreciated, and many ERP estimates rely on at least a 3 percent real trend in EPS growth. As Upton Sinclair said, “It is difficult to get a man to understand something, when his salary depends upon his not understanding it.” Still, it is true that over a single decade, real EPS growth may deviate significantly from its long-run trend, so this building block can be subject to very vigorous debates.

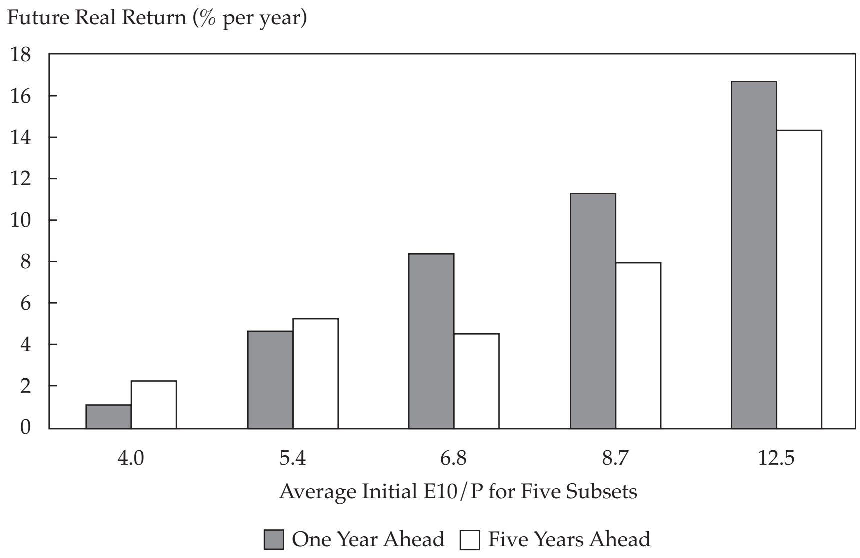
Valuation Change. I have assumed here unchanged market valuations over the coming decade. It is often a good base assumption in normal circumstances.

 ©2011 The Research Foundation of CFA Institute

One can argue, however, that current equity markets are expensive in an absolute sense. The Shiller P/E10 is near 23, more than 40 percent above its long-run average. The smoothed real earnings yield is only 4.3 percent (100/23), not far from the average of the bottom quintile over a 110-year history. **Figure 4** shows that real stock market returns have typically been modest in years following low starting yields (and high following high starting yields). Generally, Figure 4 indicates that this valuation ratio has the useful ability to predict future market returns.[6]

Other market valuation indicators suggest that equity markets are fairly valued. And in comparison with even more expensive Treasuries, the equity market may appear to be cheap.

Figure 4. Average Level of E10/P and Subsequent Returns by Periods, 1900–2009



Notes: The graph was created by sorting each month into one of five buckets based on the level of real E10/P at the beginning of the month and then computing the average level for E10/P (*x*-axis labels) and subsequent one-year and five-year real stock market returns (*y*-axis values) in five subsets of the sample history. Real return is the S&P 500 return.

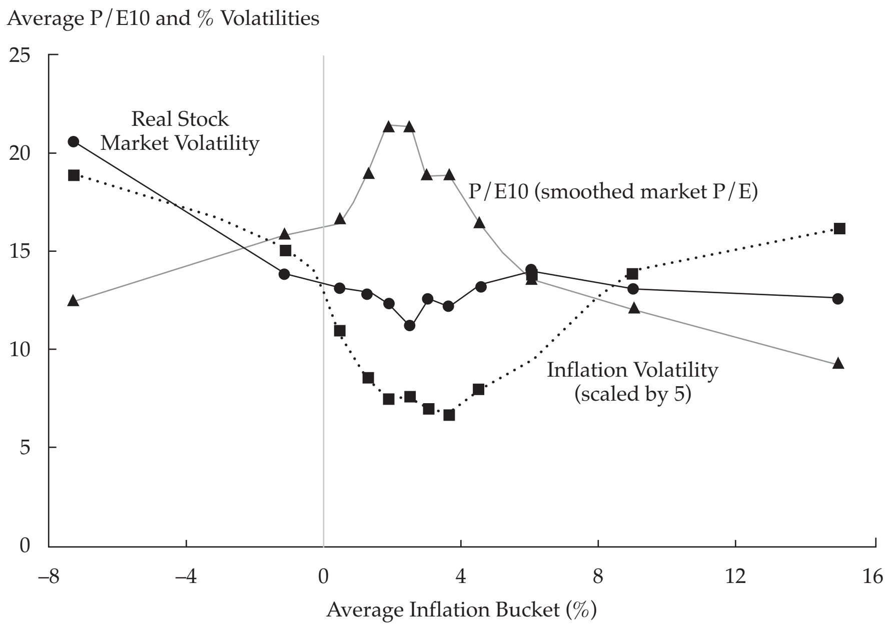
Sources: Shiller website (www.econ.yale.edu/~shiller/data.htm); Haver Analytics.

[6]The predictive ability is somewhat overstated because the sorting of months into quintiles uses in-sample information. Investors know only with hindsight that 4 percent earnings yields would be among the lowest and 12 percent yields among the highest during the full sample. The mean-reversion effect is, therefore, overstated.

In addition to market valuations, many other determinants of the outlook for growth and valuation can be considered. Bearish observers focus on debt problems, deleveraging, and unfavorable demographics. Bullish observers note that technological progress has tended to surprise on the upside and that widening knowledge and access to information may benefit from increasing returns to scale, unlike traditional capital, which tends to exhibit decreasing returns to scale.

I highlight one bearish consideration. High inflation tends to hurt equity markets, but so does deflation. Steady and low, but positive, inflation appears to be the optimal environment for real growth and risky-asset valuations. **Figure 5** shows a sombrero-shaped relationship between equity market valuation levels (P/E10) and inflation levels over the past 110 years. The sweet spot of peak valuations occurs with inflation in the 1–4 percent range. One mechanism behind this nonlinear relationship is that economic uncertainty—here measured by inflation volatility and equity market volatility—tends to be higher amid deflation and high inflation. Thus, inflation may not directly influence

Figure 5. U.S. Equity Market Valuations and Inflation, 1900–2009



Note: The graph was created by sorting each month into 1 of 12 subsets on the basis of the level of inflation during the month and then computing the average level for inflation (x-axis), the P/E10 valuation ratio, and the two volatility series (y-axis) in the 12 subsets of the sample history.

Sources: Haver Analytics; Shiller website (www.econ.yale.edu/~shiller/data.htm); author's calculations.

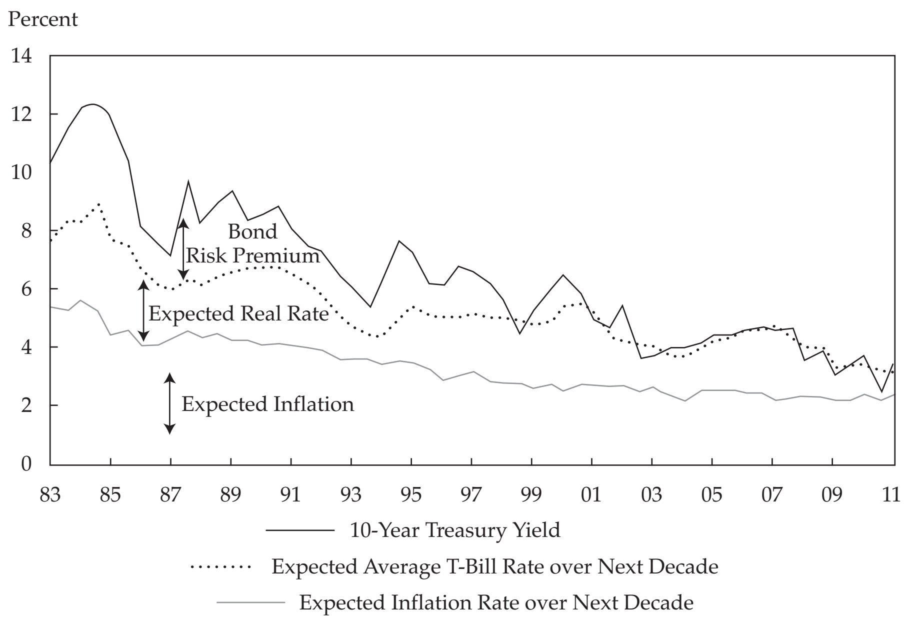
 ©2011 The Research Foundation of CFA Institute

equity market valuations, but it affects the market through its impact on economic growth and uncertainty. Whatever the reason, the pattern is bad news for market valuations because two decades have been at the sweet spot, so the likelihood of both deflation and high inflation for the coming decade has substantially increased.

Treasury Yield. This component is subtracted. Bonds appear at least as expensive as stocks when measured by historical yardsticks, especially in comparison with the past 30 or 60 years of experience. Moreover, the debt and demographic problems make many expert observers worry about inflation reaching levels not seen since the 1980s.

A perhaps surprising phenomenon is that current bond yields do not contain much of a risk premium. **Figure 6** clarifies this statement by decomposing the 10-year Treasury yield into three components: expected average inflation, expected average real T-bill rates, and the required bond risk premium over bills. The decomposition is based on consensus forecasts of next-decade average inflation and average T-bill rates. The current 10-year yield of 3.4

Figure 6. Decomposition of the 10-Year Treasury Yield Based on Survey Data, 1983–2011



Note: Each year measurement is as of March and October.

Sources: Bloomberg; *Blue Chip Economic Indicators.*

percent is close to the average expected T-bill rate, implying a bond risk premium of nearly zero. Simply put, the yield curve is exceptionally steep, but all this steepness seems to reflect the market's expectation of short rates rising sharply from the abnormal near-zero level. The expected real yield on the nominal 10-year bond is slightly more than 1 percent, well below the past 30-year average of 3 percent. The 10-year TIPS has a yield slightly under 1 percent, but this yield is an average reflecting negative real yields at the front end and clearly higher real yields further out.

The reasons for Treasuries' continued richness include still-modest inflation; the exceptional safe-haven role of Treasuries in recessions, deflations, and financial crises (which has been extremely valuable in the past decade but may not work as well in the next decade); and various exceptional sources of demand (large asset purchases by the Fed, reserve accumulation by other central banks, and purchases by pension funds seeking close asset/liability matching).

I simply assume a 1 percent real bond return for the next decade, which is broadly in line with the current market pricing of both nominal and inflation-linked Treasuries. These yields are known today.

An alternative way of computing the ERP involves comparing stock returns with the returns of constant-maturity bonds (or of long-term bond indices) over time. If such a method is used, the results thus depend on future yield changes. Unexpectedly bond-bearish outcomes would probably also hurt equity market valuations. They might leave the realized excess return of stocks and bonds broadly unchanged, but with both asset classes earning real returns lower than the now expected, respectively, 4 percent and 1 percent.

Concluding Thoughts

In this paper, I focus on the prospects of the equity risk premium over the next decade. However, it is worthwhile to think about the *term structure* of such premiums. A world of time-varying expected returns contains more than one premium number. The short-run and long-run premiums can differ significantly. How would the forecast beyond 2021 differ from the prediction for the next decade?

- The term structure effects are more obvious on the bond side of the premium. Short-dated TIPS yields are currently negative (consistent with short-dated nominal Treasuries yielding nearly zero while headline inflation is nearly 2 percent and rising). At the same time, the 10-year TIPS yield is 0.9 percent and the 20–30 year TIPS yields are approaching 2 percent. Together, these yields imply a 2.7 percent forward TIPS yield for the decade starting in 2021.

 ©2011 The Research Foundation of CFA Institute

- Abnormally high (or low) starting valuations for equity markets and related mean-reversion potential have strong implications for expected stock market returns for the next few years. When considering prospective equity returns *after* the next decade, however, it is impossible to know what the starting valuation levels will be in 2021. Thus, if one assumes below-average equity market returns for the next decade because of an expected normalization of the currently high Shiller P/E10, the best forecast for real equity market returns beyond 2021 should be close to the "unconditional" long-term return forecasts. That is, these "forward forecasts" should largely ignore starting valuations (or at least allow future higher starting yields in 2021 than in 2011).
- Many indicators in addition to valuation measures can be used to predict stock market returns. Regressions and other econometric techniques can be used to forecast returns over any investment horizon (admittedly, they have fewer independent data points in long-horizon regressions). Thus, we can estimate a full term structure of expected returns. (Such forecasts are always model specific, but such a situation is no worse than the situation with informal and judgmental forecasts.)

The following empirical fact is worth emphasizing: Although beta risk has been well rewarded across asset classes (in the sense of the capital asset pricing model, in which the stock market, with a beta near 1, has outperformed the bond market, with a beta near 0, by 3–4 percent over long time periods), the same is not true *within* stock markets. High-beta and high-volatility assets in most stock markets have hardly outperformed their low-volatility peers in the long run; often, the reverse has occurred. Such risk without reward has increasingly attracted investor attention.

This paper focuses on the equity risk premium, but I want to finish with this exhortation: LOOK MORE BROADLY! A key theme in my recent book (Ilmanen 2011) is that relying exclusively or primarily on the ERP as the source of long-run returns causes portfolios to be inadequately diversified. Investors should broaden their horizons beyond asset class perspectives to consider various dynamic strategies (value, carry, trend, volatility, illiquidity) as well as underlying risk factors. The result for investors will be smarter portfolios than they currently have and better long-run performance.

References

Allen, Franklin, and Roni Michaely. 2003. "Payout Policy." In *Handbook of the Economics of Finance*, vol. 1. 1st ed. Edited by George M. Constantinides, Milton Harris, and René M. Stulz. Amsterdam: Elsevier B.V.

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal*, vol. 58, no. 2 (March/April):64–85.

Bernstein, William J., and Robert D. Arnott. 2003. "Earnings Growth: The Two Percent Dilution." *Financial Analysts Journal*, vol. 59, no. 5 (September/October):47–55.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts. 2007. "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing." *Journal of Finance*, vol. 62, no. 2 (April):877–915.

Campbell, John Y., and Robert J. Shiller. 1988. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance*, vol. 43, no. 3 (July):661–676.

Campbell, John Y., and Samuel B. Thompson. 2008. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, vol. 21, no. 4 (July):1509–1531.

Cochrane, John. 2011. "Discount Rates." Presidential Address at the 2011 American Finance Association meeting (25 May); reprinted in *Journal of Finance*, vol. 66, no. 4 (August 2011):1047–1108 (http://faculty.chicagobooth.edu/john.cochrane/research/papers/AFA_pres_speech.pdf).

Cornell, Bradford. 2010. "Economic Growth and Equity Investing." *Financial Analysts Journal*, vol. 66, no. 1 (January/February):54–64.

Diermeier, Jeffrey J., Roger G. Ibbotson, and Laurence B. Siegel. 1984. "The Supply of Capital Market Returns." *Financial Analysts Journal*, vol. 40, no. 2 (March/April):74–80.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns.* Princeton, NJ: Princeton University Press.

———. 2011. *Credit Suisse Global Investment Returns Yearbook 2011.* Zurich: Credit Suisse Research Institute.

Fama, Eugene F., and Kenneth R. French. 2001. "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?" *Journal of Financial Economics*, vol. 60, no. 1 (April):3–43.

Ibbotson, Roger G., and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):88–98.

Ilmanen, Antti. 2011. *Expected Returns: An Investor's Guide to Harvesting Market Returns.* Chichester, U.K.: John Wiley & Sons.

MSCI Barra. 2010. "Is There a Link between GDP Growth and Equity Returns?" MSCI Barra Research Bulletin (May).

Welch, Ivo, and Amit Goyal. 2008. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, vol. 21, no. 4 (July):1455–1508.

 ©2011 The Research Foundation of CFA Institute

Will Bonds Outperform Stocks over the Long Run? Not Likely

Peng Chen, CFA
President, Global Investment Management Division
Morningstar Investment Management

Given the poor performance of stocks in the past decade, ample discussion has concerned the relative performance of stocks and bonds. Some even argue that investors should allocate assets entirely to bonds, not only because bonds are the safer investment but also because they believe bonds will outperform stocks over the long run. In other words, if bonds can deliver higher returns than stocks with less risk, why bother with stocks?

The impressive performance of the stock market in the 1980s and 1990s and the resulting rise in investor expectations spurred numerous articles that called attention to the historical market return and cautioned investors about overly optimistic expectations. Many studies forecasted equity returns that would be much lower when compared with the historical average. A few even predicted that stocks would not outperform bonds in the future. Later, after the bear markets of 2000–2002 and 2007–2009, the reverse happened. Investors tended to have very pessimistic expectations for stock returns. A study of the historical returns is, therefore, useful for bringing sense to either situation, whether overly optimistic or overly pessimistic expectations.

Table 1 shows the performance of the S&P 500 Index, the Barclays Capital (BarCap; formerly, Lehman Brothers) U.S. Aggregate Bond Index, the Ibbotson U.S. Intermediate-Term Government Bond Index, and the Ibbotson U.S. Long-Term Government Bond Index over various time periods. Average annual stock returns have been poor relative to bonds not just for the past 10 years; stock returns look mediocre for the past 20, 30, and even 40 years relative to bond returns. According to returns over the past 40 years, the argument that bonds might outperform stocks in the long run appears to be valid. But one should view these data with skepticism. Note that over the 20-, 30-, and 40-year periods, stocks actually performed quite well. In fact, stocks have outperformed their long-run average return since 1926. Only during the past 10 years have stocks significantly underperformed both the long-term average and bonds. We should also note that bonds over the past 40 years, in particular relative to stocks over the past 10, have done extremely well. Bonds have significantly outperformed their long-term averages since 1926.

©2011 Ibbotson Associates, Inc. All rights reserved.

Table 1. Compound Annualized Total Returns Ending December 2010

Span and Start Date	S&P 500	BarCap U.S. Aggregate	Ibbotson U.S. Intermediate-Term Government	Ibbotson U.S. Long-Term Government
1 Year: Jan 2010	15.06%	6.54%	7.12%	10.14%
5 Years: Jan 2006	2.29	5.80	6.06	5.58
10 Years: Jan 2001	1.41	5.84	5.64	6.64
20 Years: Jan 1991	9.14	6.89	6.56	8.44
30 Years: Jan 1981	10.71	8.92	8.51	10.18
40 Years: Jan 1971	10.14	8.32[a]	7.81	8.57
Jan 1926–Dec 2010	9.87	—	5.35	5.48

[a]The BarCap U.S. Aggregate goes back only to January 1976.

Over the very long term, however, it is no longer a contest. **Figure 1** shows the hypothetical value of $1 invested at the beginning of 1926 for the major capital market asset classes. Over this 85-year period, stocks easily beat bonds.

Consider these various long-term histories of U.S. stocks' compounded total returns:

January 1825–December 1925[1]	7.3%
January 1926–December 2010	9.9%
January 1825–December 2010	8.5%

The returns on the stock market have been consistently high for almost two centuries. The returns over the past 40 years are roughly comparable to the returns from the more distant past. Long-term history provides two major insights:

1. Stocks have outperformed bonds.
2. Stock returns are far more volatile than bond returns and are thus riskier. Given the additional amount of risk, it is not surprising that stocks do not outperform bonds in every period—even over extended periods of time.

Stocks vs. Bonds in the Future

How likely are stocks to outperform bonds in the future? As a first attempt to figure out the future, let's look in more detail at what happened during the past 40 years. We can decompose the stock and bond returns into several components:

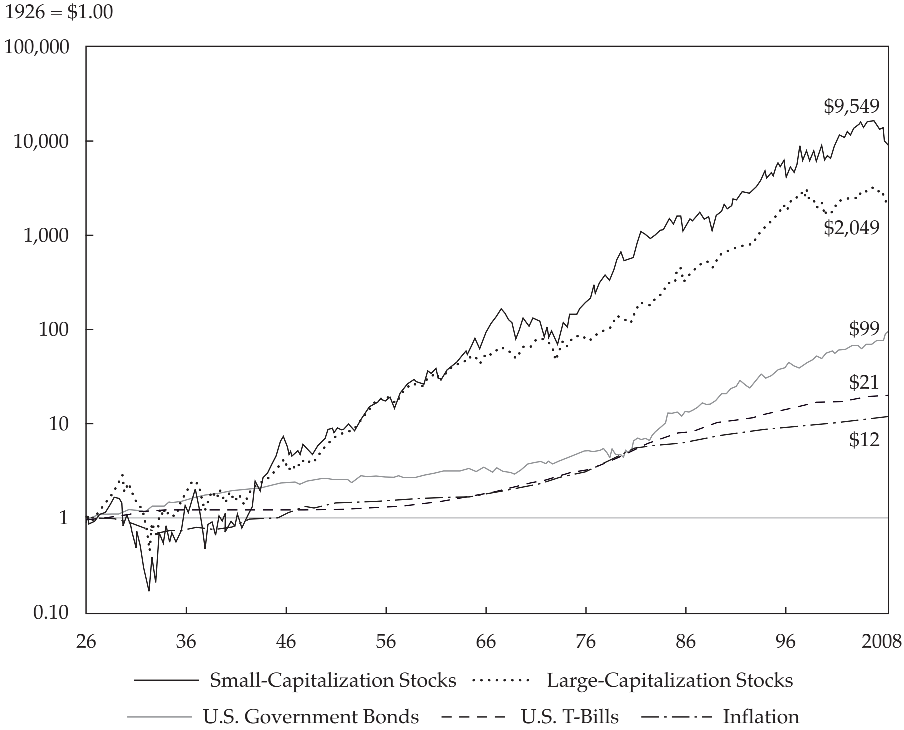
Bond return = Current yield + Capital gain;

Stock return = Current yield + Earnings growth + P/E change.

Despite the substantial decline in yields over the past 40 years, and thus substantial capital gains on bonds, **Figure 2** shows that the bulk of returns on

[1]Stock returns for 1825–1925 are from Goetzmann, Ibbotson, and Peng (2001). For 1926–2010, returns are from Ibbotson Associates (2011).

Figure 1. Stocks, Bonds, Bills, and Inflation, 1926–2008
(lognormal)

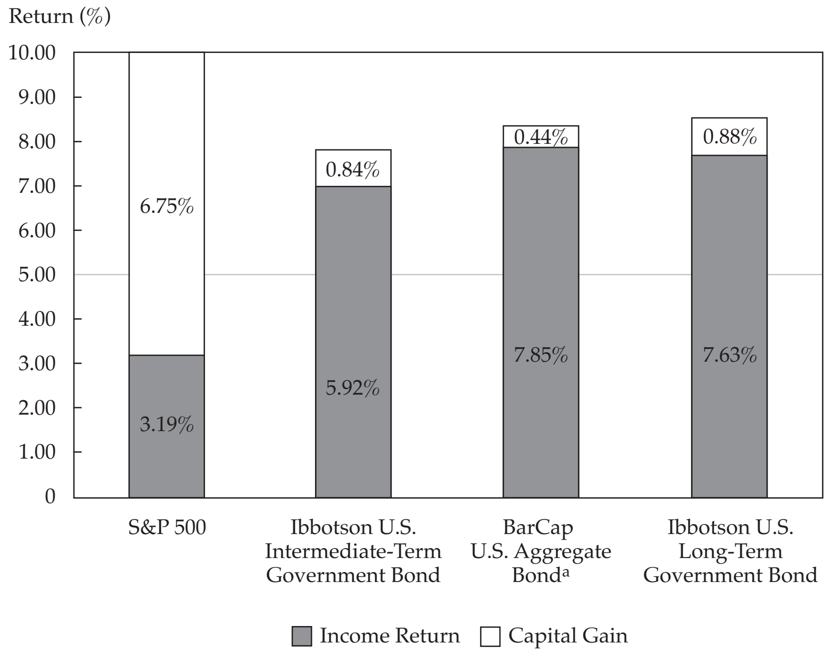


Note: Results assume reinvestment of income and no transaction costs or taxes.

the bond indices over the past 40 years came from the income return portion, or yield. On average, the bond income return from coupon payments was more than 7 percent. Capital gains caused by the yield decline made up the additional return. In contrast, over the past 40 years, stock returns consisted of 3.2 percent from dividend yield and 6.8 percent from capital gains. Next, let's look at what these components would look like going forward.

Today, bond yields are much lower than those shown in Figure 2. **Table 2** compares current bond yield information with yields at the beginning of 1971. As of the end of 2010, the Ibbotson long-term government bond yield was 4.14 percent and the Ibbotson intermediate-term government bond yield was only 1.70 percent. For bonds to continue to enjoy the same amount of capital gains over the next 40 years, their yields, especially the yield on intermediate-term government bonds, would probably have to move into negative territory. Such

Figure 2. Decomposition of Historical Returns, January 1971–December 2010



[a]BarCap U.S. Aggregate goes back only to January 1976.

a development would be impossible because it implies that investors would be willing to pay for the privilege of lending their money to a borrower. Over the past 40 years, bond investors have enjoyed abundant returns because of a high-yield environment at the beginning of the period followed by a steady decline in yields. Going forward, these conditions are not likely to repeat; we are currently experiencing a much lower-yield environment with a higher likelihood of yield increases than decreases.

Table 2. Bond Yields

Bond Index	January 1971	December 2010	Change
Ibbotson U.S. Long-Term Government	6.12%	4.14%	−1.98
BarCap U.S. Aggregate[a]	7.92	2.97	−4.95
Ibbotson U.S. Intermediate-Term Government	5.70	1.70	−4.00

Note: Change is in percentage points.

[a]The BarCap U.S. Aggregate goes back only to January 1976, so average yield was calculated as starting from that date.

Given the current low-yield environment, it would be almost impossible for bonds to generate the same amount of capital gains as they did in the past. In fact, a reasonable estimate might be that no more capital gains will be available in the near future because yields are at least as likely to rise as to fall.[2] If no future fall in yields were to occur, all of the return would have to come from the coupon return. That means the total return for bond investments would likely be 3–4 percent.

For stocks, the current dividend yield from January to December 2010 for the S&P 500 was 2.03 percent, which is a good baseline forecast of the future dividend yield levels. If stocks produce more than 2 percent in capital gains per year on average, adding the 2.03 percent dividend yield would result in a total stock return of 4 percent. Thus, just from simply looking at the decomposition of the past returns and making some simple forward-looking assumptions, one should expect that stocks will likely beat bonds going forward.

Let's elaborate some more on stocks' capital gains portion. Stocks' capital gain or price increase can be decomposed into nominal earnings growth and change in the P/E (see Ibbotson and Chen 2003). Historically, U.S. long-term nominal earnings growth has been roughly 4.65 percent, which is comparable to U.S. long-term nominal GDP growth. If we assumed that the market valuation level (the P/E of the S&P 500) would stay at the same level today over the next 40 years, then we would have an equity return of around 7 percent by adding the current dividend yield and nominal earnings growth. This means that the stock return will be in the 7 percent neighborhood, and the bond return will be around 3–4 percent. Even if we forecasted a decline in the valuation level, the 10-year average P/E would need to fall from its current level of about 20 to below 5 to result in average equity returns around 3 percent over the next 40 years. The lowest level of the P/E on the S&P 500 since 1926 was recorded at 7.1 in 1948; it has never gotten to a level less than 5, even through the Great Depression during the 1920s and 1930s and the 2008–09 global financial crisis. Again, this shows that it is unlikely that stocks will underperform bonds over the next 40 years.

Forecasting Expected Returns

The previous section showed a simple return decomposition and included some observations on future stock and bond returns. The following section will use the building block method to derive the expected returns on bonds and the supply-side equity risk premium model to derive expected returns on stocks.

[2] Some would even argue that bond yields are likely to increase over time, thus producing capital losses for bonds.

Building Block Method. The building block method was first introduced in Ibbotson and Sinquefield (1976). This approach uses current market yields as its foundation and adds estimated risk premiums to build expected return forecasts. This approach separates the expected return of each asset class into the three components shown in **Exhibit 1**.

Exhibit 1. Building Block Approach to Generating Expected Returns

Component	Description
Real risk-free rate	Return that can be earned without incurring any default or inflation risk
Expected inflation	Additional reward demanded to compensate investors for future price increases
Risk premium	Additional reward demanded for accepting uncertainty associated with a given asset class

When choosing a risk-free rate, Ibbotson Associates uses U.S. Treasury yield-curve rates with a maturity to match the investment period. **Table 3** outlines the risk-free rates that are applied to various time horizons. In this paper, because we are mostly interested in the long-term expected returns, we use the long-term (20-year) risk-free rate.

Table 3. Risk-Free Rates for Various Time Horizons

Time Horizon	Years to Maturity	Yield
Short term	5	2.01%
Intermediate term	10	3.30
Long term	20	4.13

Note: All data are from the U.S. Treasury Department website as reported for 31 December 2010.

Some risk premiums can be derived by subtracting the historical average return of one asset class from another or by subtracting the risk-free rate from the return of an asset class. In this way, past data are incorporated into the forecast of future returns; the assumptions are that the financial market is relatively efficient over time and that the realized return differential is a good measure of what investors are expecting to be compensated for in order to take on the various risk levels among different asset classes. Various premiums are added to the current risk-free rate to forecast the expected return unique to each asset class.

Historical returns are calculated over annual periods and may, depending on the nature of the benchmark, use income or total returns. In general, total returns are used for equity forecasts, whereas income returns are used for fixed-income forecasts. Total return is composed of capital appreciation and income (interest payments or dividends). For fixed-income asset classes, the realization of capital gains and losses is assumed to sum to zero over the time horizon of the investment. (In other words, coupon-paying bonds are assumed to be bought at par and are expected to mature at par.) The assumption is that the current market yield is the best forecast of *expected* returns on bonds (i.e., when investors buy bonds, they are expecting neither capital gain nor capital loss).

Expected Return for Bonds. For bond asset classes, Ibbotson Associates identifies three risk premiums that can impact the returns—a horizon premium, a default premium, and a mortgage prepayment premium, as shown in **Table 4.** The horizon premium measures the excess yield that investors in long-term fixed income expect to receive in exchange for accepting additional uncertainty and potential loss of liquidity. Ibbotson Associates estimates the horizon premium as the difference (in the income return) between two government bonds. The first government bond (which is called the "government bond proxy") has the same maturity as the asset class being modeled; the second government bond is the risk-free rate.

Table 4. Detailed Methodology on Expected Return Estimations, 31 December 2010

				Fixed Income		
Benchmark	Expected Return, Geometric	Long-Term Risk-Free Rate	Equity Risk Premium	Horizon Premium	Corporate Default Premium	Mortgage Prepayment Premium
Stocks (S&P 500)	7.61%	4.13%	3.34%			
BarCap U.S. Aggregate	4.45	4.13	—	−0.34%	0.26%	0.40%
Ibbotson U.S. Long-Term Government	4.13	4.13	—	—	—	—
Ibbotson U.S. Intermediate-Term Government	3.61	4.13	—	−0.52	—	—
T-bills	2.49	4.13	—	−1.64	—	—

The corporate default premium measures the historical reward received for holding corporate bonds rather than government bonds of the same maturity. The corporate default premium is equal to the difference between a pure corporate benchmark and a government bond of the same maturity. This difference is multiplied by the corporate exposure in the particular bond asset class.

The mortgage prepayment premium depends on early delivery of mortgage payments that may subsequently change the cash flow and total return received by an investor. The premium is calculated as the difference between the arithmetic mean income return of an index of pure mortgage-backed securities and the arithmetic mean income return of a government bond proxy with the same maturity as the mortgage-backed index. This difference is then multiplied by the percentage of mortgage exposure found in the asset class benchmark:

The specific fixed-income-premium calculations are as follows:

Horizon premium	=	Ibbotson government bond proxy[a] income return	−	Ibbotson government bond proxy[b] income return		
Corporate default premium	=	Corporate bond index income return	−	Ibbotson government bond proxy[a] income return	×	Percent corporate bond exposure
Mortgage prepayment premium	=	Mortgage bond index income return	−	Ibbotson government bond proxy[a] income return	×	Percent mortgage bond exposure

[a]Same maturity (average or current) as the asset class benchmark.
[b]Same maturity as the time horizon (i.e., 20 years).

The resulting estimated expected returns for various bond asset classes are shown in Table 4.

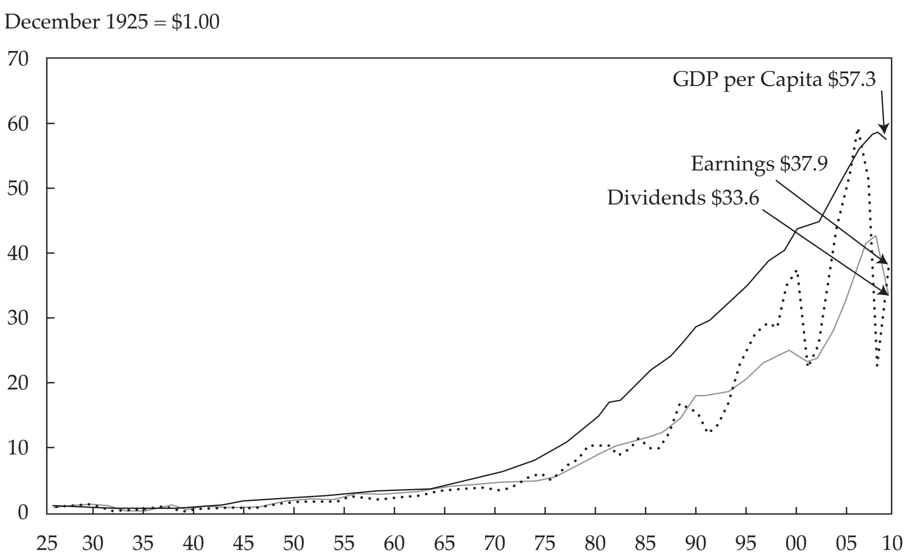
Long-Term Expected Return for Stocks and Equity Risk Premium. The expected return of stocks over bonds has been estimated by a number of authors using various approaches. Such studies can be categorized into four groups based on the approaches they have taken. The first group of studies derives the ERP from historical returns between stocks and bonds. By taking the long-term bond returns (5.48 percent) from the stock returns (9.87 percent) from Table 1, we arrive at a historical compounded equity risk premium estimate of 4.16 percent. The second group uses supply-side models to measure the expected ERP. These models incorporate fundamental information, such as earnings, dividends, and overall productivity. A third group adopts demand-side models that derive the expected return of equities through the payoff demanded by equity investors for bearing additional risk. The fourth group relies on the opinions of financial professionals through broad surveys.

Ibbotson Associates establishes an equity risk premium by following the supply-side approach outlined in Ibbotson and Chen (2003). Their work combined the first and second approaches to arrive at a forecast of the ERP. By proposing a new supply-side methodology, the Ibbotson–Chen study challenges current arguments that future returns on stocks over bonds will be negative or close to zero. The results affirm the relationship between the stock market and the overall economy. They also provide implications for investors creating a policy for allocating assets between stocks and bonds. The following section will briefly explain the methodology presented in Ibbotson and Chen (2003). For detailed explanations, please refer to the original article.

■ *Supply model.* Long-term expected equity returns can be forecasted by using supply-side models. The supply of stock market returns is generated by the productivity of corporations in the real economy. Investors should not expect a much higher or lower return than that produced by the companies in the real economy. Thus, over the long run, equity returns should be close to the long-run supply estimate.

Earnings, dividends, and capital gains are supplied by corporate profitability. **Figure 3** illustrates that earnings and dividends have historically grown in tandem with the overall economy (GDP per capita), adjusting for inflation. So, if one assumes that the economy will continue to grow, dividends and earnings should also continue to grow, thus continuing to drive stock performance. Capital gains did not, however, outpace the stock market—primarily because the P/E increased by a factor of 2 during the same period. In other words, investors' appetite to pay for per unit of earnings has increased roughly two times over the period.

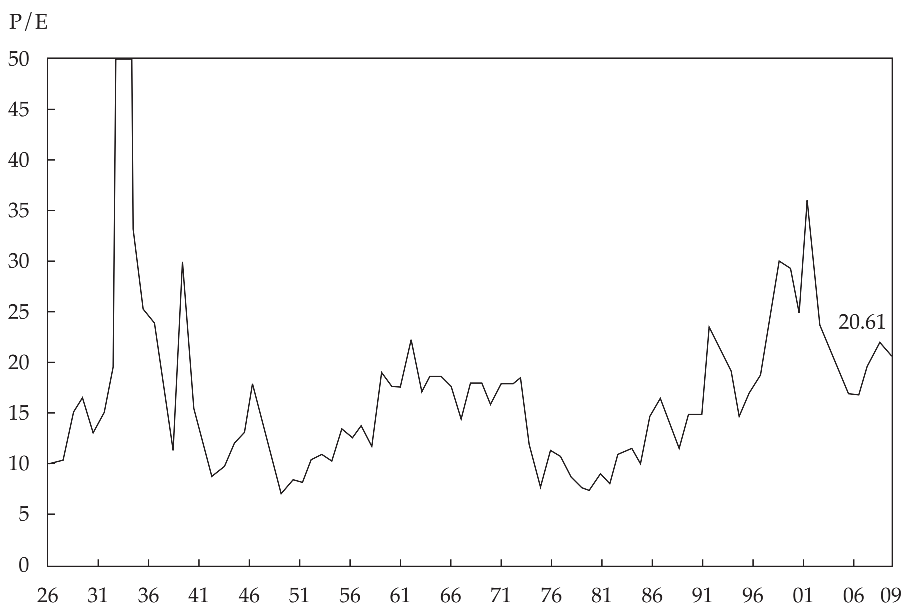
Figure 3. Growth of $1.00 in GDP per Capita, Earnings, and Dividends, 31 December 1925 to December 2010



■ *Forward-looking earnings model.* Two main components make up the supply of equity returns: current returns in the form of dividends and long-term productivity growth in the form of capital gains. The discussion that follows identifies and analyzes components of the earnings model that are tied to the supply of equity returns. This discussion leads to an estimate of the long-term sustainable equity return based on historical information about the supply components.

The Ibbotson Associates earnings model breaks the historical equity return into four components. Only three—inflation, income return, and growth in real earnings per share—have historically been supplied by companies. The growth in P/Es, the fourth piece, is a reflection of investors' increased appetite to pay the price per unit of earnings produced. We believe that the past supply of corporate growth (through dividend and earnings growth) is forecasted to continue but that a continued increase in investors' appetite to pay for per unit of earnings is not. The P/E rose dramatically over the past 80 years because investors believed that corporate earnings would grow faster in the future. This growth in P/E accounted for a small portion of the total return on equities during the period. **Figure 4** depicts the P/E from 1926 to 2009. The P/E was 10.22 at the beginning of 1926 and 20.61 in 2009—an average increase of 0.84 percent

Figure 4. P/E, 1926–2009



Note: The P/E in 1932 went off the chart to 136.50.

per year. The highest P/E was 136.50, recorded in 1932, and the lowest was 7.07, recorded in 1948. (The P/Es in Figure 4 may differ from some of the others presented in this book because of varying definitions of earnings.)

Ibbotson Associates subtracts the historical P/E growth rate from the equity risk premium forecast because we do not believe that the P/E will continue to increase in the future. The market serves as the cue. The current P/E is the market's best guess regarding the future of corporate earnings, and we have no reason to believe, at this time, that the market will change its mind. Thus, the supply of equity return includes only inflation, the growth in real EPS, and income return. Instead of using one-year earnings in calculating the P/E, as in Ibbotson and Chen (2003), we use three-year average earnings in this calculation. The reason is that reported earnings are affected not only by long-term productivity but also by "one-time" items that do not necessarily have the same consistent impact year after year.[3] For example, the 2003 earnings used in this calculation are the average reported earnings from 2002, 2003, and 2004. For 2009, the earnings are the average of reported earnings in 2008 and 2009 and the estimated earnings for 2010. Using a three-year average rather than year-by-year numbers is more reflective of the long-term trend.

The historical P/E expansion is calculated to be roughly 0.82 percent per year; therefore, by subtracting the 0.82 percent from the 4.16 percent historical equity risk premium estimate, we obtain the forward-looking equity risk premium estimate of 3.34 percent. Adding this ERP estimate to the 4.13 percent bond yield, we estimate the forward-looking equity nominal compounded return to be 7.61 percent. In other words, we expect stocks to beat bonds by 3.34 percent per year over the next 20 years.

At the end of 2010, the 20-year Treasury inflation index yield was 1.64 percent, the nominal 20-year bond yield was 4.13 percent, and expected inflation was 2.45 percent. Therefore, the forecasted real stock return is 5.04 percent—again outperforming the forecasted real bond return of 1.64 percent by 3.34 percent compounded per year. The final results are presented in Table 4 and **Table 5**.

Implications for the Investor

For the long-term investor, asset allocation is the primary determinant of the variability of returns. Of all the decisions investors make, therefore, the asset allocation decision is the most important.

[3]Effective March 2009, Ibbotson Associates began using a blend of operating and reporting earnings for the period 1988 to the present when calculating P/Es. This approach mitigates the impact of severe write-downs of reported earnings and the resulting P/Es.

Table 5. Expected Return (20-Year Horizon), 31 December 2010

Benchmark	Geometric Return	Standard Deviation
Stocks (S&P 500)	7.61%	20.39%
BarCap U.S. Aggregate	4.45	6.59
Ibbotson U.S. Long-Term Government	4.13	11.73
Ibbotson U.S. Intermediate-Term Government	3.61	6.59
T-bills	2.49	3.43

The most important asset allocation decision is the allocation between stocks and bonds. Thus, the expected return between stocks and bonds, or the equity risk premium, is the most important number. A negative ERP implies that the investor should favor allocations to fixed income, whereas a positive ERP indicates an allocation to equities. (Of course, in addition to the ERP, the investor's risk tolerance, investment goals, time horizon, etc., need to be considered.) Therefore, the asset allocation decision is only as good as the accuracy of the investor's forecast of the expected equity risk premium.

Ibbotson Associates believes that stocks will continue to provide significant returns over the long run. We calculate the geometric, or the compounded, ERP based on applying the supply-side earnings model with three-year average earnings to be 3.34 percent—82 bps lower than the straight historical estimate. This forecast for the market is in line with both the historical supply measures of public corporations (i.e., earnings) and overall economic productivity (GDP per capita).

Conclusion

Not only have bonds outperformed stocks over various recent periods because of the financial crisis, but they also have roughly matched stock performance over the past 40 years. This fact raises the question, will bonds continue to outperform stocks?

This paper demonstrated that a close examination of history shows that stock returns over the last 40 years were virtually in line with the long-term historical average. Bond returns, however, were not only much higher than their historical averages but also higher than their current yields. This high bond return is the result of high interest rates in the 1970s and a subsequent declining interest rate environment. Given today's low-interest-rate environment, this scenario for bonds is very unlikely to repeat itself in the future. Investors hoping that bonds will outperform stocks in the coming years are likely to be disappointed.

Stocks tend to outperform bonds over time but are much riskier, even over longer periods. Bonds can outperform stocks over a long period, but investors need almost perfect timing to get in and out of the market to realize such returns. Ibbotson Associates believes the right strategy is to follow a disciplined asset allocation policy that considers the return–risk trade-offs by taking advantage of the diversification benefits over time provided by investing in both stocks and bonds.

Ibbotson Associates, Inc., is a registered investment advisor and wholly owned subsidiary of Morningstar, Inc. The Ibbotson name and logo are either trademarks or service marks of Ibbotson Associates, Inc. The information contained in this document is for informational purposes only and is the proprietary material of Ibbotson Associates. Reproduction, transcription, or other use, by any means, in whole or in part, without the prior written consent of Ibbotson, is prohibited. Opinions expressed are as of the current date; such opinions are subject to change without notice. Ibbotson Associates, Inc., shall not be responsible for any trading decisions, damages, or other losses resulting from, or related to, the information, data, analyses or opinions or their use.

References

Goetzmann, William N., Roger G. Ibbotson, and Liang Peng. 2001. "A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability." *Journal of Financial Markets*, vol. 4, no. 1 (January):1–32.

Ibbotson Associates. 2011. *Stocks, Bonds, Bills, and Inflation 2010 Yearbook*. Chicago: Ibbotson Associates.

Ibbotson, Roger G., and Peng Chen. 2003. "Long-Run Stock Returns: Participating in the Real Economy." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):88–98.

Ibbotson, Roger G., and Rex A. Sinquefield. 1976. "Stocks, Bonds, Bills, and Inflation: Simulations of the Future (1976–2000)." *Journal of Business*, vol. 49, no. 3 (July):313–338.

Price-to-Earnings Ratios: Growth and Discount Rates

Andrew Ang
Ann F. Kaplan Professor of Business
Columbia University

Xiaoyan Zhang
Associate Professor of Finance
Krannert School of Management, Purdue University

In a present-value model, movements in price-to-earnings ratios must reflect variations in discount rates (which embed risk premiums) and growth opportunities (which involve the cash flow and earnings-generating capacity of the firm's investments).[1] We decomposed P/Es into a no-growth value, defined to be the perpetuity value of future earnings that are held constant with full payout of earnings, and the present value of growth opportunities (PVGO), which is the value of the stock in excess of the no-growth value. To accomplish this decomposition, we used a dynamic model that accounts for time-varying risk premiums and stochastic growth opportunities.

An important aspect of our work is that we took into account a stochastic investment opportunity set with time-varying growth and discount rates. P/Es can be high not only when growth opportunities are perceived to be favorable but also when expected returns are low. For example, during the late 1990s and early 2000s, P/Es were very high. The cause might have been high prices incorporating large growth opportunities, but Jagannathan, McGrattan, and Scherbina (2000) and Claus and Thomas (2001), among others, have argued that during this time, discount rates were low. In contrast to our no-growth and PVGO decompositions, in which both discount rates and growth rates are stochastic, in the standard decompositions of no-growth and PVGO components, discount rates and growth rates are constant. Other standard analyses in the industry, such as the ratio of the P/E to growth (often called the "PEG ratio"), implicitly assign all variations in P/Es to growth opportunities because the analyses do not allow for time-varying discount rates.

---

[1]This approach decomposes the value of a firm into the value of its assets in place plus real options (or growth opportunities). This decomposition was recognized as early as Miller and Modigliani (1961).

 ©2011 The Research Foundation of CFA Institute

Static Case

An instructive approach is to consider first the standard decomposition of the P/E into the no-growth and growth components that is typically done in an MBA-level finance class. The exposition here is adapted from Bodie, Kane, and Marcus (2009, p. 597).

Suppose earnings grow at rate g, the discount rate is δ, and the payout ratio is denoted by po. The value of equity, P, is then given by

$$P = \frac{EA \times po}{\delta - g}, \tag{1}$$

where EA is expected earnings next year. The P/E—that is, $P/E = P/EA$—is then simply

$$\begin{aligned} P/E &= \frac{P}{EA} \\ &= \frac{po}{\delta - g}. \end{aligned} \tag{2}$$

We can decompose market value P into a no-growth component and a growth component. The growth component is considered to be the PVGO. The no-growth value, P^{ng}, is defined as the present value of future earnings with no growth (so, $g = 0$ and $po = 1$):

$$P^{ng} = \frac{EA}{\delta}. \tag{3}$$

The growth component is defined as the remainder:

$$\begin{aligned} PVGO &= \frac{EA \times po}{\delta - g} - \frac{EA}{\delta} \\ &= \frac{EA\left[g - (1 - po)\delta\right]}{\delta(\delta - g)}, \end{aligned} \tag{4}$$

and the two sum up to the total market value:

$$P = P^{ng} + PVGO. \tag{5}$$

The decomposition of firm value into no-growth and PVGO components is important because, by definition, the no-growth component involves only discount rates whereas the PVGO component involves both the discount rate and the effects of cash flow growth. Understanding which component dominates gives insight into what drives P/Es. The static case cannot be used to decompose P/Es into no-growth and PVGO values over time, however, because it assumes that earnings growth (g), discount rates (δ), and payout ratios (po) remain constant over time. Clearly, this assumption is not true. Thus, to examine the no-growth and PVGO values of P/Es, we need to build a dynamic model.

The Dynamic Model

We made two changes to the static case to handle time-varying investment opportunities. First, we put "t" subscripts on the variables to indicate that they change over time. Second, for analytical tractability, we worked in log returns, log growth rates, and log payout ratios.

We defined the discount rate, δ_t, as

$$\delta_t = \ln E_t\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right), \tag{6}$$

where P_t is the equity price at time t and D_t is the dividend at time t. Earnings growth is defined as

$$g_t = \ln\left(\frac{EA_t}{EA_{t-1}}\right), \tag{7}$$

where EA_t is earnings at time t. Finally, the log payout ratio at time t is

$$po_t = \ln\left(\frac{D_t}{EA_t}\right). \tag{8}$$

In this notation, if $\delta_t = \bar{\delta}$, $g_t = \bar{g}$, and $po_t = \overline{po}$ are all constant, then the familiar P/E in Equation 2 written in simple growth rates or returns becomes

$$\frac{P}{EA} = \frac{\exp\left(\overline{po}\right)}{\exp\left(\bar{\delta} - \bar{g}\right) - 1}. \tag{9}$$

Factors. We specified factors X_t that drive P/Es. The first three factors in X_t are the risk-free rate, r_t^f; the earnings growth rate, g_t; and the payout ratio, po_t. We included two other variables that predict returns: the growth rate of industrial production, ip_t, and term spreads, $term_t$. We selected these variables after considering variables that, on their own, forecast total returns, earnings growth, or both. We also included a latent factor, f_t, that captures variation in expected returns not accounted for by the observable factors. We specified latent factor f_t to be orthogonal to the other factors. Thus, $X_t = (r_t^f \; g_t \; po_t \; ip_t \; term_t \; f_t)'$.

We assumed that state variables X_t follow a vector autoregression (VAR) with one lag:

$$X_{t+1} = \boldsymbol{\mu} + \Phi X_t + \Sigma\varepsilon_{t+1}, \tag{10}$$

 ©2011 The Research Foundation of CFA Institute

where ε_t follows a standard normal distribution with zero mean and unit standard deviation. The companion form, Φ, allows earnings growth and payout ratios to be predictable by both past earnings growth and payout ratios and other macro variables.

The long-run risk model of Bansal and Yaron (2004) incorporates a highly persistent factor in the conditional mean of cash flows. Our model accomplishes the same effect by including persistent variables in X_t, especially the risk-free rate and payout ratio, which are both highly autocorrelated.

To complete the model, we assumed that discount rates δ_t are a linear function of state variables X_t:

$$\delta_t = \delta_0 + \boldsymbol{\delta}_1' X_t. \tag{11}$$

Equation 11 subsumes the special cases of constant total expected returns by setting $\delta_1 = 0$ and subsumes the general case of time-varying discount rates when $\delta_1 \neq 0$. Because f_t is latent, we placed a unit coefficient in δ_1 that corresponds to f_t for identification.

The Dynamic P/E. Under the assumptions shown in Equation 10 and Equation 11, the dynamic P/E can be written as

$$P/E_t = \sum_{i=1}^{\infty} \exp\left(a_i + \boldsymbol{b}_i' X_t\right). \tag{12}$$

The coefficients a_i and b_i are given in Appendix A.[2]

Our model of the P/E belongs to the asset-pricing literature that builds dynamic valuation models. The approaches of Campbell and Shiller (1988) and Vuolteenaho (2002) to model the price/dividend ratio (P/D) and the P/E, respectively, require log-linearization assumptions. In contrast, our model produces analytically tractable solutions for P/Es. Recently, Bekaert, Engstrom, and Grenadier (2010) and van Binsbergen and Koijen (2010) examined dynamic P/Ds, but not P/Es, in models with closed-form solutions. Our model is more closely related to the analytical dynamic earnings models of Ang and Liu (2001) and Bakshi and Chen (2005), in which cash flows are predictable and discount rates vary over time. Ang and Liu, however, modeled price-to-book ratios instead of P/Es, and Bakshi and Chen's model of the P/E requires the payout ratio to be constant.

Growth and No-Growth Components. The no-growth P/E can be interpreted as a perpetuity, where at each time, a unit cash flow is discounted by the cumulated market discount rates prevailing up until that time. In the full P/E in Equation 12, growth occurs by plowing earnings back into the firm. In the no-growth P/E, earnings are fully paid out; consequently, the payout ratio

[2]A full derivation is available in the online appendix at www.columbia.edu/~aa610.

does not directly influence the no-growth P/E value. The payout ratio is relevant in the no-growth P/E, however, because the payout ratio is a state variable and its dynamics are allowed to influence future earnings through the VAR process.

The no-growth P/E, P/E_t^{ng}, where earnings growth is everywhere 0 and the payout ratio is equal to 1, can be written as

$$P/E_t^{ng} = \sum_{i=1}^{\infty} \exp\left(a_i^* + \boldsymbol{b}_i^{*\prime} X_t\right), \tag{13}$$

where a_i^* and b_i^* are given in Appendix A.

The present value of growth opportunities is defined as the difference between the P/E, which incorporates growth, and the no-growth P/E:

$$P/E_t = P/E_t^{ng} + PVGO_t. \tag{14}$$

Empirical Results

We used data on dividend yields, P/Es, price returns (capital gains only), and total returns (capital gains and dividends) on the S&P 500 Index from the first quarter (Q1) of 1953 to the fourth quarter (Q4) of 2009.

Panel A of **Figure 1** plots the log index of the S&P 500 Total Return Index across our sample. The decline during the mid-1970s recession, the strong bull market of the 1990s, the decline after the technology bubble in the early 2000s, and the drop resulting from the 2008–09 financial crisis are clearly visible. Panel B graphs the P/E, which averages 18.5 over the sample period. The P/E suddenly increased in Q4:2008 to 60.7 and reached a peak of 122 in Q2:2009. In Q4:2009, the P/E came down to 21.9. The large increase in the P/E from Q4:2008 through Q3:2009 is the result of large, negative reported earnings in Q4:2008 during the financial crisis. This development caused the moving four-quarter average of earnings to sharply decrease. While prices were declining during the financial crisis, an even greater decrease was occurring in reported earnings, which caused the increase in the P/E. Panel C of Figure 1 reports S&P 500 dividend yields, which reached a low at the end of the bull market in 2000.

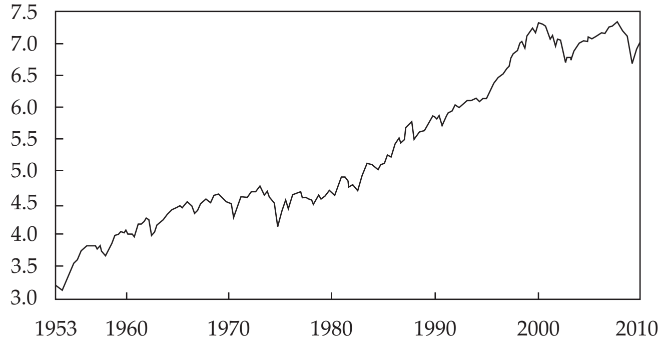
Estimation Results. **Table 1** reports the parameter estimates of the model. The two most significant predictors of the discount rate are earnings growth, g, with a coefficient of 0.38, and the growth rate of industrial production, ip, with a coefficient of −1.28. The estimated VAR parameters show that all factors are highly persistent, and this persistence dominates: No other factor except the variables themselves Granger-causes risk-free rates, earnings growth, or payout ratios.[3]

[3] Estimation of the model is discussed in the online appendix at www.columbia.edu/~aa610.

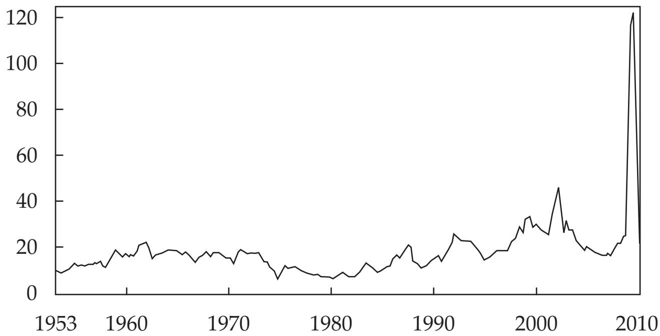
 ©2011 The Research Foundation of CFA Institute

Figure 1. Log Index Levels, Payout Ratios, and Dividend Yields for S&P 500 Total Return Index, Q1:1953–Q4:2009

A. Log of the Index Level



B. P/E



C. Dividend Yield

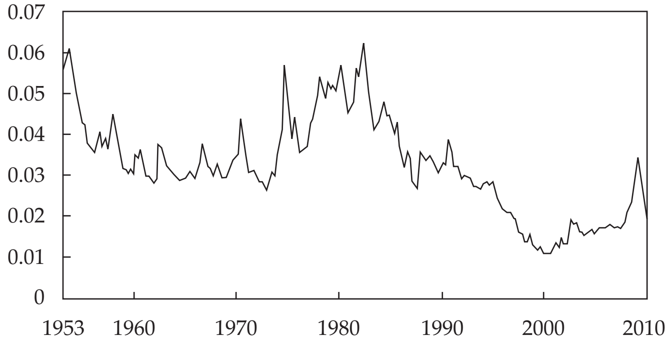


Table 1. Parameter Estimates
(*p*-values in parentheses)

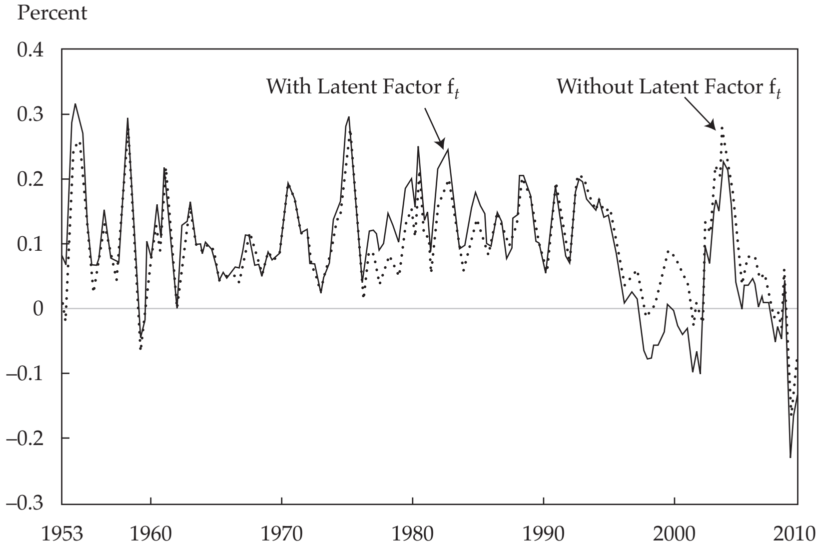
	rf	g	po	ip	$term$	f
Discount rate parameters δ_1:	0.325	0.381	0.164	−1.283	1.203	1
	(0.775)	(0.121)	(0.088)	(0.238)	(1.728)	—
VAR parameter Φ						
rf	0.863	0.26	0.012	−0.005	0.088	0
	(0.089	(0.008)	(0.012)	(0.033)	(0.191)	—
g	0.917	0.628	0.650	0.115	3.677	0
	(1.385)	(0.353)	(0.426)	(0.362)	(3.446)	—
po	−0.771	−0.514	0.303	0.045	−2.805	0
	(1.292)	(0.328)	(0.415)	(0.360)	(3.131)	—
ip	−0.244	0.096	0.071	−0.169	0.908	0
	(0.237)	(0.057)	(0.041)	(0.108)	(0.737)	—
$term$	0.021	−0.017	−0.003	−0.025	0.502	0
	(0.036)	(0.005)	(0.007)	(0.019)	(0.092)	—
f	0	0	0	0	0	0.904
	—	—	—	—	—	(0.003)

We plotted the estimated discount rates in **Figure 2**. The full discount rate (solid line) is overlaid with the implied discount rate without the latent factor, f_t (dotted line). The two discount rates have a correlation of 0.91. Thus, the observable factors capture most of the variation in expected returns. Without the latent factor, the observable factors $z_t = (r_t^f \; g_t \; po_t \; ip_t \; term_t)$ account for 18.0 percent of the variance of total returns; adding the latent factor brings the proportion up to 27.5 percent.

Figure 2 shows that discount rates declined noticeably in the 1990s—from 14.5 percent in Q1:1991 to −14.5 percent in Q1:2002. The −14.5 percent corresponds to what was at that time the all-time-high P/E in the sample, 46.5. The latent factor was very negative during this time; the model explains the high P/E as coming from low discount rates. Recently, during the financial crisis, discount rates were again negative. For example, in Q4:2008, the discount rate was −16.3 percent. Q4:2008 was characterized by pronounced negative reported earnings. The P/E increased to 60.7 at this time because of the low earnings relative to market values. The model again explains the high P/E by the low discount rate. The low discount rates at this time were caused by the large decrease in earnings growth. Subsequent returns over the 2008–09 period were indeed extremely low.

©2011 The Research Foundation of CFA Institute

Figure 2. Discount Rates, Q1:1953–Q4:2009



Drivers of the P/E. In Table 2, we report variance decompositions of the P/E. We computed the variance of the P/E implied by the model through the sample, where the factor z was held constant at its unconditional mean, $\mathrm{var}_z(P/E)$. The variance decomposition resulting from factor z is given by $1 - \mathrm{var}_z(P/E)/\mathrm{var}(P/E)$, where $\mathrm{var}(P/E)$ is the variance of the P/E in the data. These decompositions do not sum to 1.0 because the factors are correlated. Table 2 shows that the macro variables play a large role in explaining the dynamics of P/Es. Risk-free rates, earnings growth, and payout ratios explain, respectively, 18 percent, 38 percent, and 66 percent of the variance of P/Es.

Table 2. Variance Decompositions of the P/E

Parameter	Variance Explained
rf	17.8%
g	38.3
po	65.9
ip	–38.6
term	7.5
f	70.5

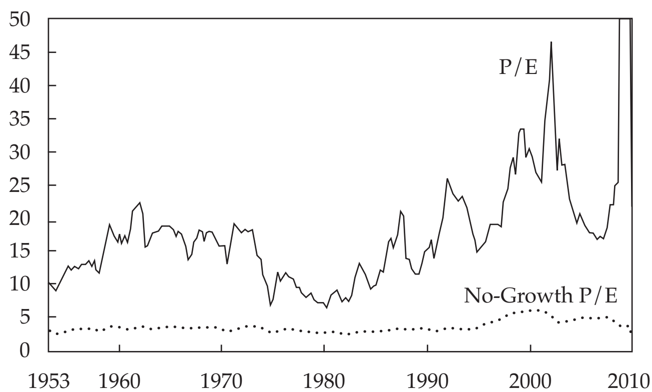
The variance attribution for growth in industrial production is negative because diminished industrial production results in more volatile discount rates and greater volatility of P/Es. The latent factor, f, plays an important role in matching P/Es, with a variance attribution of 71 percent. This finding is consistent with Figure 2, where some occasionally pronounced differences are visible between discount rates produced only with macro variables and discount rates estimated with the latent factor.

Growth and No-Growth Decompositions. Figure 3 plots the no-growth components together with the P/E. Most of the variation in the P/E is a result of growth components. The average no-growth P/E defined in Equation 13 is 3.8, compared with an average P/E in the data of 18.5. Thus, no-growth components account for, on average, 20.7 percent of the P/E; most of the total P/E is a result of the PVGO. The no-growth component is remarkably constant (as is clearly shown in Figure 3) and has a volatility of 0.853, compared with a volatility of 12.7 for the P/E. A variance decomposition of the P/E is

$$\underset{100\%}{\operatorname{var}\left(P/E_t\right)} = \underset{0.5\%}{\operatorname{var}\left(P/E_t^{ng}\right)} + \underset{94.8\%}{\operatorname{var}\left(PVGO_t\right)} + \underset{4.7\%}{2\operatorname{cov}\left(P/E_t^{ng}, PVGO_t\right)}. \tag{15}$$

Thus, 95 percent of P/E variation is explained by growth components, or the PVGO term. The perpetuity value of no-growth is relatively constant because discount rates are highly mean reverting: The year-on-year autocorrelation of discount rates over the sample is 0.34. Thus, the discounted earnings in the no-growth P/E rapidly revert to their long-term average.

Figure 3. No-Growth and Growth Components of the P/E, Q1:1953–Q4:2009



 ©2011 The Research Foundation of CFA Institute

In **Table 3**, we report various correlations of the no-growth and PVGO P/Es. The no-growth and PVGO components have a correlation of 0.363, but this correlation has only a small effect on total P/E variation because of the low volatility of no-growth P/E values. Thus, most of the variation in the total P/E is caused by growth opportunities, and not surprisingly, the PVGO P/E and the total P/E are highly correlated, at 0.998. Both the growth P/E and the total P/E decrease when risk-free rates and earnings growth increase. The correlation of the total P/E with earnings growth is particularly strong at −0.766. High earnings growth by itself increases earnings, which is the denominator of the P/E, and causes P/Es to decrease, resulting in the high negative correlation between earnings growth and the P/E. But another discount rate effect occurs because high earnings growth causes discount rates to significantly increase (see Table 1). This effect also causes P/Es to decrease. High payout ratios, as expected, are positively correlated with the P/E at 0.713. Finally, the latent factor, f, is negatively correlated with the P/E because it is only a discount rate factor: By construction, P/Es are high when f is low.

Table 3. Correlation of Growth (PVGO) and No-Growth Components of the P/E

	No Growth P/E	PVGO P/E
PVGO P/E:	0.363	
Data P/E:	0.421	0.998
rf	−0.353	−0.426
g	−0.051	−0.766
po	−0.292	0.713
ip	0.114	−0.303
term	0.027	0.390
f	−0.903	−0.538

Conclusion

We decomposed the P/E into a no-growth component (the perpetuity value of future earnings held constant with full payout) and a component termed PVGO that reflects the growth opportunities and real options a firm has to invest in the future. We valued both components in a dynamic stochastic environment where risk premiums and earnings growth are stochastic. We found that discount rates exhibit significant variation: 27.5 percent of the variation in total returns is caused by persistent, time-varying expected return components. However, although the variation of discount rates is large, these rates are highly

©2011 The Research Foundation of CFA Institute

mean reverting. The result is that the no-growth value of earnings exhibits relatively little volatility. The PVGO component dominates; it accounts for the bulk of the level and variation of P/Es in the data: Approximately 80 percent of the level and 95 percent of the variance of P/Es are a result of time-varying growth opportunities.

We thank Geert Bekaert, Sigbjørn Berg, and Tørres Trovik for helpful discussions.

Appendix A

Here, we provide the coefficients a_i and b_i and the definition of the P/E as used by the S&P 500. All the formulas are derived in the online appendix at www.columbia.edu/~aa610.

Full and No-Growth P/Es. The coefficients a_i and b_i for the P/E in Equation 12 are given by

$$a_{i+1} = -\delta_0 + a_i + (\mathbf{e}_2 + \boldsymbol{b}_i)' \boldsymbol{\mu} + \frac{1}{2}(\mathbf{e}_2 + \boldsymbol{b}_i)' \Sigma\Sigma' (\mathbf{e}_2 + \boldsymbol{b}_i)$$

and

$$b_i = -\boldsymbol{\delta}_1 + \Phi'(\mathbf{e}_2 + \boldsymbol{b}_i),$$

where $\mathbf{e}_n$ is a vector of 0s with a 1 in the nth position. The initial conditions are

$$a_1 = -\delta_0 + (\mathbf{e}_2 + \mathbf{e}_3)' \boldsymbol{\mu} + \frac{1}{2}(\mathbf{e}_2 + \mathbf{e}_3)' \Sigma\Sigma' (\mathbf{e}_2 + \mathbf{e}_3)$$

and

$$b_1 = -\boldsymbol{\delta}_1 + \Phi'(\mathbf{e}_2 + \mathbf{e}_3).$$

The coefficients in the no-growth P/E, P/E_t^{ng}, in Equation 13 are given by

$$a_{i+1}^* = -\delta_0 + a_i^* + \boldsymbol{b}_i^{*\prime} \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{b}_i^{*\prime} \Sigma\Sigma' \boldsymbol{b}_i^*$$

and

$$\boldsymbol{b}_{i+1}^* = -\boldsymbol{\delta}_1 + \Phi' \boldsymbol{b}_i^*,$$

where a_i^* and b_i^* have initial values $a_i^* = -\delta_0$ and $\boldsymbol{b}_i^* = -\delta_{\mathbf{1}}$.

©2011 The Research Foundation of CFA Institute

Data. The P/E defined by Standard & Poor's is the market value at time t divided by trailing 12-month earnings reported from t to $t - 1$. To back out earnings growth from P/Es, we used the following transformation:

$$\begin{aligned}\exp\left(g_{t+1}\right) &= \frac{EA_{t+1}}{EA_t} \\ &= \left(\frac{P/E_t}{P/E_{t+1}}\right)\left(\frac{P_{t+1}}{P_t}\right),\end{aligned}$$

where P_{t+1}/P_t is the price gain (capital gain) on the market from t to t +1.

The dividend yield reported by Standard & Poor's is also constructed from trailing 12-month summed dividends. We computed the log payout ratio from the ratio of the dividend yield, $dy_t = D_t/P_t$, to the inverse P/E:

$$\begin{aligned}\exp\left(po_t\right) &= \frac{dy_t}{1/\left(P/E\right)_t} \\ &= \frac{D_t}{EA_t}.\end{aligned}$$

For the risk-free rate, r_t^f, we used one-year zero-coupon yields expressed as a log return, which we obtained from the Fama Files derived from the CRSP U.S. Government Bond Files. For the macro variables, we expressed industrial production growth, *ip*, as a log year-on-year growth rate for which we used the industrial production index from the St. Louis Federal Reserve. We defined the term spread, *term*, as the difference in annual yields between 10-year and 1-year government bonds, which we obtained from CRSP.

Bibliography

Ang, A., and J. Liu. 2001. "A General Affine Earnings Valuation Model." *Review of Accounting Studies*, vol. 6, no. 4 (December):397–425.

Bakshi, G., and Z. Chen. 2005. "Stock Valuation in Dynamic Economies." *Journal of Financial Markets*, vol. 8, no. 2 (May):115–151.

Bansal, R., and A. Yaron. 2004. "Risk for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, vol. 59, no. 4 (August):1481–1509.

Bekaert, G., E. Engstrom, and S.R. Grenadier. 2010. "Stock and Bond Returns with Moody Investors." *Journal of Empirical Finance*, vol. 17, no. 5 (December):867–894.

Bodie, Z., A. Kane, and A.J. Marcus. 2009. *Investments*. 8th ed. New York: McGraw-Hill/Irwin.

Campbell, J.Y., and R.J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, vol. 1, no. 3 (July):195–228.

Claus, J., and J. Thomas. 2001. "Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets." *Journal of Finance*, vol. 56, no. 5 (October):1629–1666.

Fama, E.F., and K.R. French. 2002. "The Equity Premium." *Journal of Finance*, vol. 57, no. 2 (April):637–659.

Jagannathan, R., E.R. McGrattan, and A. Scherbina. 2000. "The Declining U.S. Equity Premium." *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 24, no. 4 (Fall):3–19.

Miller, M.H., and F. Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *Journal of Business*, vol. 34, no. 4 (October):411–433.

van Binsbergen, J., and R.S.J. Koijen. 2010. "Predictive Regressions: A Present-Value Approach." *Journal of Finance*, vol. 65, no. 4 (August):1439–1471.

Vuolteenaho, T. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance*, vol. 51, no. 1 (February):233–264.

 ©2011 The Research Foundation of CFA Institute

Long-Term Stock Returns Unshaken by Bear Markets

Jeremy J. Siegel
Russell E. Palmer Professor of Finance
Wharton School of the University of Pennsylvania

The first Equity Risk Premium Forum, sponsored by CFA Institute, was held on 8 November 2001, not long after the September 11 terrorist attacks and coincident with the first of two devastating bear markets in the first decade of the new millennium. At the time of the first forum, stocks had already fallen by more than half of what would become a nearly 50 percent decline from the peak reached in March 2000 to the low in October 2002. Over the four years after the low, the equity market recovered all of its losses and moved into new all-time-high territory. But the 2008 financial crisis precipitated a more severe bear market than 2000–2002 and the worst since the Great Crash of 1929–1932. In the financial crisis, the S&P 500 Index plunged 57 percent from October 2007 to March 2009 and non-U.S. equity markets fell more than 60 percent. As of this writing (May 2011), stocks worldwide have made a strong recovery and are now within 15 percent of their all-time highs.

Nevertheless, the returns for stocks during the past decade have not been good. Since the first forum was held, the stock returns on the broad-based Russell 3000 Index have averaged 5.6 percent per year; when offset against 2.5 percent annual inflation, the real return is only a little more than 3 percent per year. The nominal yields on Treasuries have averaged 2.2 percent during the decade, leaving a real return of –0.2 percent per year on those instruments. These returns mean that the realized equity premium, or excess return of stocks over T-bills, has been between 3 percent and 3.5 percent. These numbers are not far from the predictions that I made at the first forum 10 years ago. At the time, I expected real returns of equities to be 4.5–5.5 percent and an equity risk premium of 2 percent (200 bps).

As I read through my analysis from 10 years ago, I could see that the main reason I overestimated the real return on stocks was that I overestimated the price-to-earnings ratio (P/E) that investors would pay for stocks. There were good reasons back then for why the P/E of stocks should be higher than its historical average of 15, a level computed from earnings data extending back to 1871, and should instead range between 20 and 25. First, the sharp decline in transaction costs caused by the development of index funds and the plunge in commission prices gave investors a much more favorable realized risk–return

©2011 The Research Foundation of CFA Institute

trade-off than they received in earlier years. Another reason I conjectured that the P/E would be higher than its historical level was the decline in the volatility of real economy variables. This increase in macroeconomic stability was termed by economists at the time as the "Great Moderation."

Of course, the 2007–09 recession dispelled the idea that the business cycle had been tamed. It is my opinion that the Great Moderation was indeed real, but the long period of macroeconomic stability led to an excessive decline in risk premiums, particularly in housing-related securities. So, when real estate prices unexpectedly fell, the entire financial system came crashing down. The financial crisis greatly increased the risk aversion of investors, and that result brought the P/E back down to historical levels and led to the poor stock returns of the past decade.

This observation can be confirmed by examining the data. When the first forum was held in November 2001, the reported earnings of the S&P 500 over the preceding 12 months were $15.90, which yielded a P/E of 36.77. The trailing 12-month earnings on the S&P 500 at the time of the second forum in January 2011 were $81.47, more than a threefold increase. Yet the index itself was up by only 30 percent, and the P/E had fallen to 16.66. If the P/E had fallen only to 22.5, the middle of my valuation range, stock returns would have been about 3 percentage points per year higher.

Another prediction that did not materialize was my estimate of future bond yields. I believed that the real yields on bonds would remain between 3 and 4 percent, the level that prevailed when Treasury Inflation-Protected Securities (TIPS) were first issued in 1997. I also believed that the realized bond returns in the period after World War II (WWII) were biased downward because of the unanticipated inflation from the late 1960s through the early 1980s. So, I did not consider historical returns on bonds; instead, I used the current yield on TIPS in making my forecast for future bond yields.

Instead, real yields fell dramatically, especially in the wake of the financial crisis. As of early 2011, 10-year TIPS yields are less than 1 percent and 5-year TIPS yields are negative. The two primary reasons for the drop in real yields are the slowdown in economic growth and the increase in the risk aversion of the investing public, which, in turn, is caused by both the aging of the population and the shocks associated with the financial crisis. The decline in inflation has caused the yields on nominal bonds to drop even more, generating very large realized returns for nominal bond investors. Over the last decade, realized bond returns were 4.7 percent per year after inflation, swamping stock returns. Over the past 20 years, realized bond returns were 6.0 percent per year, 1 percentage point less than the 7.0 percent real returns on stocks.

 ©2011 The Research Foundation of CFA Institute

Updated Return Data

Table 1 shows historical returns for stocks, bonds, and T-bills from 1802 through April 2011. The past decade has shaved one-tenth of a percent off of the annualized real returns on stocks from 1802 through April 2001; three-tenths off of the equity returns from 1871, which is when the Cowles Foundation for Research in Economics data became available; and five-tenths off of the real return since 1926, which is the period that Ibbotson and Sinquefield popularized in their research.[1] Over all long-term periods, the real return on stocks remained in the 6–7 percent range. Over the past 30 years, the real annual return on stocks has been 7.9 percent, and over the past 20 years, the real return has been 7.0 percent. In fact, the numbers that now fill the table are almost identical to those that I calculated when I started my research in the late 1980s. In essence, the poor returns of the past 10 years just offset the very high returns of the previous decade.

Table 2 summarizes some of the important statistics about the equity market, such as the P/E, earnings growth, and dividend growth, for 1871–April 2011. The average P/E has changed very little over the past decade. In the version of Table 2 prepared for the 2001 forum, the average P/E was 14.45; adding the subsequent 10 years of data increased it by 0.06 to 14.51. The earnings yield, which is the reciprocal of the P/E, obviously also changes little.

One important issue that was in contention in the first forum is still debated today. Finance theory, particularly that of Modigliani and Miller (M&M), predicts that when the dividend payout ratio declines, the dividend yield will also decline, but this decline will be offset by an increase in the growth rate of future earnings and dividends.[2] Cliff Asness, at the 2001 forum, and Rob Arnott, at the most recent forum, cite research, which they performed together, that suggests that a lower payout ratio, in contrast to what finance theory would predict, does not actually lead to faster earnings growth.[3] At the first forum, I claimed that this finding was a result of the cyclical behavior of earnings. Asness and Arnott claimed to have run further tests to contest this point. Notwithstanding their results, my data clearly show that over long periods of time, the payout ratio is inversely correlated with dividend and earnings growth as predicted by finance theory.

---

[1] Roger G. Ibbotson and Rex A. Sinquefield, "Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926–1974)," *Journal of Business*, vol. 49, no. 1 (January 1976):11–47.

[2] Franco Modigliani and Merton H. Miller, "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, vol. 48, no. 3 (June 1958):261–297.

[3] Robert D. Arnott and Clifford S. Asness, "Surprise! Higher Dividends = Higher Earnings Growth," *Financial Analysts Journal*, vol. 59, no. 1 (January/February 2003):70–87.

Table 1. Historical Returns for Stocks, Bonds, and T-Bills, 1802–April 2011

	Real Return						Stocks' Excess Return Over			
	Stocks		Bonds		T-Bills		Bonds		T-Bills	
	Geometric	Arithmetic	Geometric	Arithmetic	Geometric	Arithmetic	Geometric	Arithmetic	Geometric	Arithmetic
Periods										
1802–2011	6.7%	8.2%	3.6%	3.9%	2.7%	2.9%	3.1%	4.3%	3.9%	5.3%
1870–2011	6.5	8.2	3.0	3.3	1.6	1.7	3.5	4.9	4.9	6.5
Major subperiods										
1802–1870	7.0%	8.3%	4.8%	5.1%	5.1%	5.4%	2.2%	3.2%	1.9%	2.9%
1871–1925	6.6	7.9	3.7	3.9	3.2	3.3	2.9	4.0	3.5	4.7
1926–2011	6.4	8.4	2.5	3.0	0.6	0.7	4.0	5.4	5.8	7.7
After World War II										
1946–2011	6.4%	8.3%	1.8%	2.2%	0.5%	0.6%	4.6%	6.0%	6.0%	7.6%
1946–1965	10.0	11.4	–1.2	–1.0	–0.8	–0.7	11.2	12.3	10.9	12.1
1966–1981	–0.4	1.4	–4.2	–3.9	–0.2	–0.1	3.8	5.2	–0.2	1.5
1982–1999	13.6	14.3	8.5	9.3	2.9	2.9	5.1	5.0	10.7	11.4
1982–2011	7.9	9.1	7.5	7.9	1.8	1.7	0.4	1.2	6.1	7.4
1991–2011	7.0	8.5	6.0	6.3	0.9	0.9	0.9	2.1	6.1	7.6
2001–2011	0.8	2.8	4.7	4.7	–0.3	–0.3	–4.0	–1.9	1.1	3.0

©2011 The Research Foundation of CFA Institute

Table 2. Historical Equity Market Statistics, 1871–April 2011

	Real Stock Return	Average P/E	Inverse of Average P/E	Real Earnings Growth	Real Dividend Growth	Dividend Yield	Real Capital Gains	Average Payout Ratio
1871–2011	6.51%	14.51	6.89%	1.81%	1.22%	4.47%	1.55%	59.92%
1871–1945	6.39	13.83	7.23	0.67	0.74	5.31	1.11	70.81
1946–2011	6.44	15.29	6.54	3.14	1.76	3.50	2.85	47.42

In fact, the evidence in favor of M&M has been strengthened by the addition of the past 10 years of data. In the 1871–1945 data, annual real per share earnings growth was only 0.67 percent per year and the payout ratio averaged nearly 72 percent. In the post-WWII period, real earnings growth was 3.14 percent and the payout ratio was only 47.42 percent.[4]

It is true that adding the past 10 years increases post-WWII real per share dividend growth only marginally because the payout ratio is still declining and has not yet reached a new "steady state" in which dividend growth will increase to the level of earnings growth.

Projections for the Next Decade

I hope a third forum will be held in 2021 so we can look back on our predictions in 2011, either nursing our wounds or congratulating ourselves on our astuteness. Using the current P/E as a basis, I expect real stock returns to be between 6 and 7 percent. But I will not be surprised if they are higher because the same factors that influenced my prediction of P/Es in the range of 20–25 are as operative in 2011 as they were at the time of the first forum in 2001.

Real bond returns are on track to be much lower. Ten-year TIPS are now yielding about 1 percent, so the excess returns of stocks over bonds should be in the 5–6 percent range, which is higher than the historical average. And the bias, if any, will be toward a higher equity premium if real bond yields rise from their extremely low levels, as I think they should. In short, relative to bonds, stocks look extraordinarily attractive, and I expect stock investors will look back a decade from now with satisfaction.

[4]Note that the 3.14 percent growth rate is more than 1 percentage point higher than the post-WWII real earnings growth rate presented at the first forum; the addition of the past 10 years also reduces the post-WWII average payout ratio from 50.75 percent to 47.42 percent.

©2011 The Research Foundation of CFA Institute

The Equity Premium Puzzle Revisited

Rajnish Mehra

E.N. Basha Arizona Heritage Chair Professor of Finance and Economics, Arizona State University

Research Associate, NBER

In the two and a half decades since "The Equity Premium: A Puzzle" (Mehra and Prescott 1985) was published, attempts to successfully account for the equity premium have become a major research impetus in finance and economics. In an effort to reconcile theory with observations, I will elaborate on the appropriateness of three crucial abstractions in that article. In particular, I will argue that our finding (i.e., the premium for bearing nondiversifiable aggregate risk is small) is not inconsistent with the average equity premium over the past 120 years.

The three abstractions that I address here are

- using T-bill prices as a proxy for the expected intertemporal marginal rate of substitution of consumption;
- ignoring the difference between borrowing and lending rates (a consequence of agent heterogeneity and costly intermediation);
- abstracting from life-cycle effects and borrowing constraints on the young.

I examine each of these in detail below.

Using T-Bill Prices as a Proxy for the Expected Intertemporal Marginal Rate of Substitution of Consumption

An assumption implicit in Mehra and Prescott (1985) is that agents use both equity and the riskless asset to smooth consumption intertemporally. This assumption is a direct consequence of the first-order condition (see Equation 1) for the representative household in our model. It implies that agents save by optimally allocating resources between equity and riskless debt.

$$0 = E_t\left[\frac{U_c\left(c_{t+s}\right)}{U_c\left(c_t\right)}\left(r^e_{t,t+s} - r^d_{t,t+s}\right)\right]. \tag{1}$$

Author Note: This paper draws widely on my collaborations with George Constantinides, John Donaldson, and Edward Prescott. Quite independently of our joint work, they have made substantial contributions to the literature on the equity premium puzzle. Consequently, the views expressed in this paper do not necessarily reflect their views.

©2011 Rajnish Mehra

Equation 1 is the standard asset-pricing equation in macroeconomics and finance. $U_c(c_{t+s})$ is the marginal utility of consumption at time $t + s$; $r^e_{t,t+s}$ and $r^d_{t,t+s}$ are, respectively, the return on equity and the return on the riskless asset over the period t, $t + s$; and E_t is the expectation conditional on the agent's information set at time t.

If the results from the model are to be compared with data, it is crucial to identify the empirical counterpart of the riskless asset that is actually used by agents to smooth consumption. In Mehra and Prescott (1985), we used the highly liquid T-bill rate, corrected for expected inflation, as a proxy for this asset. But one might ask: Is it reasonable to assume that T-bills are an appropriate proxy for the riskless asset that agents use to save for retirement and smooth consumption? Do households actually hold T-bills to finance their retirement? *Only if this question is empirically verified would it be reasonable to equate their expected marginal rate of substitution of consumption to the rate of return on T-bills.*

This question cannot be answered in the abstract without reference to the asset holdings of households, so a natural next step is to examine the assets held by households. **Table 1** details these holdings for U.S. households. The four big asset-holding categories of households are tangible assets, pension and life insurance holdings, equity (both corporate and noncorporate), and debt assets.

Table 1. Household Assets and Liabilities as a Fraction/ Multiple of GDP
(average of 2000 and 2005)

Assets (GDP)		Liabilities (GDP)	
Asset	GDP (×)	Liability	GDP (×)
Tangible household	1.65	Liabilities	0.7
Corporate equity	0.85	Net worth	4.15
Noncorporate equity	0.5		
Pension and life insurance reserves	1.0		
Debt assets	0.85		
Total	4.85		4.85

In 2000, privately held government debt was only 0.30 times GDP, a third of which was held by foreigners. The amount of interest-bearing government debt with maturity less than a year was only 0.085 times GDP, which is a small fraction of total household net worth. Virtually no T-bills are directly owned by households.[1] Approximately one-third of the T-bills outstanding are held by foreign central banks, and two-thirds are held by U.S. financial institutions.

[1] See Table B-89, *Economic Report of the President* (2005).

Although large amounts of debt assets are held, most of these are in pension fund and life insurance reserves. Some are in demand deposits, for which free services are provided. Most government debt is held indirectly; a small fraction is held as savings bonds.

Thus, much of intertemporal saving is in debt assets, such as annuities and mortgage debt, held in retirement accounts and as pension fund reserves. Other assets, not T-bills, are typically held to finance consumption in retirement. *Hence, T-bills and short-term debt are not reasonable empirical counterparts to the risk-free asset priced in Equation 1*, and it would be inappropriate to equate the return on these assets to the expected marginal rate of substitution for an important group of agents.

An inflation-indexed, default-free bond portfolio with a duration similar to that of a well-diversified equity portfolio would be a reasonable proxy for a risk-free asset used for consumption smoothing.[2] For most of the 20th century, equity has had an implied duration of about 25 years, so a portfolio of TIPS (Treasury Inflation-Protected Securities) of a similar duration would be a reasonable proxy.

Because TIPS have only recently (1997) been introduced in U.S. capital markets, it is difficult to get accurate estimates of the mean return on this asset class. The average return for the 1997–2005 period is 3.7 percent. An alternative (though imperfect) proxy would be to use the returns on indexed mortgages guaranteed by Ginnie Mae (Government National Mortgage Association) or issued by Fannie Mae (Federal National Mortgage Association). I conjecture that if these indexed default-free securities are used as a benchmark, the equity premium will be closer to 4 percent than to the 6 percent equity premium relative to T-bills. By using a more appropriate benchmark for the riskless asset, I can account for 2 percentage points of the "equity premium."

Ignoring the Difference between Borrowing and Lending Rates

A major disadvantage of the homogeneous household construct is that it precludes the modeling of borrowing and lending among agents. In equilibrium, the shadow price of consumption at date $t + 1$ in terms of consumption at date t is such that the amount of borrowing and lending is zero. However, there is a large amount of costly intermediated borrowing and lending between households, and as a consequence, borrowing rates exceed lending rates. When borrowing and lending rates differ, a question arises: Should the equity premium be measured relative to the riskless borrowing rate or the riskless lending rate?

[2]McGrattan and Prescott (2003) use long-term high-grade municipal bonds as a proxy for the riskless security.

To address this question, Mehra, Piguillem, and Prescott (2011) constructed a model that incorporates agent heterogeneity and costly financial intermediation. The resources used in intermediation (3.4 percent of GNP) and the amount intermediated (1.7 percent of GNP) imply that the average household borrowing rate is at least 2 percentage points higher than the average household lending rate. Relative to the level of the observed average rates of return on debt and equity securities, this spread is far from being insignificant and cannot be ignored when addressing the equity premium.

In this model,[3] a subset of households both borrow money and hold equity. Consequently, a no-arbitrage condition is that the return on equity and the borrowing rate are equal (5 percent). The return on government debt, the household lending rate, is 3 percent. If I use the conventional definition of the equity premium—the return on a broad equity index less the return on government debt—I would erroneously conclude that in this model, the equity premium is 2 percent. The difference between the government borrowing rate and the return on equity is not an equity premium; it arises because of the wedge between borrowing and lending rates. Analogously, if borrowing and lending rates for equity investors differ, and they do in the U.S. economy, the equity premium should be measured relative to the investor borrowing rate rather than the investor lending rate (the government's borrowing rate). Measuring the premium relative to the government's borrowing rate artificially increases the premium for bearing aggregate risk by the difference between the investor's borrowing and lending rates.[4] If such a correction is made to the benchmark discussed earlier, the "equity premium" is further reduced by 2 percentage points. Thus, I have accounted for 4 percentage points of the equity premium reported in Mehra and Prescott (1985) by factors other than aggregate risk.

Abstracting from Life-Cycle Effects and Borrowing Constraints on the Young

In Constantinides, Donaldson, and Mehra (2002), we examined the impact of life-cycle effects, such as variable labor income and borrowing constraints, on the equity premium. We illustrated these ideas in an overlapping-generations exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period.

[3]There is no aggregate uncertainty in our model.

[4]For a detailed exposition of this and related issues, see Mehra and Prescott (2008).

In the article, we explored the implications of a borrowing constraint by deriving and contrasting the stationary equilibriums in two versions of the economy. In the *borrowing-constrained* version, the young are prohibited from borrowing and from selling equity short. The *borrowing-unconstrained* economy differs from the borrowing-constrained one only in that the borrowing constraint and the short-sale constraint are absent.

The attractiveness of equity as an asset depends on the correlation between consumption and equity income. Because the marginal utility of consumption varies inversely with consumption, equity will command a higher price (and consequently, a lower rate of return) if it pays off in states when consumption is high and vice versa.[5]

A key insight of ours in the article is that as the correlation of equity income with consumption changes over the life cycle of an individual, so does the attractiveness of equity as an asset. Consumption can be decomposed into the sum of wages and equity income. Young people looking forward at the start of their lives have uncertain future wage and equity income; furthermore, the correlation of equity income with consumption will not be particularly high as long as stock and wage income are not highly correlated. This is empirically the case, as documented by Davis and Willen (2000). Equity will, therefore, be a hedge against fluctuations in wages and a "desirable" asset to hold as far as the young are concerned.

The same asset (equity) has a very different characteristic for the middle-aged. Their wage uncertainty has largely been resolved. Their future retirement wage income is either zero or deterministic, and the innovations (fluctuations) in their consumption occur from fluctuations in equity income. At this stage of the life cycle, equity income is highly correlated with consumption. Consumption is high when equity income is high, and equity is no longer a hedge against fluctuations in consumption; hence, for this group, equity requires a higher rate of return.

The characteristics of equity as an asset, therefore, change depending on the predominant holder of the equity. Life-cycle considerations thus become crucial for asset pricing. If equity is a desirable asset for the marginal investor in the economy, then the observed equity premium will be low relative to an economy where the marginal investor finds it unattractive to hold equity. The *deus ex machina* is the stage in the life cycle of the marginal investor.

[5]This is precisely the reason why high-beta stocks in the simple capital asset pricing model framework have a high rate of return. In that model, the return on the market is a proxy for consumption. High-beta stocks pay off when the market return is high—that is, when marginal utility is low and, hence, their price is (relatively) low and their rate of return high.

We argued that the young, who should be holding equity in an economy without frictions, are effectively shut out of this market because of borrowing constraints. The young are characterized by low wages; ideally, they would like to smooth lifetime consumption by borrowing against future wage income (consuming a part of the loan and investing the rest in higher return equity). However, they are prevented from doing so because human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection.

Therefore, in the presence of borrowing constraints, equity is exclusively priced by middle-aged investors because the young are effectively excluded from the equity markets and a high equity premium is thus observed. If the borrowing constraint is relaxed, the young will borrow to purchase equity, thereby raising the bond yield. The increase in the bond yield induces the middle-aged to shift their portfolio holdings from equities to bonds. The increase in demand for equity by the young and the decrease in demand for equity by the middle-aged work in opposite directions. On balance, the effect is to increase both the equity and the bond return, while shrinking the equity premium.

The results suggest that, depending on the parameterization, between 2 and 4 percentage points of the observed equity premium can be accounted for by incorporating life-cycle effects and borrowing constraints.

Conclusion

I have argued that using an appropriate benchmark for the risk-free rate, accounting for the difference between borrowing and lending rates, and incorporating life-cycle features can account for the equity premium. That this can be accomplished without resorting to risk supports the conclusion of Mehra and Prescott (1985) that the premium for bearing systematic risk is small.

My projection for the equity premium is that at the end of the next decade, it will be higher than that observed in the past. During the next 10 years, the ratio of the retired population to the working-age population will increase. These retired households, in an attempt to hedge against outliving their assets, will likely rebalance their portfolios by substituting annuity-like products for equity. Because, in equilibrium, all assets must be held, this substitution will lead to an increase in the expected equity premium. Consequently, during this adjustment process, the realized equity premium will probably be lower than the historical average.

REFERENCES

Constantinides, G.M., J.B. Donaldson, and R. Mehra. 2002. "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle." *Quarterly Journal of Economics*, vol. 117, no. 1 (February):269–296.

Davis, Stephen J., and Paul Willen. 2000. "Using Financial Assets to Hedge Labor Income Risk: Estimating the Benefits." Working paper, University of Chicago.

McGrattan, E.R., and E.C. Prescott. 2003. "Average Debt and Equity Returns: Puzzling?" *American Economic Review*, vol. 93, no. 2 (May):392–397.

Mehra, R., and E.C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, vol. 15, no. 2 (March):145–161.

———. 2008. "Non-Risk-Based Explanations of the Equity Premium." In *Handbook of Investments: The Handbook of the Equity Risk Premium*. Edited by R. Mehra. Amsterdam: Elsevier.

Mehra, R., F. Piguillem, and E.C. Prescott. 2011. "Costly Financial Intermediation in Neoclassical Growth Theory." *Quantitative Economics*, vol. 2, no. 1 (March):1–36.

RESEARCH FOUNDATION CONTRIBUTION FORM

☑ **Yes,** I want the Research Foundation to continue to fund innovative research that advances the investment management profession. Please accept my tax-deductible contribution at the following level:

Contributing Research Fellow . . . .$25,000 to $49,999
Research Fellow . . . .$10,000 to $24,999
Contributing Donor . . . .$1,000 to $9,999
Donor . . . . Up to $999

I would like to donate $________________.

☐ My check is enclosed (payable to the Research Foundation of CFA Institute).
☐ I would like to donate appreciated securities (send me information).
☐ Please charge my donation to my credit card.
☐ VISA ☐ MC ☐ Amex ☐ Diners ☐ Corporate ☐ Personal

Card Number

____/____
Expiration Date

Name on card PLEASE PRINT

☐ Corporate Card
☐ Personal Card

Signature

☐ This is a pledge. Please bill me for my donation of $________________.

☐ I would like recognition of my donation to be:
☐ Individual donation ☐ Corporate donation ☐ Different individual

PLEASE PRINT NAME OR COMPANY NAME AS YOU WOULD LIKE IT TO APPEAR

PLEASE PRINT ☐ Mr. ☐ Mrs. ☐ Ms. MEMBER NUMBER________________

Last Name (Family Name) First Middle Initial

Title

Address

City State/Province Country ZIP/Postal Code

11ERP

Please mail this completed form with your contribution to:
The Research Foundation of CFA Institute • P.O. Box 2082
Charlottesville, VA 22903-0638 USA

For more on the Research Foundation of CFA Institute, please visit www.cfainstitute.org/about/foundation/.

The Research Foundation of CFA Institute Board of Trustees 2011–2012

Chair
Thomas M. Richards, CFA
Nuveen HydePark Group, LLC

Jeffery V. Bailey, CFA
Target Corporation

Renee Kathleen-Doyle Blasky, CFA
Vista Capital Ltd.

Dwight Churchill, CFA
Bedford, NH

Margaret E. Franklin, CFA†
Kinsale Private Wealth Inc.

William Fung
London Business School

James P. Garland, CFA
The Jeffrey Company

John T. "JT" Grier, CFA
Virginia Retirement System

Walter V. "Bud" Haslett, Jr., CFA†
CFA Institute

Alan M. Meder, CFA†
Duff & Phelps Investment Management Co.

Lam Swee Sum, CFA
National University of Singapore

Frank K. Reilly, CFA*
University of Notre Dame

John D. Rogers, CFA†
CFA Institute

Raymond W. So
Hang Seng Management College

Fred H. Speece, Jr., CFA*
Speece Thorson Capital Group Inc.

Wayne H. Wagner, CFA
Venice Beach, CA

Arnold S. Wood
Martingale Asset Management

*Emeritus †Ex officio

Officers and Directors

Executive Director
Walter V. "Bud" Haslett, Jr., CFA
CFA Institute

Research Director
Laurence B. Siegel
Ounavarra Capital LLC

Secretary
Tina Sapsara
CFA Institute

Treasurer
Kim Maynard
CFA Institute

Research Foundation Review Board

William J. Bernstein
Efficient Frontier Advisors

Stephen J. Brown
New York University

Sanjiv Das
Santa Clara University

Bernard Dumas
INSEAD

Stephen Figlewski
New York University

Gary L. Gastineau
ETF Consultants, LLC

William N. Goetzmann
Yale School of Management

Stephen A. Gorman, CFA
Wellington Management Company

Elizabeth R. Hilpman
Barlow Partners, Inc.

Paul D. Kaplan
Morningstar, Inc.

Robert E. Kiernan III
Advanced Portfolio Management

Robert W. Kopprasch, CFA
The Yield Book Inc.

Andrew W. Lo
Massachusetts Institute of Technology

Alan Marcus
Boston College

Paul O'Connell
FDO Partners

Krishna Ramaswamy
University of Pennsylvania

Andrew Rudd
Advisor Software, Inc.

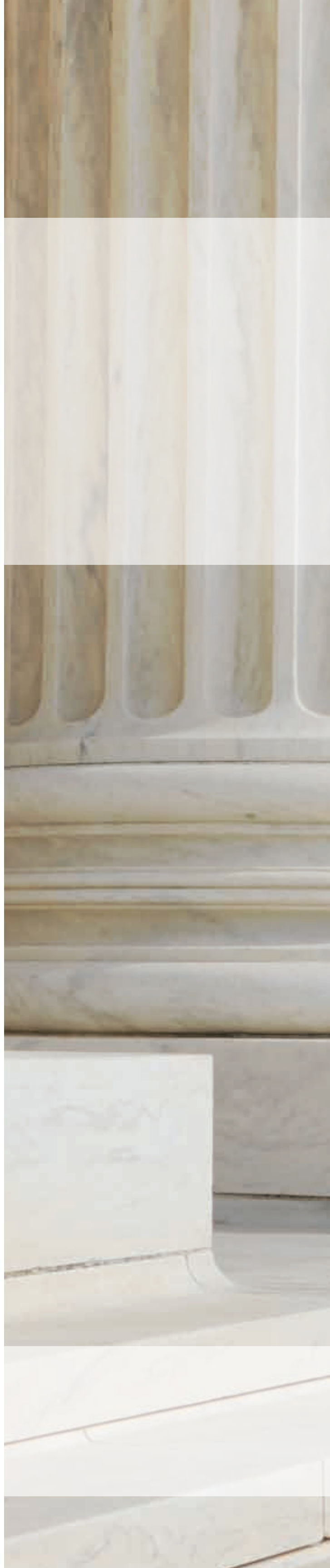
Lee R. Thomas
Pacific Investment Management Company

Robert Trevor
Macquarie University

RESEARCH FOUNDATION OF CFA INSTITUTE

available online at
www.cfapubs.org

9 781934 667446 90000



"An Opening of Minds"

"I think investors are starting to come around to the view that stocks aren't quite as special as they once thought," says Rob Arnott

By Jonathan Barnes

"My career has largely been successful as a consequence of the fact that I love to test ideas," says Rob Arnott, chairman and CEO of Research Affiliates and former editor in chief of the *Financial Analysts Journal*. Arnott's reputation for testing conventional investment wisdom made him one of the key contributors when the Research Foundation of CFA Institute gathered leading academics and practitioners in 2011 to discuss the equity risk premium (ERP), the expected return for equities in excess of a risk-free rate. He delivered a presentation titled "Equity Risk Premium Myths," which was subsequently included in the book *Rethinking the Equity Risk Premium*. In this interview with *CFA Institute Magazine*, Arnott corrects some of the misconceptions about the ERP, argues that "a cult of equities is worshipping a false idol," deconstructs the notion of a risk-free rate, and explains why "our industry, both on the practitioner and on the academic sides, has tremendous inertia, a resistance to new ideas."

Do we need a stronger definition of the equity risk premium?
All too often, the term "equity risk premium" is attached to widely different concepts. It is applied to the historical difference in returns between stocks and bonds—or between stocks and cash—and it is also applied to forward-looking expectational return differences. Really, a risk premium is an *expectational* return, so when we look at historical returns, I think it is important to use different terminology. I prefer the term "historical excess return," not risk premium.

If we turn attention from past to future, the equity risk premium should be the expected incremental return that an investor will likely earn from a willingness to hold stocks instead of bonds or cash. So, one needs to further define one's terms. The risk premium versus bonds and the risk premium versus cash are very different. Today, cash yields nothing; 30-year bonds have yields around 3%.

Which measure is more widely used?
Academia tends to think of the equity risk premium relative to a risk-free rate (never mind that there is nothing that is really risk free in life), and typically that is thought of as a cash yield. A much more relevant measure is equities versus long bonds because they both have a long investment horizon. Cash is very risky for the long-term investor!

ALL TOO OFTEN, THE TERM "EQUITY RISK PREMIUM" IS ATTACHED TO WIDELY DIFFERENT CONCEPTS.

When we look at stocks relative to long bonds, we can do some very simple arithmetic as it relates to expectational returns. Thirty-year bonds have yields around 3%, and the real return as indicated by long-term Treasury Inflation-Protected Securities (TIPS) is 0.5%, give or take.

Stocks produce returns in a real return form because earnings and dividends grow with inflation, plus a real growth kicker. Historically, going back a hundred years, you find earnings and dividends have grown a little less than 1.5% above the rate of inflation. If you add that to the current yield, you get something on the order of a 3.5% expected real return, as against 0.5% for long TIPS. That gives you a 3% risk premium. And that assumes that past rates of growth can continue, given the headwinds from our aging population, as well as our burgeoning debt and deficits.

So when we reframe the definition in terms of forward-looking return expectations for stocks (relative to forward-looking real return expectations for long bonds), we get a comparison of two relatively similar-horizon investments and a comparison that has some real economic meaning. That's my preferred way of thinking about the equity risk premium.

Is more standardization of the ERP needed?
Discussions about the equity risk premium often occur in vague terms: How much more do you expect to earn from a willingness to bear equity market risk? How much more return relative to what? Over what investment horizon? These questions are left ambiguous in all too many examinations of the equity risk premium. If they are defined with any precision, you get much more reasonable apples-with-apples comparisons. Then, you have an ability to examine the underlying assumptions.

There is an annual academic survey of estimates on the equity risk premium in which the ERP is defined as a long-term return against T-bills. But you still have to factor inflation expectations, and on a long-term basis, inflation is anyone's guess, not to mention the future real T-bill yields. So, even with studies that define their terms, if you have a gap in return horizon—cash has a horizon that is measured in weeks or months, stocks have a horizon that is measured in decades—then again, you get into ambiguous comparisons of apples and oranges and a relatively meaningless phenomenon.

Can you explain the myth that the equity risk premium is 5%?
The notion that stocks beat bonds by 5% was embraced in the 1990s by much of the consulting community (and through the consulting community, by much of the plan sponsor community). It is something of a core belief in the practitioner community. This myth is very dangerous because the long-term historical excess return—while not far from 5%—is driven in large measure by a change in valuation multiples for equities. The long-term historical average dividend yield for stocks going back a hundred or more years is about 4%. If the yield now is 2%—a rise in valuation multiples from 25 years of dividends to 50 years of dividends—that is a big change in valuation multiples. So, it creates an inflated historical excess return, which people then translate into an inflated expectational risk premium.

How does your estimate of 3% compare historically?
It's above the historic norms. In 2002, I wrote a paper with Peter Bernstein for the *Financial Analysts Journal* that showed that the reasonable historical equity risk premium—not the excess

return—but what would reasonably have been expected historically for stocks relative to long bonds—was 2.4%.

So, if we are looking at 3% today, that means that right now we have a modestly outsized equity risk premium (if future economic growth matches past growth). It's predicated on negative real yields at the long end of the bond market, so that is a big problem. If you are looking at anemic real returns on bonds (and less-anemic real returns on stocks), you get a positive risk premium through the unfortunate path of generally dismal returns.

I THINK THE MYTHS ARE A CONSEQUENCE OF INERTIA. OUR INDUSTRY, BOTH ON THE PRACTITIONER AND ON THE ACADEMIC SIDES, HAS TREMENDOUS INERTIA, A RESISTANCE TO NEW IDEAS.

Another myth is that the ERP is static over time, companies, and markets. Can you say more?

There are respected academics who build their theories on the notion that the equity risk premium must be static. Yet, on the other hand, there are those who argue that the equity risk premium varies from one stock to another. If it varies from one stock to another, why shouldn't it vary from one month or year to another? The notion of a static equity risk premium is another unfortunate myth.

The risk premium is really a function of pricing. When bond yields are high, the risk premium can get very skinny indeed. Ever so briefly in 2000, you could buy TIPS, *long-term* TIPS, extending out 20–30 years that had a yield of over 4%. I believe the top was 4.3%. A 4.3% real return guaranteed with full faith and credit of the U.S. Treasury is a marvelous default risk–free return. To have that available in bonds at a time when stocks had a yield of 1% is really quite breathtaking. So, what we find is that the risk premium is dynamic. It changes over time.

And across companies and markets.

Yes, let's look across companies. Bank of America is a huge company and comprises less than 1% of the U.S. stock market. Apple is a much smaller company that comprises over 4% of the U.S. stock market. Is it reasonable to assume that Apple—with wonderful growth, no serious competition, and viewed widely as a safe haven—should have the same risk premium as Bank of America, a company that has in recent years seemed to lose its way strategically and is facing daunting headwinds in the years ahead? Should they be priced at the same forward-looking rate of return? Probably not.

By the same token, compare the risk premium when people were worried about financial Armageddon in early 2009 and the risk premium when people felt that things were getting solidly back on track in early 2011. Should that risk premium be the same from one year to the next? Of course not.

So, yes, risk premia vary cross-sectionally, across time, across markets, across companies. Is the Greek risk premium higher than the U.S. risk premium today? Yeah, I would think so, which means that investors in Greek stocks should be expecting a higher return than investors in U.S. stocks because of the higher expected uncertainty.

Why are these myths so enduring?

I think the myths are a consequence of inertia. Our industry, both on the practitioner and on the academic sides, has tremendous inertia, a resistance to new ideas. Once people are taught a particular way of thinking, there is a resistance to questioning that way of thinking. One could characterize it even as a bit of intellectual laziness. People embrace an idea that they have been taught, and they hang on to that idea. They are reluctant to relinquish it in favor of something else.

People are taught the normal risk premium is 5%. In early 2001, Ron Ryan and I wrote a paper titled "The Death of the Risk Premium," which was first published as a First Quadrant "President's Letter" and later published in the *Journal of Portfolio Management*, where we suggested that the equity risk premium was now negative. That created a firestorm of controversy and even outrage in some quarters—to suggest that stocks would produce a lower return than bonds. But if stocks have a dividend yield of 1% and bonds have a yield of 6% in an environment of 2% inflation, that points to a negative risk premium, unless stocks can deliver long-term earnings and dividend growth north of 5%. There is nothing written into contract law in the finance community that says, "Stocks must have a positive risk premium."

WE DO OURSELVES A GREAT FAVOR IF WE ABANDON THE NOTION OF A RISK-FREE RATE AND REPLACE IT WITH A NOTION OF A RISK-MINIMIZING ASSET OR PORTFOLIO OVER A HORIZON MATCHING THE INTENDED LIABILITIES.

Why are you so interested in these myths?

My career has largely been successful as a consequence of the fact that I love to test ideas. The more widely accepted an idea is, the more I am inclined to say, "Let's test it and see if it is true."

One of the things that startled me over the course of my career is how few people pursue that line of reasoning—"If an idea is well accepted, maybe we should test it"—and how many people resist those tests when they turn out to suggest that conventional wisdom is wrong. Conventional wisdom isn't always wrong; it's just not always *right*.

How risk free is the risk-free rate?

I think the whole notion of a risk-free rate is a distraction which takes our eye off of the ball in terms of how people think about investments. First, risk free in what context?

The risk of a 30-day Treasury bill defaulting is, for all intents and purposes, zero. The risk of it producing a real return that is less than we expect—that is a much bigger risk because the uncertainty about next month's CPI has a certain standard deviation that makes that so-called risk-free asset a little less risk free than we might think or hope.

Try to persuade any investor with a long-term liability—a typical pension fund, for instance—that owning and rolling T-bills is a risk-free way to fund those pensions. Come on! We don't know what the rates are going to be over the coming years. We don't know what the inflation is going to be, and we don't know what the growth of the liability itself will be. There is *no such thing* as a risk-free rate. The sooner we abandon the notion that there is a risk-free rate, the better off we will be.

If not risk free, then what?

For most long-term investors, the risk-*minimizing* asset—not risk *free*—is something that is duration-matched to your intended spending stream and to your liabilities. If you are a pension fund, for instance, if those liabilities have an inflation kicker to them—if they are sensitive to the rates of inflation—then long TIPS are your risk-minimizing asset.

If we think in terms of risk-minimizing assets over a horizon long enough to matter, we arrive at very, very different answers. All of a sudden, what feels low risk (a cash-dominated portfolio) turns out to be very high risk measured in terms of long-term return expectations and long-term liabilities. Something that feels pretty volatile, a 30-year TIPS instrument, winds up being very low risk measured against long-term liabilities. So, I think we do ourselves a great favor if we abandon the notion of a risk-free rate and replace it with a notion of a risk-minimizing asset or portfolio over a horizon matching the intended liabilities.

Would that alter the traditional asset-pricing models that evaluate risk–return trade-offs?

Peter Bernstein and I published a paper way back in 1988 in the *Harvard Business Review* (they assigned the title "The Right Way to Manage Your Pension Fund," which I thought was a pretty arrogant title). The paper simply said, "If you redefine your efficient frontier to characterize risk as the mismatch between your assets and liabilities, you wind up with a very different efficient frontier and a very different portfolio mix." We urged consultants and pension funds to consider optimizing their holdings on the basis of a redefinition of risk. To this day, I believe that makes absolute sense, and to this day, hardly anyone does it.

How does the LIBOR scandal tie in to this?

I think that the LIBOR scandal is simultaneously a big deal and much ado about nothing, which sounds contradictory.

I say much ado about nothing because when people price swaps off LIBOR, when it is a gamed LIBOR, they figure out what they want to charge for the swap and they price it relative to that gamed LIBOR. The gaming of the LIBOR has nothing to do with the rate that they are charging. The rate that they are charging relative to LIBOR is really an outcome of setting a rate that you want to charge and subtracting the gamed LIBOR from it. So if the gaming of LIBOR is much the same from one period to the next, no one is harmed.

But it was a very big deal in the sense that people trusted that it was a fair interbank borrowing rate. We have had so many damaging body blows to the public's sense of trust in the capital markets. How useful are the capital markets if we can't trust them? How effective is the capitalist system that is predicated on trust? When we do a deal, we trust that the other side will honor their side of the deal.

You attended the CFA Institute forums on the equity risk premium in 2001 and 2011. What did you learn? What was your experience at the forums?

They were fun. As I mentioned, when Ron Ryan and I wrote the paper "The Death of the Equity Risk Premium" in 2000, we ran into a buzz saw of resistance. Today, you don't get that pushback. One thing that has changed is that people, probably by dint of the pain of the last dozen years, are beginning to recognize that the cult of equities is itself promulgating huge myths.

The notion that double-digit returns are natural for stocks, the notion that lower yields are the market's way of telling you to expect faster growth, the notion that stocks are assuredly going to produce higher returns than long bonds for those patient enough to stay the course over the course of one or two economic cycles and that stocks are less risky than bonds for the truly long-term investor—these are all myths that are fast dissipating.

My view that a cult of equities is worshipping a false idol is no longer a fringe view that gets one consigned to our industry's virtual lunatic asylum. It's becoming an acceptable view. So I think we are seeing an opening of minds. The opening of minds is unfortunately a dozen years too late to avert damage, but it is important and interesting to see that it is happening.

You've written on the necessity of challenging deeply rooted assumptions of finance theory. Can you explain?

Neoclassical finance and the capital asset pricing model are predicated on an array of powerful theories and, in many cases, mathematical proofs that demonstrate that if the market behaves in thus and such a fashion, it will have thus and such implications.

Take the capital asset pricing model. If markets are efficient and if investors share a common view on forward-looking risks and returns, if investors trade for free with no taxes and no trading costs, and if all investors have a similar utility function, then the market-clearing portfolio will be the "mean–variance-efficient portfolio" and you can't beat it on a risk-adjusted basis.

That is a very powerful conclusion—deservedly winning a Nobel Prize for Bill Sharpe—built on a foundation of heroic and clearly inaccurate assumptions. I think finance theory is *wonderful*, but I think it is important that we acknowledge that finance theory is theory. It is not the real world. Theory is designed to tell us how the world *ought* to work. The more we can learn from theory and conform theory to better match the real world, the deeper our understanding of markets.

I think, with the coming quarter century, it will be marvelous if we see a marriage—and it will be an uncomfortable marriage—of neoclassical finance with behavioral finance, a theoretical foundation for the empirical observations of behavioral finance. The big issues in finance theory are really simple. If you assume that the theory is correct and true, then we are tacitly assuming that the assumptions are correct and true. And yet nobody would argue that the assumptions are true. I think we need to back off from the notion that theory is reality.

FINANCE THEORY IS THEORY. IT IS NOT THE REAL WORLD.

Are equities worth the risk, given the potentially low equity risk premium?

I think investors are starting to come around to the view that stocks aren't quite as special as they once thought. The sad irony is that the more extravagantly expensive stocks are, the more members you will have in the cult of equities. The reason for that is simple. Stocks become extravagantly expensive by performing brilliantly. After they have performed brilliantly, it is painful to argue the case that stocks are a lousy investment. People come around to the view that stocks aren't guaranteed a premium return *after* equities have underperformed badly for a long period of time. That is unfortunate and it is ironic, but it is a simple fact.

Jonathan Barnes is a financial journalist and author of the novel *Reunion*.

WILEY-BLACKWELL

American Finance Association

The Level and Persistence of Growth Rates
Author(s): Louis K. C. Chan, Jason Karceski, Josef Lakonishok
Source: *The Journal of Finance*, Vol. 58, No. 2 (Apr., 2003), pp. 643-684
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/3094553
Accessed: 04/02/2010 09:57

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=black.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Blackwell Publishing and *American Finance Association* are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Finance.*

http://www.jstor.org

THE JOURNAL OF FINANCE • VOL. LVIII, NO. 2 • APRIL 2003

The Level and Persistence of Growth Rates

LOUIS K. C. CHAN, JASON KARCESKI, and JOSEF LAKONISHOK*

ABSTRACT

Expectations about long-term earnings growth are crucial to valuation models and cost of capital estimates. We analyze historical long-term growth rates across a broad cross section of stocks using several indicators of operating performance. We test for persistence and predictability in growth. While some firms have grown at high rates historically, they are relatively rare instances. There is no persistence in long-term earnings growth beyond chance, and there is low predictability even with a wide variety of predictor variables. Specifically, IBES growth forecasts are overly optimistic and add little predictive power. Valuation ratios also have limited ability to predict future growth.

THE EXPECTED RATE of growth in future cash flows (usually proxied by accounting earnings) plays a pivotal role in financial management and investment analysis. In the context of aggregate market valuation, for example, projections about future growth are instrumental in predicting the equity risk premium. Much current controversy surrounds the appropriate level of the equity risk premium, as well as whether recent market valuation levels (at least as of year-end 1999) can be justified (Asness (2000), Welch (2000), Fama and French (2002)). Debate also revolves around how much of the performance of equity asset classes, such as large glamour stocks, can be attributed to changes in profitability growth (Fama and French (1995), Chan, Karceski, and Lakonishok (2000)). When applied to the valuation of individual stocks, projected growth rates have implications for the cross-sectional distribution of cost of capital estimates (Fama and French (1997), Claus and Thomas (2001), Gebhardt, Lee, and Swaminathan (2001)), as well as widely followed valuation ratios like price-to-earnings and price-to-book ratios.

Common measures of expected growth in future earnings, such as valuation ratios and analysts' growth forecasts, vary greatly across stocks. In the case of price-to-earnings multiples for the IBES universe of U.S. firms, for example, at

*Chan is with the Department of Finance, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign; Karceski is with the Department of Finance, Warrington College of Business Administration, University of Florida; and Lakonishok is with the Department of Finance, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign, and NBER. We thank the editor, Rick Green; Cliff Asness; Kent Daniel; Ken French; an anonymous referee; and seminar participants at Dartmouth, Duke University, the London School of Economics Financial Markets Group, the NBER Behavioral Finance Fall 2000 workshop, the University of Illinois, Washington University, and the Western Finance Association 2001 meetings.

year-end 1999, the distribution of the stock price relative to the consensus forecast of the following year's earnings has a 90th percentile of 53.9, while the 10th percentile is 7.4, yielding a difference of 46.5. Firms with a record of sustained, strong past growth in earnings are heavily represented among those trading at high multiples. Security analysts issue positive recommendations for these stocks and forecast buoyant future prospects. Other stocks with a history of disappointing past growth are shunned by the investment community. They are priced at low multiples and analysts are unexcited about their outlook. Putting aside the possibility of mispricing, one reason for the disparity in multiples is differences in risk. At the level of individual stocks, however, the relation between risk and expected return is weak (Fama and French (1992)). It is thus unlikely that the large dispersion is driven primarily by risk (the evidence in Beaver and Morse (1978) also supports this view). Rather, if the pricing is rational, most of the cross-sectional variation reflects differences in expected growth rates. A more direct measure of the market's expectations, security analysts' forecasts of long-term growth in earnings, also displays large differences across stocks. For example, the 90th percentile of the distribution of IBES five-year forecasts is 40 percent as of year-end 1999, compared to the 10th percentile of 8.9 percent. If analysts and investors do not believe that future earnings growth is forecastable, they would predict the same growth rate (the unconditional mean of the distribution) for all companies, and it is unlikely that the dispersion in forecasts or price–earnings ratios would be as large as it actually is.

Based on market valuations and analysts' forecasts, then, there is a widespread belief among market participants that future earnings growth is highly predictable. However, economic intuition suggests that there should not be much consistency in a firm's profitability growth. Following superior growth in profits, competitive pressures should ultimately tend to dilute future growth. Exit from an unprofitable line of business should tend to raise the remaining firms' future growth rates. Some support for this logic comes from Fama and French (2002). Their evidence for the aggregate market suggests that while there is some short-term forecastability, earnings growth is in general unpredictable.

In short, there may be a sharp discrepancy between share valuations along with analysts' predictions on the one hand, and realized operating performance growth on the other. The discrepancy may reflect investors' judgmental biases or agency distortions in analysts' behavior. In any event, the divergence is potentially large, judging from current market conditions. For instance, take a firm with a ratio of price to forecasted earnings of 100. Such cases are by no means minor irregularities: based on values at year-end 1999, they represent about 11.9 percent of total market capitalization. To infer the growth expectations implicit in such a price earnings ratio, we adopt a number of conservative assumptions. In particular, suppose the multiple reverts to a more representative value of 20 in 10 years, during which time investors are content to accept a rate of return on the stock of zero (assume there are no dividends). A multiple of 20 is conservative, since Siegel (1999) argues that a ratio of 14 may not be an unreasonable long-term value. Further, an adjustment period of 10 years is not short, in light of the fact that many of the largest firms at year-end 1999 did not exist 10 years ago. These assumptions imply that earnings must grow by a factor of five, or at a rate of about 17.5 percent per year, for the next 10 years. Alternatively, suppose investors put up with a paltry 10 percent rate of return (Welch (2000), reports that financial economists' consensus expected return is considerably higher). Then earnings must grow at an even more stellar rate (29.2 percent per year) over 10 years to justify the current multiple.

The above example highlights the two questions we tackle in this paper. How plausible are investors' and analysts' expectations that many stocks will be able to sustain high growth rates over prolonged periods? Are firms that can consistently achieve such high growth rates identifiable ex ante? We begin by documenting the distribution of growth rates realized over horizons of 1, 5, and 10 years. This evidence lets us evaluate the likelihood of living up to the expectations of growth that are implicit in market valuation ratios. To justify rich valuations, investors must believe that high growth persists over many years. Accordingly, we also examine whether there is persistence in operating performance growth. Individual firms' earnings and incomes can be very erratic, so a robust empirical design is a crucial consideration. We employ nonparametric tests on multiple indicators of operating performance across a large cross section of stocks over relatively long horizons. In addition, we focus our tests for persistence by examining subsets of firms where future growth is more likely to be predictable (e.g., stocks in the technology sector and stocks which have displayed persistence in past growth). To give the benefit of the doubt to the possibility of persistence, we relax the definition of consistency in growth and redo our tests. Finally, we expand the list of variables to forecast growth beyond past growth rates. We examine whether valuation measures, such as earnings yields and ratios of book-to-market equity and sales-to-price, are associated with growth on an ex ante as well as ex post basis. Security analysts' earnings forecasts are also widely used as measures of the market's expectations of growth in future earnings. As a check on the quality of analysts' predictions, we evaluate how well realized growth rates align with IBES consensus forecasts.

Our main findings are as follows. Our median estimate of the growth rate of operating performance corresponds closely to the growth rate of gross domestic product over the sample period. Although there are instances where firms achieve spectacular growth, they are fairly rare. For instance, only about 10 percent of firms grow at a rate in excess of 18 percent per year over 10 years. Sales growth shows some persistence, but there is essentially no persistence or predictability in growth of earnings across all firms. Even in cases that are popularly associated with phenomenal growth (pharmaceutical and technology stocks, growth stocks, and firms that have experienced persistently high past growth), signs of persistent growth in earnings are slim. Security analysts' long-term growth estimates tend to be overoptimistic and contribute very little to predicting realized growth over longer horizons. Market valuation ratios have little ability to discriminate between firms with high or low future earnings growth. An expanded set of forecasting variables also has scant success in predicting future earnings growth. All in all, our evidence on the limited predictability of earnings growth suggests that investors should be wary of stocks that trade at very high multiples. Very few firms are able to live up to the high hopes for consistent growth that are built into such rich valuations.

Related prior research in the financial literature on the behavior of earnings growth is meager. Little (1962) and Little and Rayner (1966) examine the growth in earnings of a limited sample of U.K. firms in the 1950s. Early evidence for U.S. firms is provided by Lintner and Glauber (1967) and Brealey (1983). Beaver (1970) and Ball and Watts (1972) start a long line of papers that apply time-series models to earnings. However, few firms have sufficiently long earnings histories to allow precise estimation of model parameters, and the emphasis in this line of work has been on short-term forecasting. More recently, Fama and French (2002) examine the time-series predictability of aggregate earnings for the market. Our work is closest in spirit to that of Fama and French (2000), who look at the cross-sectional predictability of firms' earnings, but even they focus on one-year horizons.

A much larger number of studies by academics and practitioners rely on estimates of expected long-term earnings growth for stock valuation, or for estimating firms' cost of capital. A selective list includes Bakshi and Chen (1998), Lee, Myers, and Swaminathan (1999), Claus and Thomas (2001), and Gebhardt et al. (2001). In particular, many studies use long-term consensus IBES forecasts for expected growth rates (see, e.g., Mezrich et al. (2001)). Given the widespread use of IBES long-term estimates, it is important to evaluate their correspondence with realized growth rates.

The rest of the paper is organized as follows. Section I discusses our sample and some basics of the methodology. The cross-sectional distribution of firms' growth rates is reported in Section II. Section III presents the results of runs tests for consistency in growth of operating performance. Section IV takes up the issue of survivorship bias. Although our main focus is not on the determinants of valuation multiples, Section V examines the relation between growth and valuation ratios such as earnings yields and book-to-market ratios, on both an ex ante and ex post basis. We compare IBES long-term forecasts with realized growth rates in Section VI. Section VII uses cross-sectional regressions to forecast future growth using variables including past growth, valuation ratios, and IBES estimates. A final section concludes.

I. Sample and Methodology

Our sample of firms comprises all domestic common stocks with data on the Compustat Active and Research files. Firms are selected at the end of each calendar year from 1951 to 1997. The earlier years are included for the sake of completeness, even though there is a backfill bias in the earlier part of the sample period (see Chan, Jegadeesh, and Lakonishok (1995)), which may impart an upward bias to growth rates in the beginning of the sample. The number of eligible firms grows from 359 in the first sample selection year to about 6,825 in the last year; on average, the sample comprises about 2,900 firms.

We consider three indicators of operating performance: net sales (Compustat annual item number 12), operating income before depreciation (item 13), and income before extraordinary items available for common equity (item 237). While researchers and practitioners tend to focus exclusively on income before extraordinary items, measuring growth in this variable is beset with pitfalls. In many cases, earnings before extraordinary items is negative, so prospective growth rates are undefined (for our sample, in an average year, 29 percent of firms have negative values for earnings before extraordinary items). In other cases, firms grow from low positive values of base-year net income, introducing large outliers.[1] These include such disparate cases as beaten-down companies with depressed earnings and growing startup companies that are beginning to generate profits. To avoid hanging all our inferences on such a noisy variable, therefore, we also consider growth in net sales and growth in operating income before depreciation. These are relatively better-behaved measures of operating performance.

Researchers have adopted different conventions for calculating growth rates. Given our focus on the predictability of growth rates, we measure growth on a per share basis so as to strip out any predictability due to changes in the scale of the firm's operations. This also corresponds to the measurement convention in the investment industry.[2]

Thus, we take the perspective of an investor who buys and holds one share of a stock over some horizon and track the growth in sales or income that accrues to one share, after adjusting for stock splits and dividends. Moreover, two firms can offer the same expected return, but have different earnings growth rates because of their dividend payout policies. From an investor's standpoint, these two stocks would be considered equivalent. To put firms with different dividend policies on an equal footing, therefore, all cash dividends as well as any special distributions (such as when a firm spins off assets) are reinvested in the stock.

II. The Distribution of Growth Rates of Operating Performance

This section documents the distribution of historical growth rates over relatively long horizons (5 and 10 years). For the sake of completeness, results are also provided for 1-year horizons. At each calendar year-end over the sample period, we measure rates of growth in future operating performance for all eligible

[1] Some of these complications may be alleviated by averaging earnings over a number of years and measuring growth in these averages. Since our focus is on point-in-time growth rates, we do not explore this alternative procedure. In unreported work, we also experiment with other ways to calculate growth rates. These include value-weighted growth rates for portfolios, estimated growth rates from least-squares fits of linear and quadratic time trends through sales and income, and growth rates without dividend reinvestment. Generally speaking, the results are robust to how we measure growth rates.

[2] Lakonishok, Shleifer, and Vishny (1994) calculate growth in a firm's overall sales and earnings, while Daniel and Titman (2001) calculate growth on a per share basis. These studies focus on the impact of investor sentiment on stock returns. The hypothesis is that investors tend to favor companies with strong past performance, those in a glamorous line of business, or those which are perceived to be well managed. From this standpoint, it might be argued that it is the performance of the overall company that is relevant, and not just the profits earned per share.

stocks. Percentiles are calculated for the distribution obtained at each year-end. Table I reports the percentiles averaged across years in the sample period, as well as the most recent distribution corresponding to the last selection year of the sample period.

Several points are important as background to the results in Table I. First, since we include reinvestment of dividends and special distributions, the growth rates we report are typically higher than conventionally measured growth rates. The median dividend yield for our sample (averaged across all years) is about 2.5 percent. A second caveat is that the tabulated growth rates are based only on firms who survive for the following 1, 5, or 10 years. The survivorship bias may induce an upward bias in our reported growth rates. Moreover, we follow the conventional approach and do not calculate growth rates for operating income before depreciation or income before extraordinary items when the base-year value is negative.[3] To illustrate the potential magnitude of these complications, on average there are about 2,900 firms available for inclusion in the sample at each year-end. Of these, 2,782 firms survive at the end of the next year and have a reported value for income before extraordinary items. The calculations for 1-year growth in earnings before extraordinary items are based on 1994 of these firms; the remaining 788 firms have negative values for income in the base year. At the 5-year horizon, there are on average 1884 surviving firms. Growth rates are calculated for 1,398 of these; 486 have negative base-year values. At the 10-year horizon, there are 1,265 surviving firms: 1,002 and 263 with positive and negative base-year values, respectively. In a subsequent section, we examine the performance of nonsurviving firms.

Since negative base-year values are quite common for income before extraordinary items, valid growth rates are unavailable in many cases. These observations are symptomatic of another problem. In particular, the high frequency of cases with negative base values suggests that the neighboring portion of the distribution (with low, positive base-year values) contains a large fraction of the observations as well. These instances give rise to some very high growth rates. For growth over five years, for example, the 98th percentile value for growth in income before extraordinary items averages 62.4 percent per year. Hence, while growth in income before extraordinary items captures much of the investment community's interest, its behavior is the most questionable. While the same problem applies to operating income before depreciation, the frequency of negative base-year values is comparatively lower and growth in this variable is less problematic.[4] For growth in this variable, the 98th percentile is 51.2 percent on average. In comparison, sales growth is relatively well behaved, with a 98th percentile value of 40.5 percent on average. These comparisons suggest that looking at

[3] Note, however, that even if we are unable to calculate growth in income before extraordinary items in such a case, we still get a reading on a firm's operating performance growth from sales (or operating income before depreciation if it is positive).

[4] For example, of the firms surviving after one year and with a reported value for income before depreciation, about 14 percent on average have negative base-year values. The corresponding percentage for income before extraordinary items is 29 percent.

other indicators beyond income before extraordinary items helps to give a more robust picture of growth in operating performance.

The results in Table I serve as cautionary flags to analysts and investors who pursue stocks with rich price–earnings multiples. Take our original example of a stock with a current price–earnings multiple of 100, which declines to 20 in 10 years' time with an expected return of 10 percent per year. Earnings must grow at 29.2 percent per year over 10 years to justify the current multiple. This is a tall order by historical standards. In particular, the required growth rate corresponds to about the 95th percentile of the distribution of 10-year growth rates, even putting aside the inclusion of dividends. Put differently, suppose earnings grow at a historically more representative, but still healthy, annual rate of 14.7 percent (the 75th percentile of the distribution from Part I). Then the current ratio of 100 would be justified if the time it takes for the multiple to fall to 20 is stretched out to 38 years.

Small firms start from a smaller scale of operations and so have more room for potential growth, possibly justifying a high current multiple. However, high multiples also apply to many large, well-known firms. To see whether large firms in general can also achieve high growth, Table II reports the distribution of growth rates for large firms (companies ranked in the top two deciles of year-end equity market capitalization, based on NYSE breakpoints). Bigger firms have a larger scale of operations and, hence, are more likely to face limits on their growth, so extremely high growth rates are less prevalent in Table II compared to Table I. For example, the 90th percentiles of growth rates over 10 years for income before extraordinary items, operating income before depreciation, and sales are all close to 16 percent per year. Also, note that dividend yields are generally higher for large firms.

Our estimated median growth rate is reasonable when compared to the overall economy's growth rate. On average over the sample period, the median growth rate over 10 years for income before extraordinary items is about 10 percent for all firms. The behavior over the last 10-year period in the sample roughly matches the overall average. Growth in the other two indicators also exhibit comparable medians. After deducting the dividend yield (the median yield is 2.5 percent), as well as inflation (which averages 4 percent per year over the sample period), the growth in real income before extraordinary items is roughly 3.5 percent per year. This is consistent with the historical growth rate in real gross domestic product, which has averaged about 3.4 percent per year over the period 1950 to 1998. It is difficult to see how the profitability of the business sector over the long term can grow much faster than overall gross domestic product.

Looking forward, if we project future growth using the median of the distribution of historical growth rates, the implication is that the expected future return on stocks is not very high. For example, in a simple dividend discount model with constant growth rates and constant payout ratio, the expected return is equal to the dividend yield plus the expected future growth rate of earnings. Given the low level of current dividend yields (below 1.5 percent) and expected inflation of 2.5 percent, the expected return is only about 7.5 pecent. This is lower than the

Table I

Distribution of Growth Rates of Operating Performance over 1, 5 and 10 Years: All Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one, five, and ten years for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample selection date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Percentiles of the distribution are calculated each year-end; the simple average over the entire sample period of the percentiles is reported, along with the distribution of growth rates over horizons ending in the last year of the sample period.

	Percentile								
Sample period	2%	10%	25%	40%	50%	60%	75%	90%	98%
Part I: Annualized Growth Rate over 10 Years									
(A) Sales									
Average	− 9.6	0.1	5.5	8.7	10.2	11.5	13.8	18.0	27.6
Ending 1998	− 16.1	− 3.4	2.9	6.2	7.9	9.5	12.7	19.2	32.9
(B) Operating Income before Depreciation									
Average	− 13.3	− 2.3	4.1	7.6	9.5	11.2	14.1	19.4	31.3
Ending 1998	− 14.6	− 3.3	3.3	7.2	9.0	10.9	14.1	21.5	38.6
(C) Income before Extraordinary Items									
Average	− 15.6	− 3.1	3.9	7.7	9.7	11.6	14.7	20.4	33.4
Ending 1998	− 21.2	− 6.3	2.3	6.9	9.0	11.4	15.3	24.4	48.8
Part II: Annualized Growth Rate over 5 Years									
(A) Sales									
Average	− 18.7	− 4.1	4.3	8.2	10.2	12.0	15.3	22.1	40.5
Ending 1998	− 22.7	− 6.2	2.9	8.0	10.2	12.4	17.1	27.6	56.3
(B) Operating Income before Depreciation									
Average	− 26.8	− 8.4	1.9	7.2	9.8	12.4	17.1	26.7	51.2
Ending 1998	− 24.4	− 7.8	3.5	8.7	11.5	14.4	19.9	33.4	64.4
(C) Income before Extraordinary Items									
Average	− 30.9	− 10.3	1.5	7.4	10.5	13.4	18.8	30.4	62.4
Ending 1998	− 35.1	− 11.5	2.8	9.1	12.4	15.7	23.1	40.1	88.2

Part III: 1-Year Growth Rate									
(A) Sales									
Average	– 47.3	– 12.9	1.2	7.6	10.9	14.2	21.0	38.7	121.7
Ending 1998	– 58.3	– 20.8	– 1.4	6.3	10.3	14.5	24.9	54.1	181.9
(B) Operating Income before Depreciation									
Average	– 69.4	– 30.7	– 5.6	5.9	11.8	17.7	30.6	67.4	253.3
Ending 1998	– 74.1	– 34.7	– 4.9	6.7	12.2	18.5	32.2	76.5	273.2
(C) Income before Extraordinary Items									
Average	– 76.8	– 37.9	– 7.4	6.9	13.3	19.9	35.8	90.2	435.3
Ending 1998	– 87.3	– 48.2	– 13.7	5.4	13.7	21.3	40.4	115.0	727.2

Table II

Distribution of Growth Rates of Operating Performance over 1, 5 and 10 Years: Large Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one, five, and ten years for large firms (in the top two deciles of year-end equity market capitalization, based on NYSE breakpoints). The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Percentiles of the distribution are calculated each year-end; the simple average over the entire sample period of the percentiles is reported, along with the distribution of growth rates over horizons ending in the last year of the sample period.

	Percentile								
Sample period	2%	10%	25%	40%	50%	60%	75%	90%	98%
Part I: Annualized Growth Rate over 10 Years									
(A) Sales									
Average	− 3.4	2.5	6.8	9.4	10.7	11.7	13.3	16.3	22.0
Ending 1998	− 7.7	− 0.2	4.4	6.7	8.5	9.5	11.1	15.0	21.5
(B) Operating Income before Depreciation									
Average	− 8.3	0.6	5.4	8.1	9.5	10.8	12.9	16.1	22.6
Ending 1998	− 11.6	− 1.7	4.3	7.4	8.7	10.4	11.8	16.3	21.4
(C) Income before Extraordinary Items									
Average	− 12.8	− 0.9	4.5	7.5	9.3	10.8	13.1	16.6	23.8
Ending 1998	− 25.6	− 3.8	1.7	6.1	8.2	9.9	13.3	18.5	36.4
Part II: Annualized Growth Rate over 5 Years									
(A) Sales									
Average	− 9.7	− 0.6	6.9	9.4	10.8	11.9	14.1	18.1	27.9
Ending 1998	− 13.6	− 3.0	4.0	8.8	10.2	11.5	13.7	19.6	32.5
(B) Operating Income before Depreciation									
Average	− 16.9	− 3.5	4.3	7.9	9.8	11.5	14.3	19.3	32.1
Ending 1998	− 13.6	− 6.6	4.5	7.5	10.8	12.7	15.6	19.9	32.0
(C) Income before Extraordinary Items									
Average	− 26.4	− 6.4	2.8	7.6	9.8	12.0	15.3	21.3	37.2
Ending 1998	− 39.5	− 10.1	4.3	9.5	11.8	14.4	19.6	30.4	57.4
Part III: 1-Year Growth Rate									
(A) Sales									
Average	− 36.4	− 2.4	5.7	9.3	11.3	13.3	17.0	25.2	47.7
Ending 1998	− 49.8	− 14.7	1.5	6.6	8.9	11.8	18.1	29.1	53.0
(B) Operating Income before Depreciation									
Average	− 52.3	− 15.2	0.2	7.1	10.6	13.8	19.8	33.7	82.3
Ending 1998	− 60.0	− 30.3	− 1.9	6.6	11.1	14.0	20.8	33.4	73.1
(C) Income before Extraordinary Items									
Average	− 67.5	− 25.3	− 2.8	6.9	11.0	14.9	23.1	45.9	216.6
Ending 1998	− 80.0	− 46.9	− 13.5	4.7	11.5	15.5	27.1	56.7	213.6

consensus forecast of professional economists (see Welch (2000)), but is in line with Fama and French (2002).

III. Persistence in Growth

Differences in valuations indicate a pervasive belief that stocks with high or low future growth are easily identifiable ex ante. For example, analysts and investors seem to believe that a firm that has grown rapidly in the past for several years in a row is highly likely to repeat this performance in the future. Conversely, stocks that have done poorly over prolonged periods are shunned and trade at low multiples. This section checks whether there is consistency in growth. We examine whether past growth or other characteristics, such as industry affiliation or firm size, help to predict future growth.

A. Consistency across All Firms

Tables I and II suggest that year-to-year growth in income can take on quite extreme values. As a result, multiyear growth rate levels may look impressive because of one or two isolated years of sharp growth, although growth in other years may be unremarkable. However, many of the firms with lofty multiples grow rapidly every year for several years. Accordingly, we test for consistency in growth using a design that does not rely heavily on the level of growth rates.[5] In our first set of tests, we define consistency as achieving a growth rate above the median for a consecutive number of years: Such cases are labeled as runs.[6]

At each year-end over the sample period, we calculate how many firms achieve runs over horizons of 1 to 10 years in the future. A run over 5 years, for example, denotes a case where in each of the subsequent 5 years, a firm's growth rate exceeds the median growth rate that year. Each year's median is calculated over all growth rate observations available in that year. Again, note that survivorship bias affects our runs tests. To see how many firms achieve runs above the median for 5 years in a row, we necessarily look at firms that survive over the full 5 years. In each of these years, we compare the survivors to a median which is based on all available firms that year, including those that do not survive for the full 5 years,

[5] Brealey (1983) uses a similar procedure.

[6] We want to avoid discarding an entire sequence of observations because one year's growth rate cannot be calculated when earnings are negative. Instead, we handle such cases as follows, taking growth in operating income per share OI_t as an example. In addition to calculating the percentage growth rate of operating income as $(OI_{t+1} - OI_t)/OI_t$ for each firm, we also scale the change in operating income by the stock price as of the base year t, $(OI_{t+1} - OI_t)/P_t$. All firms in a given year are ranked by their values of change in income relative to stock price. For any firm with negative income in a base year, we find its percentile rank based on income change relative to price. We then look up the corresponding percentile value from the distribution of growth rates of income (based on firms with positive base-year values) for that year. This growth rate is then assigned to the firm with negative base-year income. At the same time, however, it would be dangerous to pin our estimates of growth over a 5- or 10-year horizon in Tables I and II on some imputed value of base-year earnings. Accordingly, we do not impute growth rates in those tables for cases with negative base-year values.

Table III

Persistence in Growth Rates of Operating Performance: All Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel A), operating income before depreciation (panel B), or income before extraordinary items available to common equity (panel C). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following ten years, the number of firms with valid growth rates, the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years, the percentage these firms represent relative to the number of valid firms, and the percentage expected under the hypothesis of independence across years, are reported. Statistics are provided for the entire sample period, and for the ten-year horizon corresponding to the last sample formation year.

	Firms with Above-Median Growth each year for Number of Years									
Variable	1	2	3	4	5	6	7	8	9	10
(A) Sales										
Average Number of Valid Firms	2771	2500	2263	2058	1878	1722	1590	1471	1364	1265
Average Number above Median	1386	721	382	209	118	70	42	26	17	11
Percent above Median	50.0	28.8	16.9	10.2	6.3	4.0	2.7	1.8	1.3	0.9
1989–1998	50.0	30.0	18.6	11.9	7.8	5.6	3.4	2.4	1.5	1.2
(B) Operating Income before Depreciation										
Average Number of Valid Firms	2730	2456	2219	2014	1833	1678	1546	1428	1322	1223
Average Number above Median	1365	628	290	136	67	34	18	10	6	4
Percent above Median	50.0	25.6	13.0	6.8	3.6	2.0	1.2	0.7	0.5	0.3
1989–1998	50.0	25.0	13.1	7.0	4.0	2.1	1.3	0.8	0.5	0.5
(C) Income before Extraordinary Items										
Average Number of Valid Firms	2782	2509	2271	2065	1884	1727	1593	1473	1365	1265
Average Number above Median	1391	625	277	125	57	28	14	7	4	2
Percent above Median	50.0	24.9	12.2	6.0	3.0	1.6	0.9	0.5	0.3	0.2
1989–1998	50.0	24.8	12.2	5.7	2.6	1.3	0.8	0.5	0.2	0.0
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

and newly listed firms. Since the survivors are likely to have better performance than the population, they tend to have a greater chance of being above the median. Section IV examines differences between the growth rates of surviving and nonsurviving firms.

Table III reports the counts of runs, averaged across the year-ends. For growth in sales (Panel A), for example, out of an average number of 2,900 firms available for sample selection at each year-end, 2,771 firms on average survive until the end of the following year. Over the following 10 years, there are on average 1,265 surviving firms. Of these, 11 have sales growth rates that exceed the median in each of the 10 years, representing 0.9 percent of the eligible firms. If sales growth is independent over time, we should expect to see 0.5^{10} (about 0.1 percent) of the surviving firms achieve runs above the median over 10 years (see the last row of the table). To give a flavor of what happens in the more recent years, we also report the percentage of firms with runs over the 10-year period ending in the last year of our sample period.

There is a great deal of persistence in sales growth. Over a five-year horizon, for example, on average 118 firms, or 6.3 percent of the 1878 firms who exist over the full five years, turn in runs above the median. The number expected under the hypothesis of independence over time is about 59 (3.1 percent of 1,878), so roughly twice more than expected achieve runs over five years.

The persistence in sales growth may reflect shifts in customer demand, which are likely to be fairly long-lasting. A firm can also sustain momentum in sales by expanding into new markets and opening new stores, by rolling out new or improved products, or by granting increasingly favorable credit terms. Persistence in sales may also arise from managers' "empire-building" efforts, such as expanding market share regardless of profitability. In all these cases, however, profit margins are likely to be shrinking as well, so growth in profits may not show as much persistence as sales growth.

While it may be relatively easy for a firm to generate growth in sales (by selling at a steep discount, for example), it is more difficult to generate growth in profits. The recent experience of Internet companies, where sales grew at the same time losses were accumulating, provides a stark example. Panel B confirms that there is less persistence in operating income before depreciation compared to sales. On average, 67 firms a year, or 3.6 percent of 1,833 surviving firms, have above-median runs for 5 consecutive years. The expected frequency of runs is 3.1 percent or 57 firms. There are, thus, 10 firms more than expected out of 1,833, so the difference is unremarkable. An average of 4 firms a year (or 0.3 percent of 1,223 survivors), which is only 3 more than expected, pull off above-median growth for 10 years in a row. The patterns in the more recent years do not deviate markedly from the averages across the entire sample period.

Any sign of persistence vanishes as we get closer to the bottom line (Panel C). On average, the number of firms who grow faster than the median for several years in a row is not different from what is expected by chance. An average of 57 firms out of 1,884 survivors (3 percent) beat the median for 5 years in a row, while 59 (3.1 percent) are expected to do so. Runs above the median for 10 years occur in 0.2 percent of 1,265 cases (or 2 firms), roughly matching the expected frequency (0.1 percent, or 1 firm). To sum up, analysts and investors seem to believe that many firms' earnings can consistently grow at high rates for quite a few years. The evidence suggests instead that the number of such occurrences is not much different from what might be expected from sheer luck. The lack of consistency in earnings growth agrees with the notion that in competitive markets, abnormal profits tend to be dissipated over time.

Table IV
Persistence in Growth Rates of Operating Performance: Selected Equity Classes

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951 to 1998, and the underlying sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of (the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following ten years, the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years is expressed as a percentage of the number of firms with valid growth rates. Statistics are provided for the following sets of stocks: technology stocks (panel A), comprising stocks whose SIC codes begin with 283, 357, 366, 38, 48, or 737; value stocks (panel B), comprising stocks ranked in the top three deciles by book-to-market value of equity; glamour stocks (panel C), comprising an equivalent number as in panel B of the lowest-ranked stocks by book-to-market value of equity; large stocks (panel D), comprising stocks ranked in the top 2 deciles by equity market value; mid-cap stocks (panel E), comprising stocks ranked in the third through seventh deciles by equity market value; and small stocks (panel F), comprising stocks ranked in the bottom three deciles by equity market value. All decile breakpoints are based on domestic NYSE stocks only.

	Percent of Firms with Above-Median Growth each Year for Number of Years									
Variable	1	2	3	4	5	6	7	8	9	10
				(A) Technology Stocks						
Sales	51.6	30.7	19.1	12.5	8.5	5.9	4.2	3.0	2.3	1.7
Operating Income	51.0	27.2	14.9	8.7	5.3	3.3	2.2	1.4	1.0	0.7
Income before Extraordinary Items	50.9	25.9	13.5	7.3	4.1	2.5	1.5	0.9	0.5	0.4
				(B) Value Stocks						
Sales	50.6	30.0	18.2	11.1	6.9	4.3	2.8	1.9	1.3	0.9
Operating Income	49.3	25.3	13.2	6.8	3.5	1.8	0.9	0.5	0.3	0.2
Income before Extraordinary Items	48.3	23.8	11.4	5.4	2.5	1.2	0.7	0.4	0.3	0.2
				(C) Glamour Stocks						
Sales	48.3	26.6	15.1	8.5	4.7	2.7	1.7	1.0	0.8	0.6
Operating Income	50.1	25.2	11.9	5.9	3.3	1.7	1.0	0.6	0.4	0.3
Income before Extraordinary Items	50.7	25.2	12.0	5.8	2.9	1.6	0.9	0.4	0.2	0.1
				(D) Large Stocks						
Sales	53.2	31.3	18.9	11.7	7.5	4.8	3.2	2.2	1.6	1.1
Operating Income	49.4	25.2	13.0	6.9	3.7	2.0	1.1	0.6	0.4	0.3
Income before Extraordinary Items	46.7	21.9	10.0	4.7	2.2	1.2	0.7	0.4	0.3	0.2
				(E) Mid-cap Stocks						
Sales	53.9	32.4	19.8	12.1	7.6	4.9	3.3	2.2	1.5	1.0
Operating Income	50.5	26.6	13.9	7.5	4.2	2.4	1.5	1.0	0.7	0.4
Income before Extraordinary Items	49.4	24.9	12.4	6.2	3.1	1.6	0.9	0.5	0.3	0.2
				(E) Small Stocks						
Sales	47.0	26.1	14.7	8.6	5.2	3.2	2.1	1.4	1.0	0.7
Operating Income	50.1	25.2	12.6	6.4	3.3	1.8	1.0	0.6	0.4	0.2
Income before Extraordinary Items	51.0	25.5	12.6	6.3	3.2	1.7	0.9	0.4	0.2	0.1
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

B. Consistency for Subsets of Firms

While Table III suggests that there may not be much consistency in growth across all firms, it is possible that consistency may show up more strongly in subsets of firms. Table IV focuses our tests by looking at the performance of subsamples of firms. For a subsample such as small stocks, we consider a "run" as a case where the firm's growth rate exceeds the median for a consecutive number of years, where each year the median is calculated across all firms in the entire sample, not just small stocks. This explains why the percentage of runs is not identically 50 percent in the first year.

Many observers single out technology and pharmaceutical firms as instances of consistently high growth over long horizons. Such firms may be able to maintain high growth rates because of their intangible assets, such as specialized technological innovations or drug patents. Panel A examines firms in these sectors. Specifically, the sample comprises firms that are relatively heavily engaged in research and development activity, and are predominantly drawn from the computer equipment, software, electrical equipment, communications, and pharmaceutical industries.[7] Growth in sales and operating income for the set of technology firms both display strong persistence. However, the percentage of runs in income before extraordinary items does not differ markedly from the expected frequency. For example, over a five-year horizon, 14 firms (or 4.1 percent of the 331 surviving technology stocks) have above-median runs. This is only 4 more than the expected number of runs (10 firms, or 3.1 percent). The recent experience of Internet companies provides numerous examples where sales grow rapidly for several years, at the same time that losses are mounting.

Panel A may exaggerate the degree of persistence in growth for technology stocks on two accounts. First, the technology stocks are evaluated against the median growth rate of the entire sample of firms, which would include, for example, utility stocks with relatively unexciting growth rates. Second, technology stocks are relatively more volatile, so survivorship bias may be a particularly acute problem in this subsample.

Technology stocks that are intensive in research and development also tend to be glamour stocks with low ratios of book-to-market value of equity. The popular sentiment regarding persistence in growth applies to glamour stocks generally. These stocks typically enjoy higher past growth in operating performance than value stocks with high book-to-market ratios (see Lakonishok et al. (1994)). The evidence from psychology suggests that individuals tend to use simple heuristics in decision making. As LaPorta et al. (1997) argue, investors may think that there is more consistency in growth than actually exists, so they extrapolate glamour stocks' past good fortunes (and value stocks' past disappointments) too far into the future. Panels B and C of Table IV test for consistency in growth for value and glamour stocks, respectively. Value stocks comprise stocks that are ranked

[7] Specifically, the sample includes all firms whose SIC codes begin with 283, 357, 366, 38, 48, or 737. See Chan, Lakonishok, and Sougiannis (2001).

in the top three deciles by book-to-market ratio based on NYSE breakpoints, while glamour stocks represent an equivalent number of stocks with the lowest positive book-to-market ratios. Growth in sales is persistent for both sets of stocks. The results for the other measures of operating performance, however, are not markedly different across the two sets of stocks.

The remaining panels perform our runs tests for large, midcapitalization, and small stocks. Large stocks include stocks in the top two deciles of market capitalization based on NYSE breakpoints as of June in the sample selection year, midcapitalization stocks fall in the next five deciles, and small stocks include the bottom three deciles. While sales growth tends to be more persistent for large firms, it does not translate into persistent growth in income. Of the large stocks, 2.2 percent achieve five-year runs in growth of income before extraordinary items, while 3.2 percent of small stocks achieve the same result (the expected fraction is 3.1 percent).

C. Runs Tests Conditional on Past Growth

It might be expected that firms that have demonstrated consistently superior past growth would be able to maintain their growth in the future. In the case of firms such as Microsoft and EMC, their valuations at year-end 1999 reflected investors' bets that these firms will beat the odds and continue the streak. Table V checks whether firms that have demonstrated consistently high (or low) past growth have continued success in the future.

Part I of Table V applies runs tests to those firms that have achieved superior past growth. In Panel A, at every year-end, we select those firms with above-median growth in each of the prior five years (or three years), and examine their subsequent growth.

Superior past growth in sales carries over into the future. In Panel A1, out of all firms whose sales grow above the median rate each year over the prior three years, on average 305 firms survive over the three years following sample selection. Of these, 70 firms have above-median growth rates in each of the three postselection years. They represent 22.8 percent of the survivors, compared to the 12.5 percent that is expected by chance. Growth in income, on the other hand, is an entirely different matter (Panels A2 and A3). For example, there are 222 firms with the impressive track record of above-median growth in income before extraordinary items in each of the three prior years and that survive over the following three years. Yet over the postselection period, only 28 or 12.5 percent manage to repeat and beat the median over all available firms each year. This matches the number expected under the null hypothesis of independence. Although sample sizes become much smaller in the case of firms with favorable growth over the past five years, the findings are similar. Starting out with roughly 2,900 eligible firms on average, 43 firms enjoy a run over the preceding five years for growth in income before extraordinary items and survive over the subsequent five years. In these five years, the percentage of firms who manage to repeat the run is 5.1 percent, while the percentage expected by chance is 3.1 percent. This corresponds to only one run more than expected, however, so the difference is not outstanding.

Table V

Persistence in Growth Rates of Operating Performance: Firms with Superior and Poor Past Growth

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to five years (or until delisting) for firms with superior (part I of the table) or inferior (part II) past growth in operating performance. Firms with superior (inferior) past growth include: firms with above-median (below-median) operating performance growth each year over the past five or past three years; firms whose average rank on growth rate each year over the past five or past three years falls in the top (bottom) quartile. The sample period is 1951 to 1998, and eligible firms include all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel 1), operating income before depreciation (panel 2), or income before extraordinary items available to common equity (panel 3). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. For each of the following five years, the number of firms with valid growth rates, the number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years, the percentage these firms represent relative to the number of valid firms, and the percentage expected under the hypothesis of independence across years are reported.

	Part I: Firms with Superior Past Growth									
	(A) Firms with Past Above-Median Run									
	Firms with Above-Median Growth each Year for Past 5 Years and Above-Median Growth each Year for Number of Future Years:					Firms with Above-Median Growth each Year for Past 3 Years and Above-Median Growth each Year for Number of Future Years:				
	1	2	3	4	5	1	2	3	4	5
	(A1) Sales									
Average Number of Valid Firms	110	103	96	90	83	355	329	305	285	265
Average Number above Median	70	42	26	17	11	209	118	70	42	26
Percent above Median	63.3	41.0	27.3	19.0	13.7	58.9	35.6	22.8	14.8	9.9
	(A2) Operating Income before Depreciation									
Average Number of Valid Firms	61	57	53	50	47	267	245	227	210	194
Average Number above Median	34	18	10	6	4	136	67	34	18	10
Percent above Median	55.9	32.3	19.4	12.2	8.0	51.1	27.2	15.1	8.8	5.3
	(A3) Income before Extraordinary Items									
Average Number of Valid Firms	53	50	47	44	43	259	240	222	207	193
Average Number above Median	28	14	7	4	2	125	57	28	14	7
Percent above Median	51.9	27.8	15.1	8.4	5.1	48.3	23.7	12.5	6.7	3.6
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	50.0	25.0	12.5	6.3	3.1

Table V—*continued*

	(B) Firms with Past Average Growth Rank in Top Quartile									
	Firms with Average Growth Rank over Past 5 Years in Top Quartile and Above-Median Growth each Year for Number of Future Years					Firms with Average Growth Rank over Past 3 Years in Top Quartile and Above-Median Growth each Year for Number of Future Years				
	1	2	3	4	5	1	2	3	4	5
(B1) Sales										
Average Number of Valid Firms	78	71	66	61	56	204	187	172	159	147
Average Number above Median	47	27	16	10	6	120	67	39	24	15
Percent above Median	60.8	37.7	24.4	16.6	11.4	58.9	35.8	22.8	14.8	9.9
(B2) Operating Income before Depreciation										
Average Number of Valid Firms	35	32	30	27	25	133	121	110	100	91
Average Number above Median	18	8	4	2	1	65	31	15	8	4
Percent above Median	50.6	26.4	15.0	8.9	5.9	49.0	25.4	13.6	7.6	4.7
(B3) Income before Extraordinary Items										
Average Number of Valid Firms	29	27	25	23	22	121	112	103	94	86
Average Number above Median	13	5	3	1	0	56	24	11	5	2
Percent above Median	44.0	19.6	10.2	4.8	2.1	46.4	21.5	10.4	5.5	2.6

Part II. Firms with Inferior Past Growth
(C) Firms with Past Below-Median Run

	Firms with Below Median Growth each Year for Past 5 Years and Above-Median Growth each Year for Number of Future Years:					Firms with Below Median Growth each Year for Past 3 Years and Above-Median Growth each Year for Number of Future Years:				
	1	2	3	4	5	1	2	3	4	5
(C1) Sales										
Average Number of Valid Firms	106	92	82	73	66	343	302	270	244	221
Average Number above Median	35	15	7	4	2	125	59	28	14	7
Percent above Median	33.0	16.3	8.6	4.9	2.5	36.4	19.4	10.6	5.9	3.4

					(C2) Operating Income before Depreciation					
Average Number of Valid Firms	39	35	32	30	28	229	206	186	170	156
Average Number above Median	20	9	5	2	1	122	58	27	13	6
Percent above Median	51.4	25.7	14.3	6.3	3.5	53.3	28.0	14.7	7.6	3.6
					(C3) Income before Extraordinary Items					
Average Number of Valid Firms	33	30	28	26	25	220	201	184	170	157
Average Number above Median	18	9	4	2	1	127	61	28	13	5
Percent above Median	56.2	30.2	14.8	6.7	3.0	57.7	30.4	15.3	7.7	3.4
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	50.0	25.0	12.5	6.3	3.1

	(D) Firms with Past Average Growth Rank in Bottom Quartile									
	Firms with Average Growth Rank over Past 5 Years in Bottom Quartile and Above-Median Growth each Year for Number of Future Years					Firms with Average Growth Rank over Past 3 Years in Bottom Quartile and Above-Median Growth each Year for Number of Future Years				
	1	2	3	4	5	1	2	3	4	5
					(D1) Sales					
Average Number of Valid Firms	86	74	65	57	51	202	175	154	137	123
Average Number above Median	29	12	6	3	1	71	32	14	6	3
Percent above Median	33.1	16.7	8.6	4.4	2.3	35.2	18.1	9.3	4.5	2.3
					(D2) Operating Income before Depreciation					
Average Number of Valid Firms	23	20	17	15	14	111	97	86	77	70
Average Number above Median	15	7	3	1	1	68	33	15	7	3
Percent above Median	63.8	34.8	19.8	8.9	4.2	61.8	33.7	17.5	8.7	4.1
					(D3) Income before Extraordinary Items					
Average Number of Valid Firms	18	16	14	13	12	100	89	80	72	66
Average Number above Median	13	7	4	2	1	68	34	16	7	3
Percent above Median	73.5	47.1	25.1	12.1	5.3	68.1	38.9	20.7	10.3	9.8

The results caution against extrapolating past success in income growth into the future.

A firm may have extraordinary past growth even though it slips below the median for one or two years, as long as growth in the other years is very high. To include such cases of successful past growth, we use a different criterion for what qualifies as superior past growth. In particular, we also classify firms by their average growth ranks. At every calendar year-end over the sample period, we assign each firm a score based on its past growth. The score is obtained by looking back over each of the preceding five (or three) years, ranking the firm's growth rate each year relative to all available firms (where the firms with the highest growth rate and the lowest growth rate get ranks of one and zero, respectively), and then averaging the ranks over five (or three) years. Firms whose average ranks fall in the top quartile are classified as firms with superior past growth in Panel B. While high past sales growth foretells high future sales growth, there are still no signs of persistence in growth of income before extraordinary items in Panel B3. Out of the firms who survive for three years following sample selection, 103 firms have an average rank based on growth over the preceding three years falling in the top quartile. Only 11 or 10.4 percent of them have above-median runs in the three postselection years, amounting to 2 less than the expected number.

In Part II of Table V, Panel C performs the same analysis for firms with below-median growth over each of the past five or past three years. However, survivorship bias is a particularly grave concern here. After a long period of lackluster performance, the firms that are left standing at the end of the following period are particularly likely to be those who post relatively high growth rates. From Panel C1, future sales growth is persistently low. The fraction of above-median runs in sales growth is notably lower than the expected percentage. On the other hand, they are not less likely to achieve favorable above-median runs with regard to future growth in income. For example, looking at firms with a below-median run for the past three years, over the following three- and five-year horizons, the actual (expected) proportions of above-median runs are 15.3 (12.5) and 3.4 (3.1) percent for growth in income before extraordinary items. While survivorship bias makes it difficult to draw a definitive conclusion, it does not appear that, going forward, the firms with disappointing past growth differ notably from the more successful firms with respect to growth in income.

D. Alternative Criteria for Consistency in Growth

Given the large transitory component of earnings, investors may consider a firm to show persistent growth even if its growth fades for a few years, as long as there is rapid growth for the rest of the time. Even a celebrated example of a growth stock such as Microsoft, for example, falls short of delivering above-median growth in income before extraordinary items for 10 years in a row.[8]

[8] In the 10-year period preceding the latest sample selection date, Microsoft's growth rank of 0.49 in 1994 narrowly misses the median that year.

In Table VI, we adopt more relaxed criteria for defining consistency in growth. In particular, we check whether a firm beats the median for most years over the horizon, but allow it to fall short of the median for one or two years. For example, looking forward from a sample selection date, 269 firms on average have sales growth rates that exceed the median in five out of the following six years. These firms represent 15.6 percent of the surviving firms, more than the expected value of 9.4 percent. In the case of income before extraordinary items, the departures from what is expected under independence are slender, especially over longer horizons. For instance, an average of 9.9 percent have income before extraordinary items growing at a rate above the median for five out of six years, which is close to the expectation of 9.4 percent. Similarly, if we let a firm falter for two years, 4.8 percent of the surviving firms have growth in income before extraordinary items that exceeds the median in 8 out of 10 years, compared to an expected value of 4.4 percent.

As another way to single out cases of sustained high growth while allowing for some slack, we require a firm to post an average annual growth rank over the subsequent five years that falls in the top quartile (where in any year a growth rank of one denotes the highest realized growth rate that year, and zero denotes the lowest rate). The results for this definition of consistency are provided in the last column of Table VI. On average, 1.4 percent of the surviving firms (27 firms) pass this criterion with respect to growth of income before extraordinary items. Assuming independence, the expected value is 2.5 percent.

In summary, analysts' forecasts as well as investors' valuations reflect a widespread belief in the investment community that many firms can achieve streaks of high growth in earnings. Perhaps this belief is akin to the notion that there are "hot hands" in basketball or mutual funds (see Camerer (1989) and Hendricks, Patel, and Zeckhauser (1993)). While there is persistence in sales growth, there is no evidence of persistence in terms of growth in the bottom line as reflected by operating income before depreciation and income before extraordinary items. Instead, the number of firms delivering sustained high growth in profits is not much different from what is expected by chance. The results for subsets of firms, and under a variety of definitions of what constitutes consistently superior growth, deliver the same verdict. Put more bluntly, the chances of being able to identify the next Microsoft are about the same as the odds of winning the lottery. This finding is what would be expected from economic theory: Competitive pressures ultimately dissipate excess earnings, so profitability growth reverts to a normal rate.

IV. The Behavior of Nonsurvivors

Survivorship bias is a serious concern in our tests. By necessity, we condition on surviving into the future in order to calculate growth rates and to carry out our runs tests. Moreover, in our runs tests, the survivors are compared each year to all firms (survivors and nonsurvivors) available that year. To gauge the poten-

Table VI

Distribution of Firms Classified by Above-Median Growth in Operating Performance over Indicated Horizon: All Firms

At every calendar year-end over the sample period, growth rates in operating performance are calculated over each of the following one to ten years (or until delisting) for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Operating performance is measured as sales (panel A), operating income before depreciation (panel B), or income before extraordinary items available to common equity (panel C). Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. The table reports the average number of firms with above-median growth in each of the indicated categories, as well as the percentage these firms represent relative to the number of valid firms; the last row reports the percentage expected under the hypothesis of independence across years. Statistics are provided for the entire sample period and for the ten-year horizon corresponding to the last sample formation year.

	Firms with Above-Median Growth							
Variable	3 out of 4 years	4 out of 5 years	5 out of 6 years	6 out of 7 years	6 out of 8 years	7 out of 9 years	8 out of 10 years	Firms with Average Growth Rank in Top Quartile over 5 Years
					(A) Sales			
Average Number	697	432	269	170	287	191	127	79
Percent	33.9	23.0	15.6	10.7	19.5	14.0	10.0	4.2
1989–1998	36.6	26.0	18.0	12.6	21.4	16.0	12.7	5.6
					(B) Operating Income before Depreciation			
Average Number	629	341	184	100	205	119	70	34
Percent	31.2	18.6	10.9	6.5	14.4	9.0	5.7	1.9
1989–1998	31.7	19.3	11.5	7.4	15.1	10.4	8.0	2.0
					(C) Income before Extraordinary Items			
Average Number	634	334	171	88	190	109	61	27
Percent	30.7	17.7	9.9	5.5	12.9	8.0	4.8	1.4
1989–1998	29.9	16.5	8.4	5.0	12.8	8.4	5.7	0.9
Expected Percent	25.0	15.6	9.4	5.5	10.9	7.0	4.4	2.5

tial magnitude of the problem, in this section, we replicate some of our tests on firms who do not survive over the entire future horizon.

Specifically, we examine two sets of stocks. Given our focus on long-horizon growth, we first select at each year-end a sample of firms who survive over the full 10-year following period. The behavior of these (the survivors) is compared to a second set (the nonsurvivors) that also includes firms who do not last for the full period. To strike a balance between the mix of survivors and nonsurvivors in this second set, we require firms to survive for the first five years after sample selection, but they may drop out between the 6th to 10th year of the postselection period.

The results are reported in Panels A and B of Table VII. The survivors have a higher chance than expected for achieving runs above the median in growth of income before extraordinary items. Conversely, the fraction of runs is lower for the set of nonsurvivors. Of the survivors, for example, 3.4 percent sustain runs for five years of growth in income before extraordinary items above the median (where the expected proportion is 3.1 percent). The corresponding percentage for nonsurvivors is 2.3 percent. Nonetheless, the differences across the two sets are generally not substantial. Panels C and D apply the same procedure to the technology stocks considered in Table IV. Here the differences across the two sets are more notable. At the five-year horizon, for example, 5.2 percent of the survivors achieve runs above the median for growth in income before extraordinary items, compared to 3.2 percent of the nonsurvivors.

Finally, Panels A and B of Part II of Table VII give the distribution of one-year growth rates for the two sets of firms (where the percentiles are averaged across all sample selection years). The results confirm that survivors realize higher growth rates than nonsurvivors. For example, the median growth in income before extraordinary items for the survivors averages 10.6 percent, compared to 8.2 percent for nonsurvivors.

V. The Predictability of Growth: Valuation Ratios

Based on the historical record, it is not out of the question for a firm to enjoy strong growth in excess of 20 percent a year for prolonged periods. The issue, however, is whether such firms are identifiable ex ante. Our attempts in the previous sections to uncover cases of persistently high future growth using information such as past growth, industry affiliation, value–glamour orientation, and firm size have limited success. In this section, we expand our search for predictability by investigating whether valuation indicators such as earnings-to-price, book-to-market, and sales-to-price ratios distinguish between firms with high or low future growth. Further, several studies suggest that investors are prone to judgmental biases, so they respond to past growth by extrapolating performance too far into the future (see, e.g., La Porta (1996) and La Porta et al. (1997)). Consequently, after a period of above- or below-average growth, the valuations of firms with high (low) realized growth may be pushed too high (or too low).

In Table VIII, stocks are sorted into deciles at each year-end on the basis of their growth rate in income before extraordinary items over the following five years (Panel A) or over the following 10 years (Panel B). Within each decile, we

Table VII

Results for Surviving versus Non-Surviving Firms: Persistence Tests and Growth Rates

At every calendar year-end over the sample period, two sets of firms are selected: firms that survive over the following ten years (survivors), and firms that survive over the following five years but thereafter fail to survive until the tenth year (nonsurvivors). For each set of firms, growth rates in operating performance are calculated over each of the following ten years. The sample period is 1951 to 1998, and all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files are eligible. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Part I provides runs tests of persistence over each of the following ten years for the two sets of firms: the average number of firms whose growth rate exceeds the median growth rate each year for the indicated number of years is expressed as a percentage of the number of firms with valid growth rates. Part II reports the distribution of annualized growth rates realized over the sixth to tenth year (or until delisting) following sample selection for the two sets of firms. The simple average over the entire sample period of the percentiles is reported.

Part I: Runs Tests for Persistence

	Percent of Firms with Above-Median Growth each Year for Number of Years:									
Variable	1	2	3	4	5	6	7	8	9	10
(A) Survivors (1265 firms)										
Sales	52.8	30.9	18.1	10.8	6.6	4.2	2.7	1.8	1.3	0.9
Operating Income before Depreciation	51.5	26.8	13.7	7.0	3.8	2.1	1.2	0.7	0.5	0.3
Income before Extraordinary Items	51.7	26.9	13.5	6.7	3.4	1.8	1.0	0.5	0.3	0.2
(B) Non-Survivors										
Number of Firms	445	445	445	445	445	344	250	165	86	0
Sales	48.7	26.6	14.6	8.1	4.5	2.8	1.7	1.1	0.8	—
Operating Income before Depreciation	50.0	24.2	11.5	5.5	2.5	1.3	0.7	0.5	0.3	—
Income before Extraordinary Items	49.1	23.8	11.1	5.1	2.3	1.1	0.6	0.3	0.1	—
(C) Survivors, Technology (195 firms)										
Sales	54.6	33.2	20.5	12.9	8.4	5.8	4.2	3.0	2.3	1.7
Operating Income before Depreciation	53.6	29.7	16.5	9.6	5.9	3.6	2.2	1.4	1.0	0.7
Income before Extraordinary Items	54.1	29.9	16.3	9.0	5.2	3.1	1.9	1.1	0.6	0.4

					(D) Non-Survivors, Technology					
Number of Firms	100	100	100	100	100	77	55	37	20	0
Sales	51.5	28.6	16.7	10.6	6.5	4.6	3.1	2.0	1.4	—
Operating Income before Depreciation	49.5	24.3	12.4	6.6	3.3	2.0	1.4	1.3	1.0	—
Income before Extraordinary Items	50.1	25.0	12.4	6.7	3.2	1.7	1.0	0.5	0.0	—
Expected Percent above Median	50.0	25.0	12.5	6.3	3.1	1.6	0.8	0.4	0.2	0.1

Part II: Annualized Growth Rates

	Percentile									
Variable	2%	10%	25%	40%	50%	60%	75%	95%	98%	
					(A) Survivors					
Sales	− 15.4	− 2.0	5.6	9.1	10.9	12.5	15.5	21.7	37.6	
Operating Income before Depreciation	− 23.3	− 6.8	2.8	7.6	10.1	12.5	16.9	25.5	48.0	
Income before Extraordinary Items	− 28.6	− 8.6	2.1	7.7	10.6	13.3	18.1	28.4	56.4	
					(B) Non-Survivors					
Sales	− 18.5	− 7.0	1.0	6.0	8.4	10.4	13.9	20.3	36.8	
Operating Income before Depreciation	− 26.1	− 12.5	− 2.6	4.7	8.1	11.5	16.3	25.7	47.9	
Income before Extraordinary Items	− 27.4	− 14.5	− 3.3	4.4	8.2	11.9	17.9	28.6	55.9	–

Table VIII
Valuation Ratios and Characteristics at Beginning and End of Horizon for Firms Classified by Growth in Income before Extraordinary Items

At every calendar year-end over the sample period, growth rates in income before extraordinary items available to common equity are calculated over the following five and ten years for all firms in the sample. The sample period is 1951 to 1998, and the sample includes all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files. Growth rates are measured on a per share basis as of the sample selection date, with the number of shares outstanding adjusted to reflect stock splits and dividends; cash dividends and special distributions are also reinvested. Firms are classified into one of ten equally-sized categories based on their realized five- and ten-year growth rates. The following statistics are calculated for firms within each category: the median realized annual growth rate over the horizon; the average size decile rank at the beginning and end of the growth horizon; median valuation ratios at the beginning and at the end of the horizon. The ratios are the prior year's income before extraordinary items to price (*EP*), net sales to price (*SP*), and book value to market value of common equity (*BM*). Results are averaged over all years in the sample period, and are also reported for the last five- or 10-year period. Panel A of the table provides results for firms classified by growth rates over five years and for firms with above-median growth each year for five consecutive years; Panel B provides results for firms classified by ten-year growth rates.

	Panel A: Classified by Annualized Growth Rate over 5 Years										
	Decile										
Variable	1	2	3	4	5	6	7	8	9	10	5-year run above median
Median Growth Rate	−18.9	−5.0	1.5	5.8	9.1	12.0	15.1	18.9	25.1	41.7	40.9
Beginning Size Decile Rank	4.118	4.773	5.087	5.423	5.447	5.526	5.338	4.989	4.273	3.272	3.699
Ending Size Decile Rank	3.526	4.414	4.831	5.275	5.452	5.668	5.652	5.482	5.056	4.243	5.163
Beginning Median EP Ratio	0.083	0.085	0.086	0.083	0.084	0.082	0.082	0.082	0.079	0.068	0.061
At Start of Last 5-year Period	0.050	0.056	0.059	0.055	0.060	0.055	0.052	0.047	0.037	0.021	0.033
Ending Median EP Ratio	0.055	0.073	0.078	0.080	0.082	0.081	0.080	0.079	0.077	0.075	0.066
At End of Last 5-year Period	0.033	0.047	0.052	0.053	0.052	0.052	0.049	0.050	0.046	0.042	0.040
Beginning Median BM Ratio	0.650	0.654	0.678	0.665	0.685	0.679	0.694	0.726	0.777	0.880	0.694
At Start of Last 5-year Period	0.465	0.485	0.476	0.465	0.494	0.430	0.458	0.437	0.452	0.537	0.446
Ending Median BM Ratio	1.115	0.927	0.845	0.789	0.755	0.700	0.669	0.610	0.574	0.560	0.369
At End of Last 5-year Period	0.549	0.495	0.501	0.461	0.402	0.367	0.350	0.337	0.291	0.292	0.200
Beginning Median SP Ratio	1.723	1.576	1.473	1.304	1.370	1.276	1.328	1.530	1.791	2.323	1.684
At Start of Last 5-year Period	0.962	1.022	1.079	0.825	0.890	0.807	0.822	1.065	1.052	1.423	0.914
Ending Median SP Ratio	2.606	2.062	1.783	1.501	1.422	1.288	1.274	1.305	1.377	1.503	1.012
At End of Last 5-year Period	1.174	0.860	0.972	0.638	0.653	0.587	0.573	0.649	0.563	0.681	0.460

Table VIII—*continued*

Panel B: Classified by Annualized Growth Rate over 10 years

Median Growth Rate	−10.8	−3.4	−0.3	2.1	3.9	5.6	7.4	9.4	12.4	19.3
Beginning Size Decile Rank	4.565	5.223	5.577	5.641	5.597	5.508	5.563	5.480	5.040	3.890
Ending Size Decile Rank	3.950	5.087	5.608	5.818	5.882	5.921	5.981	6.100	5.851	5.100
Beginning Median EP Ratio	0.088	0.088	0.087	0.087	0.087	0.086	0.085	0.081	0.080	0.069
At Start of Last 10-year Period	0.072	0.070	0.077	0.073	0.074	0.065	0.068	0.066	0.056	0.039
Ending Median EP Ratio	0.057	0.072	0.076	0.079	0.081	0.083	0.084	0.082	0.082	0.079
At End of Last 10-year Period	0.035	0.047	0.050	0.053	0.048	0.054	0.056	0.049	0.044	0.049
Beginning Median BM Ratio	0.653	0.699	0.696	0.699	0.726	0.707	0.723	0.706	0.742	0.817
At Start of Last 10-year Period	0.550	0.605	0.548	0.564	0.595	0.543	0.609	0.504	0.597	0.724
Ending Median BM Ratio	1.048	0.860	0.796	0.761	0.748	0.734	0.725	0.673	0.647	0.622
At End of Last 10-year Period	0.626	0.482	0.382	0.439	0.392	0.396	0.409	0.321	0.343	0.337
Beginning Median SP Ratio	1.664	1.560	1.470	1.392	1.429	1.399	1.415	1.408	1.503	2.022
At Start of Last 10-year Period	1.405	1.417	1.164	1.285	1.054	1.106	1.211	1.133	1.455	1.409
Ending Median SP Ratio	2.619	1.928	1.648	1.531	1.535	1.477	1.478	1.411	1.385	1.468
At End of Last 10-year Period	1.520	0.941	0.735	0.853	0.758	0.826	0.805	0.664	0.724	0.756

calculate the median realized growth rate, as well as median characteristics such as size decile rank and valuation ratios. This is done at the beginning of the 5- or 10-year growth horizon and also at the end of the horizon. We report results averaged across all sample selection years, as well as results for the most recent 5-year or 10-year growth horizon in our sample period.

We focus the discussion on Panel A of the table (the results are similar for the 10-year horizon). In line with the results from Tables I and II, the stocks in the extreme growth deciles tend to be smaller firms. The median firm in the top decile (with a growth rate of 41.7 percent a year) falls in the third size decile, while the median firm in the bottom decile (with a growth rate of -18.9 percent) ranks in the fourth size decile. Over the following 5 years, however, the high-growth firms perform relatively well, resulting in a surge in their market values. Conversely, the market values of the low-growth firms show a relative slump.

Sorting by realized future growth induces a mechanical association between growth rates and the level of earnings at the beginning and end of the growth horizon. To weaken this link, we measure earnings one year prior to the base year (or one year before the final year) of the growth horizon. The price is measured at the start or end of the horizon, so the numbers correspond to the conventional measure of trailing earnings yield that is widely used in practice and research. There is reason to be wary about relying too heavily on the earnings yield variable, however, because net income is the most problematic of our measures of operating performance. For example, a firm may have a low earnings yield because its price impounds investors' expectations of high growth in future earnings, but another reason may be its recent performance has been poor and its earnings are currently depressed. On this account, earnings-to-price ratios are not generally used in academic research, or investment industry analysis, to classify firms as "value" or "glamour" stocks. Instead other, better-behaved, indicators such as the book-to-market ratio, are favored.

The top decile of growth firms at the beginning of the growth horizon has a median earnings-price ratio (0.068) that is much lower than the others (which cluster around 0.08). The low earnings yield for this group is consistent with the notion that the market's valuation accurately incorporates future growth. On the other hand, decile portfolios 8 and 9, which also show relatively strong growth, do not have notably low earnings yields. Rather, the association for the highest-growth decile may reflect cases where firms grow from a depressed level of income. At the end of the growth horizon, only the earnings-price ratio of the bottom decile of firms is eye-catching. Contrary to intuition, however, these firms have comparatively low earnings yields so they appear to be relatively "expensive." Instead, the explanation here may also lie in their low earnings levels, since they have gone through a period of disappointing growth.

Given the shortcomings of the earnings yield variable, we also look at valuation measures that tend to be better-behaved. Table VIII provides median ratios of book-to-market and sales-to-price at the beginning and end of the growth horizon for each decile. Firms which are ranked in the highest decile by earnings growth have relatively high sales-to-price and book-to-market ratios at the beginning. For example, their median book-to-market ratio is 0.880 (compared to 0.690

averaged across the other groups) and the median sales-to-price multiple is 2.323 (compared to 1.486 for the other groups). The modest ex ante valuations suggest that the market fails to anticipate their subsequent growth.

On the other hand, ex post valuations closely track prior growth. The top decile of high-growth firms have ending book-to-market and sales-to-price ratios of 0.560 and 1.503, respectively. These are substantially lower than the averages across all the other groups. This finding fits in with earlier evidence on the existence of extrapolative biases in investors' expectations about future growth (see La Porta (1996) and La Porta et al. (1997)).

The last column in Panel A of Table VIII provides corresponding statistics for firms whose income before extraordinary items grows above the median rate for five consecutive years. The difference between these firms' valuation ratios at the beginning and end of the growth horizon is striking. At the beginning, their book-to-market and sales-to-price ratios are not too far out of line from the average, suggesting that their future performance is not foreseen by the market. However, at the end of the growth horizon, the median book-to-market and sales-to-price ratios of this group are the lowest in Table VIII. The rich ending multiples such firms command highlight the importance investors attach to consistently superior growth, and not just high growth per se. Investors handsomely reward firms that have achieved several consecutive years of strong growth, and believe they will continue the streak (counterfactually, as the results in Table V indicate).

In summary, the results suggest that market valuation ratios have little ability to sort out firms with high future growth from firms with low growth. Instead, in line with the extrapolative expectations hypothesis, investors tend to key on past growth. Firms that have achieved high growth in the past fetch high valuations, while firms with low past growth are penalized with poor valuations.

VI. Comparisons with IBES Consensus Forecasts

Security analysts' estimates of near-term earnings are widely disseminated and receive much attention. Dramatic movements in a stock's price can arise when an influential analyst issues a revised earnings estimate. Possibly, therefore, analysts' estimates of long-term earnings growth may also be useful in forecasting future growth over longer horizons. Analysts are not shy about making aggressive growth forecasts either (the dispersion between the top and bottom decile of IBES long-term forecasts is about 31 percent), so they apparently are confident in their own ability to pick the future success stories.

The current dividend yield on a stock may also have predictive power for future growth in earnings per share. Standard textbook analysis suggests that, given a firm's investment policy and ignoring tax effects, it is a matter of indifference to a shareholder whether earnings are paid out as current dividends or retained for growth in future dividends. For example, a firm may choose to raise the amount paid out from earnings as dividends to current shareholders. To maintain investment, however, it must use external financing, thereby diluting current shareholders' claims to future profits. In other words, high current dividends come at the expense of low future growth per share. To use a simple constant-growth dividend discount model as an illustration, given investors' required rate of return, there is a one-to-one trade-off between future growth per share and the dividend yield. Furthermore, a firm's dividend payout may signal whether it has attractive investment projects available to fuel future growth.

To allow a cleaner comparison with analysts' forecasts, which do not include dividends, in the remainder of the paper, we drop our convention of reinvesting dividends when we calculate growth rates. Analysts' predictions refer to growth in income before extraordinary items, but realized growth in this variable is highly prone to measurement problems (such as the exclusion of cases with negative base-year values for income). For this reason, we also report realized growth in sales and operating income before depreciation. Growth rates in these variables are correlated with growth in income before extraordinary items, but are better behaved and are available for a much larger fraction of the sample.

A. Individual Firm Growth Rates

Table IX relates IBES consensus long-term growth forecasts to realized future growth. At each year-end, we rank all domestic firms with available IBES long-term forecasts and sort them into quintiles. IBES long-term estimates do not become available until 1982, so the sample period in Table IX runs from 1982 to 1998. The breakpoints for the sort use all NYSE firms available as of the sample selection date (regardless of whether they survive in the future). In Table IX, we track the subsequent growth rates of firms who survive over the next one, three, or five years in each quintile. The median realized growth rate over firms in each quintile is then averaged across all sample selection dates.

The dispersion in IBES consensus growth forecasts is large, so analysts are boldly distinguishing between firms with high and low growth prospects. The median estimate in quintile 1 averages 6 percent, while the median estimate in quintile 5 is 22.4 percent on average.[9] Notably, analysts' estimates are quite optimistic. Over the period 1982 to 1998, the median of the distribution of IBES growth forecasts is about 14.5 percent, a far cry from the median realized five-year growth rate of about 9 percent for income before extraordinary items.[10]

Near-term realized growth tends to line up closely with the IBES estimate (Panel A). In the first postranking year, the median growth rate in income before extraordinary items is 18.3 percent on average for quintile 5, and 5.1 percent on average for quintile 1. The difference between the growth rates for the other quintile portfolios is much milder, however. Comparing quintiles 4 and 2, median growth rates in income before extraordinary items are apart by only 2.5 percent.

A naive model for predicting future growth uses the dividend yield, and is based on the trade-off between current dividends and future growth. Suppose,

[9] Note that since the breakpoints are based on NYSE stocks only, the number of stocks differs across the quintiles. In particular, many firms penetrate the top quintile.

[10] To sharpen the point, note that the median realized growth rate of nine percent (without dividends reinvested) is based on all firms, including smaller firms that tend to be associated with somewhat higher growth rates. IBES forecasts, on the other hand, predominantly cover larger firms.

Table IX

Realized Median Growth Rates of Operating Performance for Stocks Classified by IBES Long-Term Growth Forecasts

At every calendar year-end t over the sample period, stocks are ranked and classified to one of five groups based on IBES forecasts of long-term earnings growth. Results are reported for individual stocks and for portfolios. For individual stocks, growth rates in operating performance are calculated over each of the five subsequent years (years $t+1$ to $t+5$) for all firms in the sample with available data. The sample period is 1982 to 1998, and all domestic firms listed on the New York, American, and Nasdaq markets with data on the Compustat files are eligible. Operating performance is measured as sales, operating income before depreciation, or income before extraordinary items available to common equity. Growth in each variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends. The median realized growth over all stocks in each classification is calculated each year, and the simple average over the entire sample period is reported. For portfolios, a value-weighted portfolio is formed at each year-end from all the stocks in each quintile sorted by IBES forecasts. The portfolio's income before extraordinary items is calculated over each of the subsequent five years, with the proceeds from liquidating delisted stocks reinvested in the surviving stocks. Growth rates for each portfolio are calculated in each formation year, and the simple average over the entire sample period of the growth rates is reported. Also reported are the ratios of the prior year's income before extraordinary items per share to current price, and the prior year's cumulative regular dividends per share to current price.

	Quintile Based on IBES Forecast:				
Growth in:	1 (Low)	2	3	4	5 (High)
(A) Growth Rate in Year t+1					
Sales	1.4	4.5	6.3	8.3	13.7
Operating Income before Depreciation	3.6	6.8	7.6	10.3	16.0
Income before Extraordinary Items	5.1	9.5	10.1	12.0	18.3
Portfolio Income before Extraordinary Items	12.6	4.2	4.5	7.2	13.6
No. with Positive Base & Survive 1 year	242	256	266	318	584
No. with Negative Base & Survive 1 year	71	78	60	88	265
(B) Growth Rate in Year t+2					
Sales	1.7	4.5	6.4	7.8	11.6
Operating Income before Depreciation	3.2	7.0	8.4	9.9	14.0
Income before Extraordinary Items	4.7	9.9	10.5	12.2	16.4
Portfolio Income before Extraordinary Items	6.9	7.5	6.1	9.1	10.6
No. with Positive Base & Survive 2 years	225	235	244	296	497
No. with Negative Base & Survive 2 years	62	75	59	85	252
(C) Annualized Growth Rate over 3 Years					
Sales	1.1	4.0	5.6	7.3	11.3
Operating Income before Depreciation	2.5	5.2	6.8	8.1	10.9
Income before Extraordinary Items	3.1	7.4	7.0	9.0	11.5
Portfolio Income before Extraordinary Items	9.0	7.3	5.2	7.1	11.4
No. with Positive Base & Survive 3 years	202	209	230	263	439
No. with Negative Base & Survive 3 years	67	70	56	82	217
(D) Annualized Growth Rate over 5 Years					
Sales	1.2	3.4	5.1	6.9	9.9
Operating Income before Depreciation	2.2	5.1	6.8	7.3	9.2
Income before Extraordinary Items	2.0	6.5	6.5	8.0	9.5
Portfolio Income before Extraordinary Items	8.0	10.7	7.2	7.7	11.3
No. with Positive Base & Survive 5 years	182	179	201	233	356
No. with Negative Base & Survive 5 years	57	63	50	68	170
Median IBES Forecast	6.0	10.2	12.3	15.1	22.4
Median Stock Dividend Yield, %	6.0	3.4	2.7	1.5	0.1
Portfolio Dividend Yield, %	6.9	4.6	3.3	2.5	1.3
Median Stock Earnings to Price Ratio, %	10.0	8.9	7.9	7.2	5.6

as a first approximation, that all stocks have the same long-term expected return. Given this, the naive model forecasts a spread in future growth across stocks that is identical to the spread in their current dividend yields (but in the opposite direction). The naive forecast is quite successful at picking up differences in growth across the intermediate quintiles. Over the first postranking year, the difference between the dividend yields of quintiles 2 and 4 (3.4 and 1.5 percent, respectively) corresponds roughly to the difference in their growth rates. Once differences in the dividend yield are taken into account, then, IBES estimates have forecast power for realized growth over the first year only at the extremes.

In general, IBES long-term forecasts refer to a three- to five-year horizon, so the behavior of realized growth over these horizons is more interesting. Median realized growth rates over three years and over five years are reported in Panels C and D. These panels highlight the upward bias in analysts' long-term growth estimates. In every quintile, median forecasts exceed median realized growth rates, with the most pronounced bias in quintile 5. For five-year growth in income before extraordinary items, for example, the median forecast in the top quintile is 22.4 percent, much higher than the median realized growth rate, which is only 9.5 percent. Furthermore, the realized growth rate for the firms in the top quintile should be taken with a grain of salt. In the highest-ranked quintile, the percentage of firms who survive for the full five postranking years is lower than for any of the other quintiles. For example, there are 849 firms on average who survive in the first postranking year in quintile 5, but this drops to 526 by the fifth year, so about 38 percent of the firms drop out between the first and fifth years. For quintile 3, the corresponding counts are 326 and 251, respectively, so 23 percent disappear from the sample. The upshot is that realized growth in income before extraordinary items is likely to be somewhat overstated for firms in the top quintile.

Over longer horizons, analysts' growth estimates still do not add much information beyond what is contained in the dividend yield. For example, the median realized five-year growth rate is 9.5 percent for the highest-ranked quintile by IBES forecasts, compared to 2 percent for the lowest-ranked quintile. The difference of 7.5 percent is not much higher than the spread in their dividend yields. The yields are 0.1 percent and 6 percent for the highest and lowest ranked quintiles, respectively, so the dividend yield spread is 5.9 percent. The results for growth in operating income before depreciation yield similar conclusions.

To sum up, analysts forecast that long-term earnings growth for the top quintile outperforms the bottom quintile by 16.4 percent. The realized gap in five-year growth rates, however, is only 7.5 percent. Much of the spread in realized growth reflects differences in dividend yields, and some is due to survivorship bias in the top quintile. After accounting for these influences, analyst forecasts add information only over shorter horizons.

B. Portfolio Growth Rates

Issues of survivorship bias and low or negative base-year values for income before extraordinary items are major concerns. Table IX takes another approach to measuring growth rates that tries to work around these concerns. Specifically, after ranking stocks by IBES long-term forecasts at each year-end, we form a value-weighted portfolio of the stocks in each quintile. Value-weighting affords some degree of robustness to our measures, to the extent that problems in measuring growth are less severe for large companies. We then track over the postformation period the income before extraordinary items of the portfolio as a whole. If a stock is delisted in a year after portfolio formation, we assume it generates the average income of the remaining firms in that year. Then, at the end of the year, we take the proceeds from liquidating nonsurviving firms and reallocate them proportionally across the surviving stocks. As a result, we are able to use all eligible companies to calculate growth rates, regardless of whether they survive over the full growth horizon, or whether they have positive earnings in the base year.[11] The portfolio approach, however, is not without its drawbacks. As firms drop out of the sample and the funds from their liquidation are reinvested in the remaining firms, over time, the portfolio can build up large stakes in a relatively small number of surviving firms who tend to have relatively high growth rates. The implication is that long-term portfolio growth rates for cases where survivorship bias is acute, such as the fastest-growing firms in the top quintile by IBES forecasts as noted above, should be interpreted with caution.

The results for the portfolios' long-term growth rates are in line with our earlier findings. IBES long-term forecasts are essentially unrelated to realized growth in income before extraordinary items beyond one or two years out. For example, over the five postformation years (Panel D), the bottom and top quintile portfolios on average experience growth rates of 8 and 11.3 percent per year, respectively. The spread of 3.3 percent in the portfolios' growth rates is smaller than the gap between their dividend yields (5.6 percent).

One difference between our results for individual stocks' growth rates and the portfolios' growth rate concerns the performance of the bottom quintile in the first postranking year. In the year immediately following portfolio formation, the bottom quintile portfolio experiences a strong recovery. Its short-term growth rate (12.6 percent) falls slightly short of the top quintile portfolio's growth rate (13.6 percent). This difference from the earlier results based on individual stocks reflects several methodological details, specifically the use of value-weights, the inclusion in the portfolios of nonsurviving firms as well as firms with negative income, and the use of a time-series average of the yearly portfolio growth rates rather than the cross-sectional medians. In particular, since firms with low IBES forecasts generally tend to start with low or negative values of income before extraordinary items at the portfolio formation date, the growth rate over the following year is likely to be high.[12]

Analysts' forecasts substantially overstate realized long-term growth in the top three quintile portfolios. In the top-ranked quintile, for example, the median projected future growth rate is about 22.4 percent, but the portfolio's realized

[11] The portfolio approach to measuring growth rates is described further in Chan et al. (2000, 2001).

[12] Our results parallel the findings for the prospective earnings growth of beaten-down value stocks documented in Lakonishok et al. (1994).

growth is only 11.4 percent over three years and 11.3 percent over five years. These results suggest that, in general, caution should be exercised before relying too heavily on IBES long-term forecasts as estimates of expected growth in valuation studies. The bottom quintile portfolios by IBES forecasts predominantly comprise firms in mature industries whose growth prospects are relatively unexciting, so analysts' estimates come closer to the mark here. For instance, about 25 percent of the firms in the first quintile are utilities.

The long-term estimates of analysts may be overly optimistic for several reasons. One explanation draws on evidence from studies in psychology that individuals' forecasts are susceptible to cognitive biases.[13] For example, the confirmation bias suggests that individuals tend to focus on evidence that supports their beliefs, while downplaying other information that is inconsistent. In this regard, analysts' estimates will be particularly bullish for glamour stocks that have shown strong past growth and which enjoy favorable investor sentiment. In addition, an analyst is employed by a brokerage firm and is expected to make contributions beyond predicting earnings. Up-beat forecasts may encourage trading by investors and thereby raise commission income, as well as generate investment banking business from firms that receive favorable coverage. The general perception is that these aspects of the brokerage and investment banking business are larger, and their links to analysts closer, in the U.S. market than overseas. As one piece of evidence that such considerations may lead to inflated forecasts, IBES estimates as of mid-2001 for U.S. companies project long-term growth of about 18 percent on average. At the same time, in non-U.S. markets, analysts are forecasting long-term growth for companies of roughly the same size to average 11 percent. Perhaps the close ties that exist in practice between the brokerage and investment banking businesses in the U.S. market foster an environment where analysts tend to be less impartial and err on the side of optimism.

VII. Regression Models

We close out our analysis by gathering all the variables we have previously considered individually into one model in order to take our best shot at forecasting growth. Table X reports the results from cross-sectional regressions to predict future growth in operating profits. The model is

$$\begin{aligned} y_{it+j} = {} & \beta_0 + \beta_1 PASTGS5_{it} + \beta_2 EP_{it-1} + \beta_3 G_{it-1} + \beta_4 RDSALES_{it} \\ & + \beta_5 TECH_{it} + \beta_6 BM_{it} + \beta_7 PASTR6_{it} + \beta_8 IBESLTG_{it} + \beta_9 DP_{it} \\ & + \varepsilon_{it+j.} \end{aligned} \tag{1}$$

The dependent variable, y_{it+j}, is the rate of growth for firm i over year $t+j$ in sales (SALES), operating income before depreciation (OIBD), or income before extraordinary items available to common equity (IBEI). We forecast growth over the first year following sample selection, over the three and five years subsequent to sample selection, and over the second to fifth subsequent years.

[13] The evidence is discussed in Kahnemann and Riepe (1998) and Fisher and Statman (2000).

Table X

Forecasting Regressions for Growth Rates of Operating Performance

At every calendar year-end, a cross-sectional regression model is used to forecast growth rates of operating performance, y_{it+j} for firm i over the following one to five years for all firms in the sample with available data. The model is.

$$y_{it+j} = \beta_0 + \beta_1 PASTGS5_{it} + \beta_2 EP_{it-1} + \beta_3 G_{it-1} + \beta_4 RDSALES_{it} + \beta_5 TECH_{it} + \beta_6 BM_{it} + \beta_7 PASTR6_{it} + \beta_8 IBESLTG_{it} + \beta_9 DP_{it} + \varepsilon_{it+j}.$$

The dependent variable is growth in: sales (SALES); operating income before depreciation (OIBD); or income before extraordinary items available to common equity (IBEI). The variables used to forecast a firm's growth are *PASTGS5*, the growth in sales over the five years prior to the sample selection date; *EP*, the ratio of income before extraordinary items available to common equity to equity market value; *G*, the sustainable growth rate given by the product of return on equity (income before extraordinary items available to common equity relative to book equity) and plowback ratio (one minus the ratio of total dividends to common equity to income before extraordinary items available to common equity); *RDSALES*, the ratio of research and development expenditures to sales; *TECH*, a dummy variable with a value of one for a stock in the technology sector and zero otherwise; *BM*, book-to-market ratio; *PASTRGs*, the stock's prior six-month compound rate of return; *IBESLTGs*, the IBES consensus forecast for long-term growth; and *DP* the dividend yield, accumulated regular dividends per share over the last twelve months divided by current price per share.

Growth in:	*PASTGS5*	*EP*	*G*	*RDSALES*	*TECH*	*BM*	*PASTR6*	*IBESLTG*	*DP*	R^2
				(A) Growth Rate in Year t+1						
SALES	0.0890	0.1641	0.0141	0.0979	−0.0038	−0.0184	0.0365	0.3018	−0.5258	0.0709
	(3.7)	(6.0)	(1.5)	(1.6)	(−0.5)	(−4.7)	(3.0)	(6.1)	(−4.8)	
OIBD	−0.0729	−0.2400	0.0064	0.2047	−0.0045	0.0031	−0.0592	0.2334	−0.5390	0.0274
	(−1.3)	(−3.3)	(0.9)	(1.0)	(−0.3)	(0.4)	(−2.4)	(2.6)	(−3.9)	
OBEI	−0.0971	−0.3982	−0.0242	−0.0024	−0.0162	0.0093	−0.0621	0.1179	−0.9564	0.0263
	(−1.4)	(−3.3)	(−1.5)	(−0.0)	(−0.7)	(0.4)	(−2.0)	(0.9)	(−3.5)	
				(B) Annualized Growth Rate over Years t+1 to t+3						
SALES	0.0469	0.1400	0.0099	0.0974	0.0014	−0.0253	0.0311	0.1901	−0.5758	0.0984
	(1.3)	(5.4)	(1.6)	(3.1)	(0.6)	(−9.2)	(6.8)	(9.3)	(−6.4)	
OIBD	−0.0547	−0.0554	0.0014	0.3453	−0.0127	−0.0073	−0.0089	0.1147	−0.4060	0.0296
	(−1.5)	(−1.8)	(0.1)	(3.1)	(−3.2)	(−1.1)	(−1.7)	(2.0)	(−2.6)	
IBEI	0.0087	−0.1881	0.0011	0.3436	−0.0191	−0.0061	−0.0279	0.0758	−0.0630	0.0257
	(0.5)	(−6.0)	(0.1)	(2.4)	(−2.9)	(−0.4)	(−6.5)	(0.9)	(−0.3)	

Table X—*continued*

				(C) Annualized Growth Rate over Years t+1 to t+5						
SALES	0.0252	0.1074	0.0067	0.0931	0.0014	−0.0260	0.0227	0.1538	−0.5446	0.1175
	(0.7)	(10.5)	(3.6)	(6.8)	(0.4)	(−7.4)	(3.2)	(3.1)	(−16.6)	
OIBD	−0.0645	−0.0146	−0.0035	0.3476	−0.0115	−0.0069	−0.0133	0.1227	−0.2675	0.0367
	(−3.0)	(−0.6)	(−0.5)	(7.6)	(−10.3)	(−1.8)	(−2.3)	(1.5)	(−7.4)	
IBEI	−0.0163	−0.1222	−0.0098	0.2493	−0.0133	−0.0095	−0.0293	0.0729	−0.0917	0.0313
	(−4.2)	(−2.3)	(−0.6)	(3.7)	(−3.0)	(−1.0)	(−2.8)	(0.9)	(−0.7)	
SALES	0.1128	0.0351	0.0628	0.2554						0.0507
	(2.7)	(1.8)	(2.3)	(4.3)						
OIBD	−0.0080	−0.0518	−0.0166	0.3779						0.0150
	(−0.2)	(−3.3)	(−0.7)	(13.1)						
IBEI	0.0311	−0.1295	−0.0675	0.2229						0.0148
	(25.5)	(−3.8)	(−1.5)	(2.4)						
				(D) Annualized Growth Rate over Years t+2 to t+5						
SALES	0.0175	0.0983	0.0060	0.1020	0.0007	−0.0273	0.0218	0.1237	−0.5122	0.0902
	(0.5)	(5.0)	(2.9)	(5.6)	(0.2)	(−6.3)	(3.7)	(2.8)	(−20.1)	
OIBD	−0.0665	0.0136	−0.0147	0.3856	−0.0130	−0.0049	−0.0042	0.1354	−0.3197	0.0335
	(−2.1)	(1.0)	(−1.1)	(4.9)	(−7.7)	(−0.9)	(−0.3)	(1.7)	(−2.7)	
IBEI	0.0119	−0.0932	0.0018	0.2897	−0.0174	−0.0075	−0.0245	0.0809	−0.0538	0.0268
	(0.6)	(−2.6)	(0.1)	(12.8)	(−5.8)	(−0.6)	(−1.8)	(1.0)	(−0.4)	
SALES	0.0962	0.0279	0.0655	0.2515						0.0398
	(2.1)	(1.6)	(3.1)	(5.2)						
OIBD	−0.0097	−0.0255	−0.0023	0.3840						0.0144
	(−0.2)	(−1.2)	(−0.1)	(8.6)						
IBEI	0.0534	−0.1065	−0.0448	0.2310						0.0144
	(3.2)	(−3.3)	(−0.8)	(5.5)						

Growth in each operating peformance variable is measured on a per share basis as of the sample formation date, with the number of shares outstanding adjusted to reflect stock splits and dividends. Values of *PASTGS5, RDSALES, EP, G,* and *PASTR6* are Winsorized at their 5th and 95th percentiles; *IBESLTG* is Winsorized at its 1st and 99th percentiles; and *DP* is Winsorized at its 98th percentile. Stocks with negative values of *BM* are excluded. In the regressions for *OIBD* or *IBEI,* firms with negative values of the operating performance variable in the base year are excluded, as are stocks with ratios of price to the operating performance variable above 100. The reported statistics are the averages over all years of the estimated coefficients, with t-statistics in parentheses, as well as the average R^2 of the model. In panels B to D, standard errors are based on the Hansen-Hodrick (1980) adjustment for serial correlation.

To see whether high past growth is a precursor to future growth, we use *PASTGS5*, the growth rate in sales over the five years prior to the sample selection date. Sales growth is correlated with earnings growth, but is much less erratic and so should yield a relatively more reliable verdict on whether past growth helps to predict future growth.[14]

Simple theoretical models of earnings growth suggest one set of variables that, in principle, should help to predict growth. For instance, a firm's earnings-to-price ratio, *EP*, is widely interpreted as impounding the market's expectations of future growth. We measure this as the firm's income before extraordinary items in the year prior to the sample selection date, relative to its price at the sample selection date. Similarly, in the standard constant-growth valuation model, a firm's sustainable growth rate is given by the product of its return on equity and its plowback ratio. Our proxy for this measure is *G*, where return on equity is measured as the firm's earnings before extraordinary items in the year prior to sample selection, divided by book equity in the preceding year; plowback is one minus the ratio in the prior year of dividends to income before extraordinary items.[15] Finally, to capture the firm's investment opportunities, we use the ratio of research and development expenditures to sales, *RDSALES*. The intensity of R&D relative to sales is widely used in practice as an indicator of how much resources a firm is investing in future growth opportunities (see, e.g., Chan et al. (2001)). When a firm has no R&D spending, we set this variable to zero, so all firms are eligible for the regression.

The forecast equation also incorporates variables that are popularly thought to connote high growth. Firms in technologically innovative industries, or more generally, growth stocks as measured by low book-to-market ratios, are popularly associated with high growth. High past returns for a stock may signal upward revisions in investors' expectations of future growth. Analysts' long-term forecasts are another proxy for the market's expectations of future growth. Finally, the dividend yield may provide information on the firm's investment opportunities and hence ability to grow future earnings. Correspondingly, the other forecasting variables are *TECH*, a dummy variable with a value of one for a stock in the pharmaceutical and technology sectors (defined as in Panel A of Table IV) and zero otherwise; *BM*, the firm's book-to-market value of equity; *PASTR6*, the stock's prior six-month compound rate of return; *IBESLTG*, the *IBES* consensus forecast of long-term growth; and *DP*, the ratio of dividends per share cumulated over the previous 12 months to current price. To be eligible for inclusion in the regression at a given horizon, a firm must have nonmissing values for all the predictors. In addition it must have a positive base-year value for the operating performance indicator in question, so as to calculate a growth rate. To screen out

[14] Results using past five-year growth in *OIBD* or *IBEI* as predictor variables indicate that these variables do a worse job in capturing any persistence in growth.

[15] Firms with negative value of book equity are dropped from the sample for the regression. In cases where the measure for sustainable growth is negative (when income is negative, or when dividends to common exceed income so the plowback ratio is negative), we set the sustainable growth rate variable *G* to zero.

outliers due to low values in the base year, we exclude cases where the ratio of the price to the operating performance variable exceeds 100 in the base year.

The model is estimated each year-end, yielding a time series of estimated coefficients and the adjusted R^2. Means for the time series, and t-statistics based on the standard error from the time series, are reported in Table X. Standard errors from the overlapping regressions in Panels B to D use the Hansen–Hodrick (1980) correction for serial correlation.

The results in Table X deliver a clear verdict on the amount of predictability in growth rates. In line with our earlier results, it is much easier to forecast growth in sales than growth in variables such as *OIBD* and *IBEI*, which focus more on the bottom line. For example, the forecasting model that has the highest adjusted R^2 in Table X is the equation for five-year growth in sales (11.75 percent; Panel C). By comparison, the adjusted R^2 in the equations for *OIBD* and *IBEI* barely exceed 3 percent, so there is relatively little predictability for growth in these variables. If anything, our results may be overstating the predictability in growth. Our cross-sectional regressions are reestimated monthly, so we let the coefficients in the model change over time. As a check on the robustness of our results, we also replicated the regressions in the table using growth rate ranks (ranging from zero for the firm with the lowest growth rate in that year to one for the firm with the highest growth rate). The results from the growth rank regressions echo the findings in Table X.

Our full model includes a total of nine predictors, and the correlations between some of them are quite high. As a result, sorting out the relative importance of each variable is not straightforward. Focusing on the models for *OIBD* and *IBEI*, no variable has coefficients that are statistically significant across all forecasting horizons. The coefficient of past sales growth *PASTGS5* is generally negative, suggesting that there are reversals in growth rates. When past sales have been declining, income levels tend to be low in the base year, resulting in relatively higher future growth rates.[16]

At least over longer horizons (Panels B to D), R&D intensity, *RDSALES*, has the strongest forecast power. In accordance with economic intuition, firms that are investing heavily in R&D, and thereby building up their intangible capital base, on average tend to be associated with elevated future growth. Specifically, a firm that spends 10 percent of its sales on R&D tends to have higher five-year growth in *IBEI* by about 2.5 percent, compared to a firm with no R&D (Panel C). However, the high correlation between *RDSALES* and variables like *TECH* or *DP* suggests caution is warranted in interpreting this result.

The variable *IBESLTG* is provided by supposed experts, and is widely used as a proxy for expected future growth. Its coefficient has the expected positive sign, but it is not statistically significant in the equations for *IBEI*. This variable does somewhat better in the equations for *OIBD*, especially over shorter horizons. In general, however, *IBESLTG* does not have higher forecast power than the divi-

[16] The effect of extremely low base-year values is mitigated to some extent because we drop from the regression cases where the ratio of the price to operating performance indicator exceeds 100 in the base year. However, this is only a partial solution.

dend yield, *DP*, which can be viewed as another proxy for the firm's investment opportunities.[17] In terms of predicting long-term growth, the forecasts of highly paid security analysts are about as helpful as the dividend yield, a piece of information that is readily available in the stock listings of most newspapers.

In line with the results in Table VIII, a low earnings yield *EP* is associated with higher future growth rates, especially for *IBEI*. However, the association is driven by a relatively small number of cases with unusually low base-year earnings. Low values of the earnings base result in a low earnings yield, and given that the firm survives, in an unusually high future growth rate. This explanation agrees with the results in Table VIII, where the relation between *EP* and future growth is confined to companies with the highest growth rates. As further confirmation of this line of reasoning, when we use growth in a variable such as *OIBD*, which is less prone to the problem of a low base level, *EP* does a poor job of forecasting in Table X.

The coefficient of the technology dummy *TECH* is highly significant in many cases, but it generally has an unexpected sign. This may be due to the high correlation between *TECH* and *RDSALES*. For example, dropping *RDSALES* from the model substantially reduces the *t*-statistics for *TECH* (although its coefficient retains a negative sign).

Neither the book-to-market ratio nor our proxy for sustainable growth *G* reliably predicts growth in *OIBD* and *IBEI*. Contrary to the conventional notion that high past returns signal high future growth, the coefficient of *PASTR6* is negative. The explanation for this result echoes our explanation for our findings with respect to *EP*. When a firm's near-term prospects sour and current earnings are poor, stock returns tend to be disappointing as well. Once again, these cases of low base levels of earnings may induce a negative association between past return and future growth.

Panels *C* and *D* also provide results that are based on a simple textbook model for predicting growth. Here the predictor variables are earnings yield, sustainable growth, and R&D intensity. The textbook model has weak forecast power. For example, over a five-year horizon, the adjusted R^2 from the equation for *IBEI* is only 1.48 percent.

VIII. Summary and Conclusions

We analyze historical long-term growth rates across a broad cross section of stocks using a variety of indicators of operating performance. All the indicators yield a median growth rate of about 10 percent per year (with dividends reinvested) over the 1951 to 1998 period. With dividends taken out, the median estimate is the same magnitude as the growth rate of gross domestic product over this period, between 3 and 3.5 percent in real terms. Given the survivorship bias underlying the growth rate calculations, the expected growth rate is likely to be lower. Based on these historical values and the low level of the current dividend

[17] Forecasting models with *IBESLTG* and DP as the only predictors yield qualitatively similar conclusions. In particular, the dividend yield does at least as well as the consensus forecast in forecasting growth.

yield, looking forward, the expected return on stocks in general does not appear to be high. In particular, the expected return using a constant-growth dividend valuation model is about 7.5 percent, assuming there is no mispricing.

Expectations about long-term growth are also crucial inputs in the valuation of individual stocks and for estimating firms' cost of capital. At year-end 1999, a sizeable portion of the market commanded price–earnings multiples in excess of 100. Justifying such a multiple under some relatively generous assumptions requires that earnings grow at a rate of about 29 percent per year for 10 years or more. Historically, some firms have achieved such dazzling growth. These instances are quite rare, however. Going by the historical record, only about 5 percent of surviving firms do better than a growth rate of 29 percent per year over 10 years. In the case of large firms, even fewer cases (less than 1 percent) would meet this cutoff. On this basis, historical patterns raise strong doubts about the sustainability of such valuations.

Nonetheless, market valuation ratios reflect a pervasive belief among market participants that firms who can consistently achieve high earnings growth over many years are identifiable ex ante. The long-term growth expectations of one influential segment of the market, security analysts, boldly distinguish between firms with strong and weak growth prospects. To see whether this belief that many firms can achieve persistently high growth holds up in reality, we use an experimental design that singles out cases where a firm consistently delivers favorable growth for several years in a row. Our results suggest that there is some persistence in sales revenue growth. The persistence in sales does not translate into persistence of earnings, however. Even though we measure consistency against a hurdle that is not particularly challenging (the median growth rate), there are few traces of persistence in growth of operating income before depreciation, or in income before extraordinary items. For example, on average three percent of the available firms manage to have streaks in growth above the median for five years in a row. This matches what is expected by chance. The evidence for persistence is still slim under more relaxed criteria for consistency in growth. All in all, the evidence suggests that the odds of an investor successfully uncovering the next stellar growth stock are about the same as correctly calling coin tosses.

A skeptic might argue that while there is little persistence for the population at large, specific segments of the market are able to improve earnings steadily over long periods. In particular, popular sentiment views firms in the pharmaceutical and technology sectors, along with glamour stocks, as being able to maintain consistently high growth rates. To accommodate this argument, we narrow our search to these subsets of firms. While there is persistence in sales growth, when it comes to growth in bottom-line income, over long horizons, the likelihood of achieving streaks is not much different from sheer luck. Conversely, value firms who are out of favor do not seem to do much worse, although survivorship bias makes it difficult to deliver a definitive verdict. To narrow the search even more, we check whether firms with consistently high past growth manage to maintain their performance going forward. While past growth carries over to future sales growth, the income variables do not display strong persistence.

There is a widespread belief that earnings-to-price ratios signal future growth rates. However, the cross-sectional relation between earnings yields and future growth is weak, except possibly in the cases of firms ranked highest by realized growth. For these firms, an inverse association between ex ante earnings yields and growth may arise because they start from a battered level of earnings in the base year, so future growth is high. In light of the noisiness of the earnings yield measure, academic and practitioner research mainly focuses on other valuation ratios such as book-to-market and sales-to-price. These multiples, which are better behaved, show little evidence of anticipating future growth. On the other hand, firms that enjoy a period of above-average growth are subsequently rewarded by investors with relatively high ratios of sales-to-price and book-to-market. Conversely, investors tend to penalize firms that have experienced poor growth. These results are consistent with the extrapolation hypothesis of La Porta (1996) and La Porta et al. (1997).

Additionally, it is commonly suggested that one group of informed participants, security analysts, may have some ability to predict growth. The dispersion in analysts' forecasts indicates their willingness to distinguish boldly between high- and low-growth prospects. IBES long-term growth estimates are associated with realized growth in the immediate short-term future. Over long horizons, however, there is little forecastability in earnings, and analysts' estimates tend to be overly optimistic. The spread in predicted growth between the top and bottom quintiles by IBES forecasts is 16.4 percent, but the dispersion in realized five-year growth rates is only 7.5 percent. On the basis of earnings growth for portfolios formed from stocks sorted by IBES forecasts, the spread in realized five-year growth rates is even smaller (3.3 percent). In any event, analysts' forecasts do not do much better than a naive model that predicts a one-for-one tradeoff between current dividend yield and future growth per share.

A regression forecasting model which brings to bear a battery of predictor variables confirms that there is some predictability in sales growth, but meager predictability in long-term growth of earnings. Only about three percent of the variation in five-year earnings growth rates is captured by the model. One variable that stands out is the level of research and development intensity, suggesting that a firm's intangible assets may have an important influence on its future performance. On the whole, the absence of predictability in growth fits in with the economic intuition that competitive pressures ultimately work to correct excessively high or excessively low profitability growth.

REFERENCES

Asness, Clifford, 2000, Bubble logic, Working paper, AQR Capital Management.

Bakshi, Gurdip, and Zhiwu Chen, 1998, Stock valuation in dynamic economies, Working paper, University of Maryland.

Ball, Ray, and Ross Watts, 1972, Some time-series properties of accounting income, *Journal of Finance* 27, 663–682.

Beaver, William H., 1970, The time series behavior of earnings, *Journal of Accounting Research* 8 (Suppl.), 62–99.

Beaver, William H., and Dale Morse, 1978, What determines price-earnings ratios?, *Financial Analysts Journal* 34, 65–76.

Brealey, Richard A., 1983, *An Introduction to Risk and Return from Common Stocks* (2nd edition) (MIT Press, Cambridge, MA).

Camerer, Colin, 1989, Does the basketball market believe in the "hot hand"? *American Economic Review* 79, 1257–1261.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1995, Evaluating the performance of value versus glamour stocks: The impact of selection bias, *Journal of Financial Economics* 38, 269–296.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, 2000, New paradigm or same old hype in equity investing? *Financial Analysts Journal* 56, 23–36.

Chan, Louis K. C., Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *Journal of Finance* 56, 2431–2456.

Claus, James, and Jacob Thomas, 2001, Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stocks, *Journal of Finance* 56, 1629–1666.

Daniel, Kent, and Sheridan Titman, 2001, Market reactions to tangible and intangible information, Working paper, Kellogg School of Management, Northwestern University.

Fama, Eugene F., and Kenneth R. French, 1992, The cross section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–155.

Fama, Eugene, F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fama, Eugene F., and Kenneth R. French, 2000, Forecasting profitability and earnings, *Journal of Business* 73, 161–175.

Fama, Eugene F., and Kenneth R. French, 2002, The equity premium, *Journal of Finance* 57, 637–659.

Fisher, Kenneth L., and Meir Statman, 2000, Cognitive biases in market forecasts, *Journal of Portfolio Management* 27, 72–81.

Gebhardt, William R., Charles M. C. Lee, and Bhaskaran Swaminathan, 2001, Toward an implied cost of capital, *Journal of Accounting Research* 39, 135–176.

Hansen, Lars, and Robert Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988, *Journal of Finance* 48, 93–130.

Kahneman, Daniel, and Mark W. Riepe, 1998, Aspects of investor psychology, *Journal of Portfolio Management* 24, 52–65.

La Porta, Rafael, 1996, Expectations and the cross section of stock returns, *Journal of Finance* 51, 1715–1742.

La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *Journal of Finance* 52, 859–874.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Lee, Charles M. C., James Myers, and Bhaskaran Swaminathan, 1999, What is the intrinsic value of the Dow? *Journal of Finance* 54, 1693–1741.

Lintner, John, and Robert Glauber, 1967, Higgledy piggledy growth in America, reprinted in James Lorie and Richard Brealey, eds., *Modern Developments in Investment Management* (Dryden Press, Hinsdale, IL).

Little, I. M. D., 1962, Higgledy piggledy growth, *Bulletin of the Oxford University Institute of Economics and Statistics* 4, 387–412.

Little, I. M. D., and Arthur C. Rayner, 1966, *Higgledy Piggledy Growth Again* (Basil Blackwell, Oxford).

Mezrich, Joseph, Qi Zeng, David Nordquist, and Lakshmi Seshadri, 2001, Rebroadcast: Changing of the guard, *Equity Research Quantitative Strategy* (Morgan Stanley, New York, NY).

Siegel, Jeremy J., 1999, The shrinking equity premium, *Journal of Portfolio Management* 26, 10–17.

Welch, Ivo, 2000, Views of financial economists on the equity premium and on professional controversies, *Journal of Business* 73, 501–537.

CFA Institute

Why so Much Error in Analysts' Earnings Forecasts?
Author(s): Vijay Kumar Chopra
Source: *Financial Analysts Journal*, Vol. 54, No. 6 (Nov. - Dec., 1998), pp. 35-42
Published by: CFA Institute
Stable URL: http://www.jstor.org/stable/4480122
Accessed: 17/03/2010 12:14

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=cfa.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

CFA Institute is collaborating with JSTOR to digitize, preserve and extend access to *Financial Analysts Journal.*

http://www.jstor.org

Why So Much Error in Analysts' Earnings Forecasts?

Vijay Kumar Chopra

Wall Street analysts tend to be too optimistic about the earnings prospects of companies they follow. The average consensus 12-month EPS growth forecast is 17.7 percent, which is more than twice the actual growth rate. In aggregate, forecasts are 11.2 percent above actual earnings at the start of a year and are revised downward continuously in the course of the year. For the full study period reported here, the percentage of 12-month earnings estimates revised downward exceeded the percentage revised upward, on average, by 4.4 percent every month. Since 1993, however, the quality of analyst forecasts seems to have improved. This article provides an intuitive explanation of the change and suggests ways in which analysts can use the explanation to improve portfolio performance.

Use of earnings estimates is an integral part of equity valuation by fundamental and quantitative analysts, and the estimates have even become an integral part of financial reporting in the popular press. The behavior and uses of earnings estimates have been widely studied. I/B/E/S International has published an excellent bibliography of earnings expectation research (Brown 1996). Studies that have shown that analysts tend to overestimate earnings include Clayman and Schwartz (1994), Dreman and Berry (1995), and Olsen (1996). Clayman and Schwartz attributed the positive bias to analysts' tendency to "fall in love" with their stocks. In addition, they proposed that investment banking relationships of investment houses and the prospect of being cut off from access to company managers make issuing negative or critical reports on companies difficult for analysts. Dreman and Berry examined quarterly earnings estimates and found that the average forecast errors tend to be high; in their study, only a small percentage of estimates fell into an acceptable error range. Olsen ascribed the positive bias and lack of accuracy in earnings estimates to herding behavior among forecasters. Francis and Philbrick (1993) argued that analysts make optimistic forecasts to maintain relationships with company managers.

The data for the studies reported here are from the I/B/E/S Global Aggregates database, which aggregates bottom-up analyst earnings forecasts to create forecasts at the market level. The specific forecasts analyzed were for the earnings of the S&P 500 Index. I/B/E/S uses market-capitalization weights to combine the mean earnings forecasts for each company in the S&P 500 into an index of earnings estimates. The data are available on a monthly basis beginning with January 1985; the cutoff point for this study is December 30, 1997.

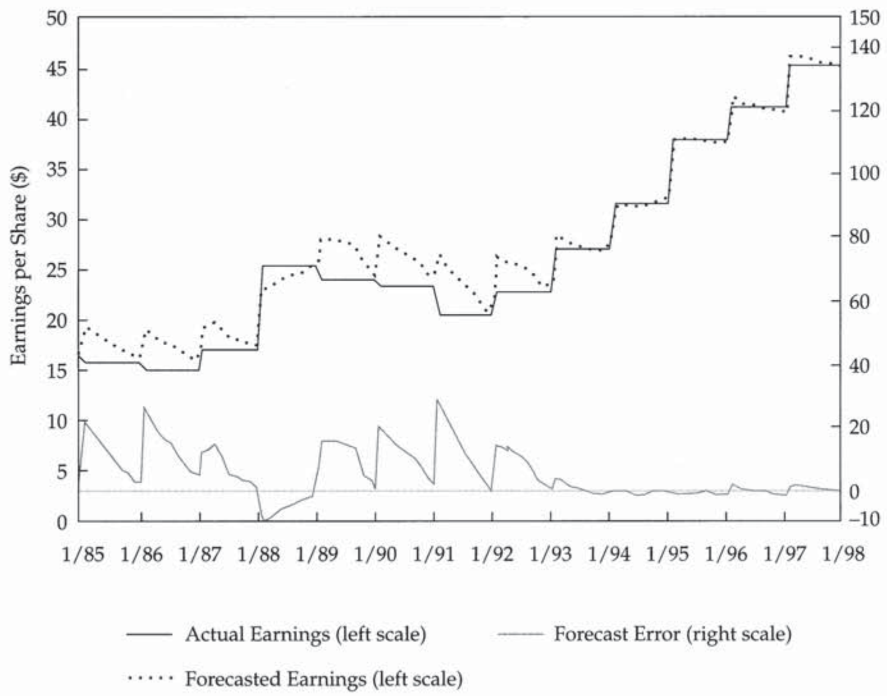
Vijay Kumar Chopra, CFA, is a vice president and senior quantitative portfolio manager in the Global Equity Group at Bankers Trust Company.

Forecast Changes during a Year

This study focused on how the forecasts for the S&P 500 earnings for the current fiscal year vary over the course of the year. **Figure 1** shows the "calendarized" current fiscal year (Calendar FY1 in I/B/E/S terminology) forecasts and actual earnings per share for the entire study period, January 1985 through December 1997.[1] Because of the delay in reporting earnings, the actual earnings are not known until after the year has ended. To make sure that all companies have reported, I used the actual earnings for a calendar year from the I/B/E/S computation made in July of the following year. Therefore, the July 1996 calculation of calendarized 1995 earnings is taken to be the actual earnings for calendarized 1995.

The calendarized actual earnings follow a stair-step pattern. The long-term upward trend and the cyclicality in actual earnings are both evident from Figure 1: Earnings tend to increase over the long run. The cumulative annualized growth rate in earnings for the period is 8 percent, but earnings

Figure 1. Calendarized FY1 Actual Earnings, Forecasted Earnings, and Forecast Errors for the S&P 500: 1985–97



have declined in some periods, such as 1986 and 1989–1991. The earnings recovery since 1992 has produced a steady step-up pattern.

In general, Figure 1 shows that earnings forecasts are very optimistic at the start of the year and decline toward actual values as the year progresses. The decline in full-year forecasts occurs as quarterly numbers are released and an increasing portion of the fiscal-year earnings becomes known. In addition, as the year progresses, company managers comment on the outlook for their companies in future quarters and analysts gather additional information that may lead them to revise their estimates. On rare occasions, analysts underestimate earnings, such as in 1988. For most years, however, analysts revise their initial estimates downward. Future research will have to separate the effect of time from the effect of better visibility for the late quarters of each year.

On average, the Street overestimated current-year earnings by 6.1 percent in the 1985–97 period. In some periods, such as around February 1991, the overestimation was as high as 30 percent, and in other periods, such as February 1988, earnings were underestimated by more than 8 percent. The average overestimation in the 1985–92 period was 9.4 percent.

Since 1993, analyst forecasts have been much closer than in the past to actual earnings. The average forecast error since January 1993 has been remarkably small, an average overestimation of less than 1 percent.

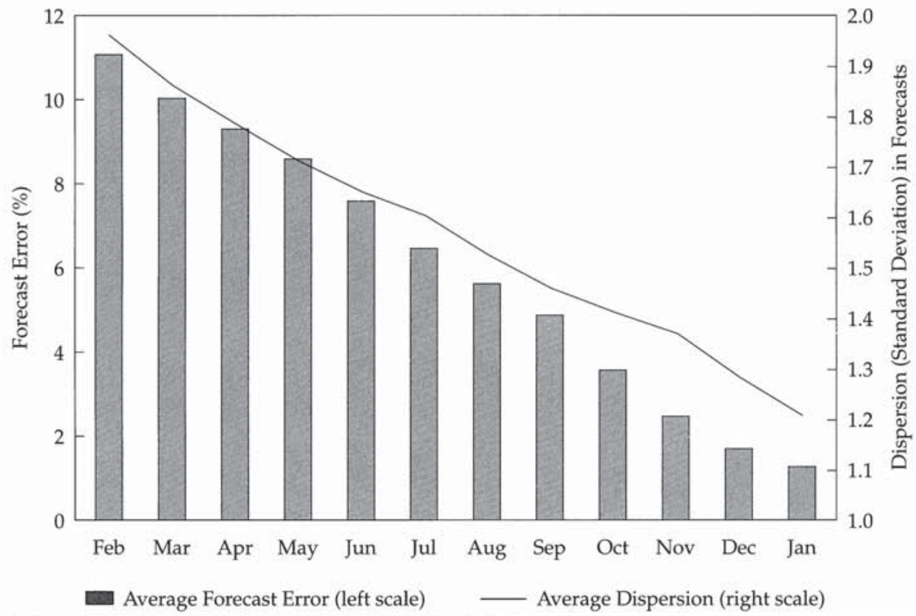
Overestimations typically correct in the course of a year. **Figure 2** shows the decline toward reality of analyst optimism. On average, earnings are overestimated by about 11.2 percent at the start of the fiscal year. (The largest forecast errors occur in February because of the I/B/E/S convention of rolling over a calendar year at the end of January instead of at the end of December.) The overestimation declines to 8.7 percent three months later. Another quarter later, the estimate declines to only 6.6 percent above the actual. By the end of the third quarter, the overoptimism is only 3.6 percent. With attention shifting to the next fiscal year, the final overestimation is only slightly more than 1 percent on average. (Complete convergence does not occur at year end because of the delay in reporting earnings.)

The pattern of declining overestimation was more pronounced before 1993; in the pre-1993 period, the average forecast errors in February were almost 17 percent. At the end of July, they were still well over 10 percent. Since 1993, the error has been as low as 2 percent in February, fading to small negative values from September on.

Another perspective on analyst optimism can be gained by looking at the percentage of estimates of 12-month-forward earnings that are revised upward or downward every month.[2] **Figure 3**

©Association for Investment Management and Research

Figure 2. Analyst Overoptimism and Dispersion in EPS Estimates: Monthly Pattern, Averages for 1985–97



Note: Estimates are from February of a calendar year to January of the following year because of the I/B/E/S February rollover. The initial estimate for Calendar FY1 is made in February, and the final estimate is made in January of Calendar FY2.

Figure 3. Net EPS Estimate Revisions

shows the net positive revisions of 12-month-forward earnings.[3] This series is volatile, but its overall trend is important. Most of the net revisions are negative, which is to be expected; analysts are constantly adjusting their estimates downward because the initial estimates are too optimistic. The average net revision for the entire period, indicated by the shaded line in Figure 3, is –4.4 percent—that is, the percentage of estimates revised downward exceeds the percentage revised upward by 4.4 percent each month. Since 1994, however, net revisions have been close to zero, which confirms the other evidence that analyst forecasts have improved in accuracy since that time.

Consider now another interesting aspect of analyst forecasts—the degree of disagreement among the estimates. Figure 2 shows the decline in the dispersion of estimates over the course of a typical year. The dispersion is greatest in February and declines systematically to its lowest value the following January. This decline can be attributed to quarterly earnings releases and the resulting increase in the visibility of the company's prospects. For the whole study period, dispersion in estimates at the level of the S&P 500 exhibits the sawtooth pattern shown in **Figure 4**. Analyst estimates of Calendar FY1 earnings show the greatest disagreement at the start of the year. As companies report interim quarterly results, the proportion of the fiscal year for which earnings have to be forecasted declines, which reduces the divergence in Calendar FY1 estimates as the year proceeds. This pattern has been particularly strong since 1988 and does not show any signs of fading in recent years. Although analysts may have gotten better at estimating the year's overall level of earnings, the disagreement among analysts over earnings estimates has not diminished over the years.

Forecasted versus Actual EPS Growth

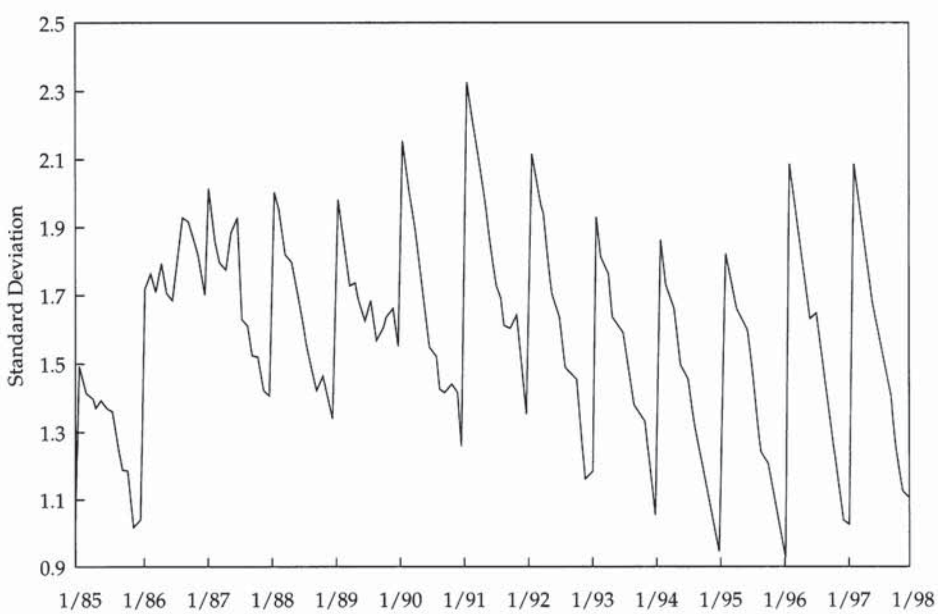
Analysts' earnings growth rate forecasts provide another perspective on the overoptimism evident in their forward estimates of EPS. **Figure 5** shows the rolling 12-month-forward actual and forecasted growth in S&P 500 earnings. For example, the 12-month forecasted growth rate in March 1986 was 16.6 percent whereas the actual growth rate for the subsequent 12 months was –2 percent.

Figure 5 provides three key insights into analyst behavior. First, earnings growth forecasts are always positive. The forecasts lie roughly in the 10–30 percent range, with an average of 17.7 percent, whereas actual growth averages 8.6 percent, almost 9 percent below the forecasts on an annual basis. Therefore, on average, analysts' forecasts are double the actual growth rate in earnings.

Second, actual earnings growth rates vary a lot more than the forecasted rates. Actual earnings growth varies between –15 percent and 40 percent, whereas the forecasts lie within a much narrower range, 10–30 percent. The standard deviation of forecasted growth rates is only 5.4 percent, compared with a 12 percent standard deviation for actual earnings growth rates. Note that, in aggregate, analysts never forecast an absolute decline in earnings, but actual earnings have fallen for extended periods of time (e.g., January 1985 to June 1986, which coincided with a rapid decline in the pace of economic activity and a collapse in the price of oil, and again from January 1989 through June 1991, which was a time of brief economic recession).

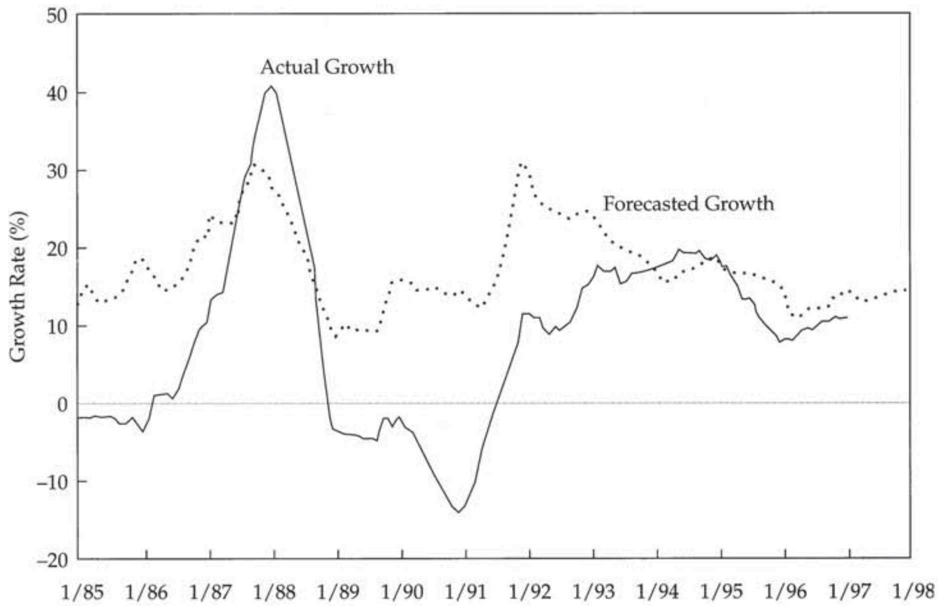
Third, Figure 5 shows that, as with EPS levels, actual and forecasted EPS growth rates have been much closer since January 1993. **Table 1** summarizes the forecasting behavior of analysts for the

Figure 4. Dispersion in Analyst EPS Estimates over Time



©Association for Investment Management and Research

Figure 5. Forecasted versus Actual EPS Growth Rates



Note: The actual growth rates end in December 1996, whereas the forecasted growth rates are available through the end of 1997 because the actual growth rate is not known until 12 months after a given month-end. For example, the actual growth rate for March 1986 comes from March 1987 data.

Table 1. Twelve-Month-Forward Forecasted and Actual Earnings Growth Rates: Summary Statistics

Period/Statistic	Forecasted Growth Rate	Actual Growth Rate	Difference in Rates
January 1985 to December 1996			
Mean	17.7%	8.6%	9.1%
Standard deviation	5.4	12.0	9.3
Maximum	31.2	41.0	28.7
Minimum	8.4	–14.0	–13.1
January 1993 to December 1996			
Mean	16.5	14.4	2.1
Standard deviation	3.2	3.9	2.8
Maximum	24.3	19.5	8.3
Minimum	10.9	7.7	–2.9

Note: The difference between forecasted and actual growth rates is a new series. The last column shows the mean, standard deviation, maximum, and minimum for this series.

whole study period and the post-1993 periods. The average forecasted growth rate of 16.5 percent since January 1993 reported in Table 1 is only about 2 percent higher than the actual increase of 14.4 percent. The standard deviations have also been closer, at 3.2 for the forecast versus 3.9 for the actual.

The correlation between average forecasted and actual EPS growth rates for the total period is 0.67, which indicates that analysts have done a moderately good job of capturing changes in EPS growth rates over time. The correlation for the 1993–97 period was 0.70.

Does the recent convergence between analyst forecasts and actual EPS indicate a sudden increase in analyst forecasting ability? Possibly, but the more likely explanation is that analysts have continued to predict optimistic growth rates but those predictions turned out to be in line with actual rates that were high by historical standards. That is, because of restructurings during the previous decade, when the economy started strengthening in 1992, earnings per share grew strongly to match the usual analyst optimism. This explanation is supported by a comparison of rates since January 1993 with rates for the whole period. The forecasted growth rates are very close, 16.5 for the recent period and 17.7 for the whole period, which indicates that analyst optimism did not decline; the actual growth rate for the recent period, however, was almost 6 percentage points higher than growth for the whole period. In short, the actual growth rate for January 1993 through December 1997 has been close to the long-term average growth forecast in what has been one of the longest economic expansions in the history of the United States.

Economic Growth and Earnings Growth

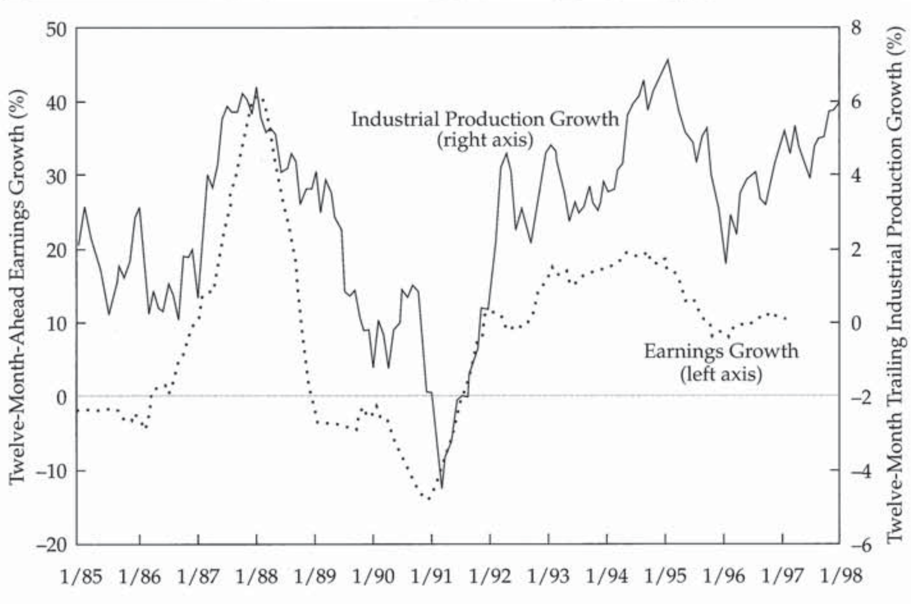
At the aggregate level, company earnings are likely to be tied to the state of the economy. Strong economic growth should, therefore, lead to strong growth in EPS, and indeed, a comparison of growth in industrial production with earnings growth for the S&P 500 supports that expectation.[4]

Figure 6 provides plots of the year-on-year growth in industrial production and the year-on-year growth in actual earnings. Earnings growth lags industrial production growth by between 9 and 18 months, with an average of about 12 months. In order to highlight the close link between growth in industrial production and EPS growth, the earnings growth has been shifted back by 12 months; that is, for example, the June 1996 growth in industrial production is the growth for June 1995 to June 1996 and the June 1996 earnings growth is the growth from June 1996 to June 1997.

Figure 6 suggests that investment analysts could predict aggregate earnings using industrial production data. The correlation between the growth of the two series is 0.77. When industrial production is lagged by one additional month to account for the late release of the data, the correlation is still very high, 0.73. In comparison, the correlation between forecasted and actual earnings growth rates has been averaging 0.67.

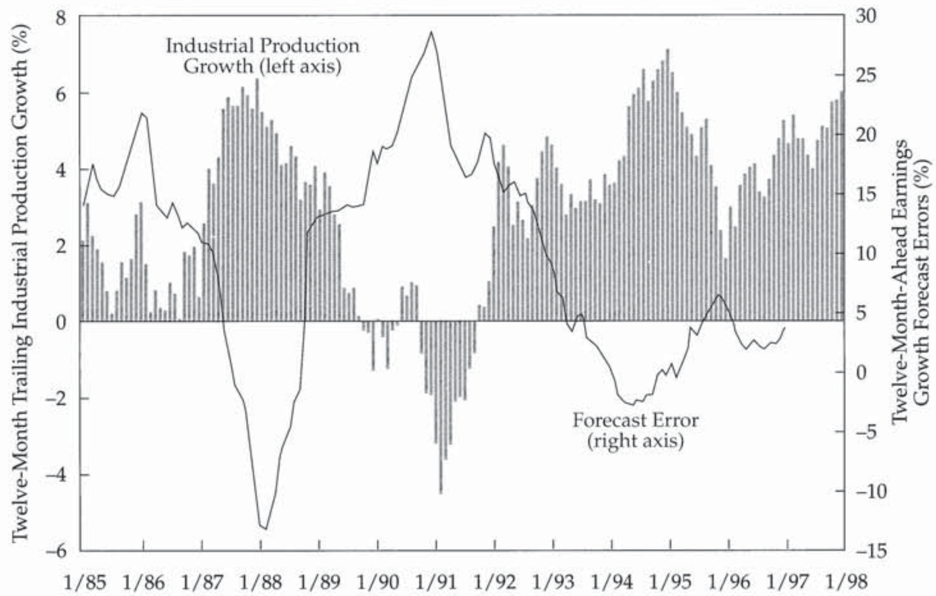
An exploration of the link between the strength of the economy and earnings growth estimates will shed considerable light on why earnings estimates are consistently off the mark and why they have been closer to actual earnings since 1993. **Figure 7** shows the year-on-year growth in industrial production and plots the error in the 12-month-forward earnings growth forecast (the difference between the 12-month-forward forecasted earnings growth and actual earnings growth). The clear inverse relationship between the two series indicates that forecast errors are greatest when industrial production growth is at a peak or trough. Furthermore, when industrial production growth accelerates, forecast errors decline, and when industrial production decelerates, forecast errors increase. When growth in industrial production accelerates, earnings grow strongly and the gap between the optimistic growth forecasts and actual earnings growth narrows, which results in more-accurate forecasts. When growth in industrial production decelerates, earnings growth declines

Figure 6. Industrial Production Growth and Aggregate EPS Growth



©Association for Investment Management and Research

Figure 7. Industrial Production Growth and Errors in EPS Growth Forecasts



(with a 12-month lag) and the gap between the optimistic forecasts and actual earnings growth widens, which results in inaccurate forecasts. When industrial production growth is at its peak, the forecast errors overshoot on the downside and are large but negative. An example is the fourth quarter of 1987 through the first quarter of 1988. On the other hand, when the growth in industrial production started declining in January 1988 from 6.4 percent down to –4.5 percent in March 1991, the forecast errors went from –13 percent to almost 29 percent.

In light of this evidence on growth in the economy and analysts' forecasts, the aggregate behavior of analysts can be described as follows: They are normally very optimistic. When economic growth strengthens, actual earnings accelerate toward the normally optimistic forecasts, so forecast errors decline. If economic growth is very strong, earnings rise well beyond the forecasts, so analysts end up underforecasting earnings for a while. When the economy slows down, earnings start declining but the analysts' optimism prevents them from reducing their estimates far enough. Therefore, the size of forecast errors increases. If forecast errors are negative when the economy starts to slow down, as in January 1988, the errors become less negative at first; then, as the economy continues to decelerate and moves into a recession, the forecast errors move into the positive range and continue to grow. In December 1990, the errors hit a peak of almost 29 percent.

This behavior implies that analysts are likely to be most accurate in an environment of continuing strong economic growth, when earnings growth will approach the analysts' usually bullish forecasts—as has been the case since early 1992. The worst economic environment for aggregate analyst forecasts is one of an accelerating or decelerating economy, and the faster the pace of acceleration or deceleration, the greater the deviation between forecasts and actual earnings growth. The bottom line is that analysts will continue to forecast inaccurately as long as business cycles exist.[5]

Investment Implications

Users of EPS estimates will clearly benefit from recognizing the extent of analyst optimism. Valuation models that rely on earnings forecasts are likely to be biased, but if the extent of optimism is similar across industries and sectors, these valuation models will still be useful in evaluating stocks relative to each other.

The finding that forecast errors vary systematically with the business cycle suggests that analysts may focus too much on firm-specific issues and not enough on the overall macroeconomic environment. Portfolio managers could improve portfolio performance, therefore, by adjusting consensus earnings for systematic biases in forecasts.

One of the uses of aggregate estimate data is in global asset allocation, and conventional asset allocation approaches rely on comparing earnings yields with interest rates. Emanuelli and Pearson (1994) described an approach to global asset allocation that relies on estimate revisions. Recognizing that biases in earnings forecasts are linked to the business cycle and adjusting earnings forecasts to reduce the bias will improve the performance of such global asset allocation strategies.

Conclusion

Analysts' forecasts of EPS and growth in EPS tend to be overly optimistic. Calendarized earnings estimates overstate actual earnings by about 11 percent at the start of the year. These estimates are revised downward monotonically as a typical year unfolds. On average, the percentage of 12-month earnings estimates revised downward exceeds the percentage revised up by 4.4 percent a month. Analyst forecasts of 12-month earnings growth rates average 17.7 percent, more than twice the actual growth rate in the past 13 years.

Industrial production is a good predictor of earnings growth for a year in the future; the correlation is 0.77 percent. The analyst forecast for aggregate EPS growth is also a good predictor of actual growth (with a correlation of 0.67), but the forecasted growth rates are generally too optimistic and lie in a narrow (10–30 percent) range whereas the actual growth rates have varied from –10 percent to 40 percent.

Analysts' usual optimism, their tendency to forecast in a narrow and comfortable range, and the business cycle prove to be the bane of their forecasts. Acceleration or deceleration in economic growth tends to catch analysts off-guard. The forecasts are most accurate in an environment of continued strong growth, such as the one the U.S. economy has been in since 1992. Therefore, although the quality of forecasts has improved since 1992, it will deteriorate if and when the U.S. economy slows down and reverts to its historical cyclical pattern.

Notes

1. I/B/E/S uses the "Compustat rule" to calendarize company-level data prior to aggregation. Data for fiscal years ending between January and May are included in the aggregate for the prior calendar year. Data for the fiscal years ending between June and December of the current calendar year are included in the current calendar-year aggregate (Calendar FY1). For example, data for a company with a fiscal year ending in March 1996 are in the 1995 aggregate; data for a company with a fiscal year ending August 1996 are in the 1996 aggregate. I/B/E/S applies a February "rollover"; that is, when the calendar year ends and a new calendar year begins, the data for Calendar FY1 should shift or roll over from the year just ended to the new year, but I/B/E/S lags the shift by one month. Therefore, the current calendar year is not considered Calendar FY1 until February. The rationale for the lag is, presumably, that a majority of the companies with fiscal years ending in December do not report by the end of January.
2. I/B/E/S calculates 12-month-forward estimates for a company by prorating the current and next fiscal year estimates using the formula $[(a/12)(\text{Current fiscal year EPS}] + [(12 - a)]/[12(\text{Next fiscal year EPS})]$, where a is the number of months remaining in the current year. I/B/E/S then aggregates 12-month-forward company estimates to the index level.
3. Net revisions are defined as (Number of estimates revised upward – Number of estimates revised downward)/Total estimates, over the preceding four weeks, in percentage terms.
4. I used industrial production as a measure of economic activity instead of GDP because of the monthly availability of production data. Using GDP produced qualitatively similar results.
5. This link between forecast errors and the business cycle contrasts with the findings of Dreman and Berry, who found that forecast errors are not meaningfully affected by the business cycle.

References

Brown, Lawrence, ed. 1996. *I/B/E/S Research Bibliography*. 5th ed. New York: I/B/E/S International:xvii–xviii.

Clayman, Michelle, and Robin Schwartz. 1994. "Falling in Love Again—Analysts' Estimates and Reality." *Financial Analysts Journal*, vol. 50, no. 5 (September/October):66–68.

Dreman, David, and Michael Berry. 1995. "Analyst Forecasting Errors and Their Implications for Security Analysis." *Financial Analysts Journal*, vol. 51, no. 3 (May/June):30–41.

Emanuelli, Joseph, and Randal Pearson. 1994. "Using Earnings Estimates for Global Asset Allocation." *Financial Analysts Journal*, vol. 50, no. 2 (March/April):60–72.

Francis, Jennifer, and Donna Philbrick. 1993. "Analysts' Decisions as Products of a Multi-Task Environment." *Journal of Accounting Research*, vol. 31, no. 2 (Autumn):216–30.

Olsen, Robert. 1996. "Implications of Herding Behavior for Earnings Estimation, Risk Assessment, and Stock Returns." *Financial Analysts Journal*, vol. 52, no. 4 (July/August):37–51.

 ©Association for Investment Management and Research

Available online at www.sciencedirect.com
SCIENCE DIRECT
International Review of Financial Analysis
14 (2005) 1–22
IRFA INTERNATIONAL REVIEW OF Financial Analysis

Trends in analyst earnings forecast properties

Stephen J. Ciccone*

Whittemore School of Business and Economics, University of New Hampshire, McConnell Hall, 15 College Road, Durham, NH 03824, USA

Abstract

Forecast dispersion, error, and optimism are computed using 120,022 quarterly observations from 1990 to 2001. Forecast dispersion, error, and optimism all decrease steadily over the sample period, with loss firms showing an especially striking decrease. By the end of the sample period, dispersion and error differences between profit and loss firms are relatively minor, optimism for loss firms is around an unbiased 50%, and pessimism dominates profit firms. Additionally, loss firm earnings appear more difficult to forecast. The reduction in dispersion, error, and optimism does not appear fully attributable to earnings management, earnings guidance, or earnings smoothing. The trends are consistent with increased litigation concerns.
© 2004 Elsevier Inc. All rights reserved.

JEL classification: D84; G14; G24
Keywords: Earnings forecasts; Analysts; Earnings management; Forecast dispersion; Forecast accuracy

1. Introduction

A major responsibility of analysts is to make earnings forecasts. Professionals, such as investment bankers, financial advisors, and stockbrokers, rely on these forecasts to make their decisions, as do many individual investors. The forecasts serve as critical inputs into stock valuation models. Earnings announcement period returns are influenced by the forecasts (e.g., Imhoff & Lobo, 1992), and forecast dispersion is even related to monthly or annual stock returns (Ang & Ciccone, 2001; Diether, Malloy, & Scherbina, 2002; Dische, 2002). Forecasts are now publicly available on many investment-related web sites, providing free access to millions of investors all over the world.

* Tel.: +1-603-862-3343; fax: +1-603-862-3383.
E-mail address: stephen.ciccone@unh.edu (S.J. Ciccone).

1057-5219/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.irfa.2004.06.001

For a long period of time, the ability of analysts to forecast earnings was questioned. Analysts were biased some argued, optimistic and unresponsive to earnings changes (Abarbanell & Bernard, 1992; DeBondt & Thaler, 1990). They tended to herd, making forecasts or recommendations similar to other analysts (Hong, Kubik, & Solomon, 2000; Olsen, 1996; Stickel, 1990; Trueman, 1994; Welch, 2000). They were better than time-series earnings estimates, but only slightly (Fried & Givoly, 1982; O'Brien, 1988).

Recent studies have found that analyst forecasts have changed, perhaps even improved. Analysts have reduced both the size of their forecast errors and their optimism (Brown, 1997; Matsumoto, 2002; Richardson, Teoh, & Wysocki, 2001). Unfortunately for the analysts, many attribute this trend, not to better forecast accuracy, but to increases in earnings guidance, management, or smoothing (e.g., Degeorge, Patel, & Zeckhauser, 1999; Matsumoto, 2002).

The purpose of this study is twofold, both to document trends in forecast properties and to differentiate among theories as to why the trends exist. Several trends are investigated; some revisited, some new: (1) the trends of dispersion, error, and optimism; (2) the trend of wrongly forecasted profits or losses; (3) the trend of naïve forecast performance versus analyst forecast performance; (4) the trend of earnings volatility; and (5) the trend of Street versus GAAP earning differences. In addition, the influence of Regulation FD on the trends is examined. Quarterly data is used during a 1990 to 2001 sample period. As previous research has shown that analysts have greater difficulty forecasting the earnings of firms with losses (Brown, 2001; Butler & Saraoglu, 1999; Ciccone, 2001; Dowen, 1996; Dreman & Berry, 1995), firms with profits and losses are separated and examined independently in much of the testing.[1]

There are several possible explanations for changes in forecast properties: legal liability (e.g., Skinner, 1994), earnings guidance (e.g., Matsumoto, 2002), earnings management (e.g., Degeorge et al., 1999), earnings smoothing (consistent with Bartov, 1993), or information flow improvements (consistent with Asthana, 2003). The testing investigates the validity of these reasons.

The results are quite remarkable. Forecast properties have undergone an extraordinary change, perhaps best called a transformation, during the sample period. Forecast dispersion and error both decrease throughout the sample period, with most of the decrease due to loss firm forecasts. Although analysts still do not forecast loss firms with the same degree of accuracy as profit firms, the differences in forecasting performance are steadily eroding.

Optimism also decreases as analysts moved from being optimistically biased to being pessimistically biased during the sample period. The pessimism associated with profit firms is astonishing. Near the end of the sample period, almost three quarters of the

[1] Several related studies exist. Brown (1997), Richardson et al. (2001), and Matsumoto (2002) all show a decreasing trend in signed earnings surprise or optimism, although they do not separate firms by profitability. Gu and Wu (2003) evaluate forecast differences between profit and loss firms but do not examine trends in performance. Dreman and Berry (1995) and Butler and Saraoglu (1999) do separate firms by profitability while examining trends, but both rely on sample periods ending in 1991. Brown (2001) uses the signed, earnings surprise of the last forecast made prior to the earnings release date to examine shifts in the trend of the median surprise for profit and loss subsamples.

quarterly forecasts for profit firms are pessimistic. Analysts still tend to be optimistic toward loss firms, but this optimism has decreased dramatically over the sample period, hovering around an unbiased 50% at the end of the period. The decrease in the optimistic biases is so pronounced that the still-lingering legend of analyst earnings optimism (e.g., Easterwood & Nutt, 1999; Gu & Wu, 2003) is clearly no longer true, even for loss firms. If anything, analysts have a new concern: earnings pessimism for profit firms.

Additional results show that analysts have gotten much better at predicting the sign of earnings when firms report losses. Moreover, forecasting loss firm earnings appears to be much more difficult than forecasting profit firm earnings. Given this difficulty, analysts actually seem to provide greater value to the market when forecasting for loss firms.

Finally, the results suggest that the trends in forecast properties are unlikely to be fully attributable to earnings guidance, management, or smoothing. Firms unlikely to manage earnings—those with negative surprises, earnings declines, and losses—experience similar reductions in dispersion and error as the sample of all firms. So do firms considered unlikely to be guiding firms toward a specific earnings target, those with high dispersion. Furthermore, Street versus GAAP earnings differences and earnings volatility do not affect the results. The trends in forecast properties are consistent with litigation concerns, especially those surrounding loss reporting. In addition, although not specifically tested, analysts, aided by new information technology, may have simply improved in their forecasting abilities.

2. Forecast property changes

One of the most prominent explanations for the changing trends in forecast properties centers on earnings management. In the financial press, managers are often thought to play an ‘‘earnings game,’’ manipulating reported earnings (and hence the surprise) to reap various benefits: increased stock prices, favorable publicity, and bonuses (Vickers, 1999). Fox (1997) tells of a Microsoft 1997 quarterly earnings release in January, the 41st time in 42 consecutive quarters that Microsoft met or beat the Wall Street consensus. The earnings game is often considered dangerous: when played long-term prospects are sacrificed by concern with short-term profits. Corporate decisions are altered, accounting rules are stretched, and investors lose faith in both financial statements and stock prices (Collingwood, 2001).

Academics have intensively investigated the issue of earnings management. Burgstahler and Dichev (1997) and Degeorge et al. (1999) find that firms manage earnings to meet analyst expectations, avoid losses, and avoid earnings declines. These studies mention several reasons why executives manage earnings, including increased job security, increased bonuses, and bolstered investor interest. Furthermore, anecdotal evidence suggests that firms like the favorable publicity of positive surprises, profits, and earnings increases. Of the three objectives identified by Degeorge, Patel, and Zeckhauser, the positive profit objective proves predominant. However, missing a consensus earnings estimate can be very costly to a firm. For example, Skinner and Sloan (2002) find that, all else equal, the price decline after a negative surprise is greater than the price increase following a positive surprise.

Another way of managing earnings entails "smoothing" or making earnings less volatile through time (e.g., Bartov, 1993). There are several theories that attempt to explain this behavior. Healy (1985) and Holthausen, Larcker, and Sloan (1995) find smoothed earnings are related to management bonus arrangements. Degeorge et al. (1999) use these findings to argue that managers may reduce high earnings levels to make future earnings objectives easier to meet. Fudenberg and Tirole (1995) argue that managers will boost earnings in bad times to increase the probability of retaining their jobs. Trueman and Titman (1988) believe that firms smooth earnings to lower their perceived bankruptcy risk and thus lower their cost of debt.

A cheaper way of playing the earnings game involves forecast guidance. Firms guide analysts toward a pessimistic target and then beat that target (Matsumoto, 2002), an easy way to garner favorable publicity.

An additional perspective on earnings guidance is rooted in legal liability issues. Firms face scrutiny when reporting large, unexpected losses. The consequent stock price decrease angers investors, who then might sue the firm for damages, consistent with Skinner (1994, 1997). Kasznik and Lev (1995) provide support for this argument by showing that firms increased their tendency to warn investors of impending losses. By warning of losses, firms are not necessarily playing an earnings game. As such, guiding analysts toward pessimistic targets and warning analysts of losses, although related, are considered two distinct concepts in this study.

Simpler explanations also exist to explain forecasting trends. For example, an alternative viewpoint looks at data availability and the information revolution, consistent with Asthana (2003). Forecasting techniques might be improving, aided in part by more precise and timelier economic information. Communications channels between firm managers and analysts may be better. Perhaps even the recent proliferation of freely available financial information on the Internet makes analysts more careful as they strive to add value and provide information above and beyond what is known by individual investors.

3. Data and methodology

The First Call summary database is used to obtain the forecast properties. Quarterly forecasts are used to present all results. The results using annual forecasts are similar to the quarterly results and do not require separate analysis. The last mean forecast available prior to the fiscal period end is used as the consensus forecast. All conclusions are similar if median forecasts are used instead of the mean forecasts or if the last mean forecasts prior to the earnings release are used instead of the last mean forecasts prior to fiscal period end.

Forecast dispersion is defined as the standard deviation of the forecasts divided by the absolute value of the mean forecast. This measure requires at least two forecasts.[2] Forecast error is defined as the difference between the actual earnings and the mean forecasted

[2] Although the procedure sharply reduces the sample size, the results for dispersion are similar if only companies with five or more analysts are included.

earnings, divided by the actual earnings. The absolute value is taken to obtain the final error number. A ''raw error'' is also computed as the absolute value of the difference between actual and forecasted earnings (i.e., the error is not deflated).[3] A forecast is considered optimistic if the mean forecast is greater than the corresponding actual earnings. The error and optimism measures require at least one forecast.

Many studies deflate the forecast properties by the stock price rather than the deflators described above. Thus, as a check, trends in dispersion and error are reexamined using price at the beginning of the fiscal year as the deflator. These results are qualitatively similar to the presented results, although the trends are not quite as obvious.[4]

Forecast dispersion is sometimes thought to signify herding. With this interpretation, low dispersion would be undesirable as it suggests greater herding. However, in this study, low dispersion is considered a desirable property. At least two reasons suggest this is true: (1) firms with losses or earnings declines, potential candidates to hide bad information, tend to have highly dispersed forecasts in previous studies (Ciccone, 2001), and (2) the high positive correlation between dispersion and error.[5]

An important component of this research is the separation of firms with losses and profits. A loss is defined as when the actual earnings per First Call are less than zero. A profit is defined as when actual earnings are greater than or equal to zero. First Call earnings, frequently referred to as ''Street'' or ''operating'' earnings (among other names), are often different from earnings under generally accepted accounting principles or GAAP (Abarbanell & Lehavy, 2000; Bradshaw & Sloan, 2002). The results are similar if GAAP earnings are used to determine profitability. The Compustat database is used to obtain GAAP earnings.

To alleviate problems with small denominators, a firm with a divisor less than US\$0.02 in absolute value terms has the problem divisor set to US\$0.02. Two procedures are used to reduce the influence of large observations. Firms with dispersion or error numbers greater than 10 and firms with earnings per share greater than an absolute value of US\$20 are eliminated from their respective sample. Combined, the two procedures eliminate a total of 220 quarterly observations with no effect on the conclusions.

The final sample includes the years 1990 through 2001, a 12-year or 48-quarter period.[6] The total sample includes 120,022 firm quarters: 94,194 with profits and 25,828 (21.5%) with losses. The number of observations varies by the forecast property being examined.

[3] The raw error, often called the ''earnings surprise'' (although usually with the sign or direction of the error), is important because this number is often reported by the news media. It is important to note that ''error'' and ''raw error'' have two distinct meanings in this study.

[4] Using price as a deflator, average profit firm dispersion decreases from 0.0027 in the early (1990–1995) sample period to 0.0015 in the later sample period (1996–2001). Loss firm dispersion decreases from 0.0128 to 0.0069. Profit firm error decreases from 0.0052 to 0.0041, while loss firm error decreases from 0.0409 to 0.0333. All differences are significant with 99% confidence.

[5] To illustrate the latter point, the correlation between the dispersion and error is computed as 0.22 (0.24 if a log transform is performed). In a related test, every quarter each firm is placed into 1 of 10 portfolios based on its ranking of dispersion and 1 of 10 portfolios based on its ranking of error. The correlation between the group placement (1–10) is then computed. The correlation between the dispersion and error groupings is .47.

[6] The year 1990 contains considerably less sample firms than the other 11 years. Caution is thus recommended when evaluating the 1990 data.

The dispersion measure has the fewest number of observations: 84,919 quarterly observations.

Portfolio analyses are used to communicate the results in an easily accessible manner. The included tables present the results year-by-year and also during two sample periods: an ‘‘early’’ sample period from 1990 through 1995 and a ‘‘later’’ sample period from 1996 through 2001. Each period contains half the sample years. In addition, regression models controlling for size and book-to-market ratio are used to support the major conclusions reached.

4. Forecasting trends

Table 1 presents, by year, the forecast properties and maximum number of observations (recall there are sample size differences among the various properties). Dispersion, error, raw error, and optimism all steadily decrease throughout the sample period. The trend for optimism is interesting as the forecasts changed from being optimistic more than 50% of the time in the first couple of sample years to being optimistic less than 50% of the time after 1992. The amount of optimism continues to decrease during the sample period, reaching a low of 34.27% in 2000.

Table 1
Forecast dispersion, error, and optimism

	Quarterly forecasts				
	Maximum number of observations	Dispersion	Error	Raw error	Percent optimistic
All years	120,022	0.22	0.44	0.09	40.27
1990–1995	40,949	0.27	0.48	0.11	45.90
1996–2001	79,073	0.20	0.42	0.09	37.36
Difference		0.07*	0.06*	0.02*	8.54*
1990	1373	0.31	0.58	0.16	57.70
1991	2929	0.38	0.59	0.15	53.77
1992	6497	0.30	0.46	0.11	46.36
1993	8411	0.26	0.46	0.12	46.64
1994	10,249	0.25	0.46	0.10	43.33
1995	11,490	0.24	0.47	0.09	43.88
1996	14,002	0.23	0.44	0.09	39.27
1997	14,942	0.19	0.41	0.08	38.86
1998	15,184	0.20	0.41	0.08	38.71
1999	13,638	0.20	0.43	0.09	34.95
2000	12,314	0.17	0.42	0.10	34.27
2001	8993	0.21	0.42	0.09	37.46

This table reports mean analyst quarterly forecast properties over the sample period 1990 through 2001. Dispersion is defined as the standard deviation of the quarterly forecasts divided by the absolute mean forecast. Raw error is defined as the absolute value of the actual earnings less the forecasted earnings. Error is defined as the absolute value of the actual earnings less the forecasted earnings, divided by the absolute actual earnings. A firm’s forecast is considered optimistic if the mean forecast is greater than the corresponding actual earnings. As the sample size varies by the forecast property in question, the maximum number of observations is reported.

* Difference is significantly different from zero with 99% confidence.

Table 2 shows the same forecast properties after separating firms by profitability. The dispersion and error of loss firms is considerably greater than the dispersion and error of profit firms. This occurs in every sample year and, although not tabulated, in every sample quarter. However, loss firms show greater reductions in dispersion and error throughout the sample period. The average dispersion of loss firms decreases from a high of 1.12 in 1990 to 0.30 in 2000 and 0.33 in 2001. Thus, the typical forecast dispersion of a loss firm today is roughly a quarter of what it was just 10 years ago. The story is similar for forecast error. The mean forecast error of loss firms decreases from a high of 1.16 in 1990 to 0.63 in 2000 and 0.55 in 2001. The error reduction for profit firms is not nearly as large, decreasing from a high of 0.48 in 1991 to 0.33 in 2000 and 0.35 in 2001.

The first two charts in Fig. 1 show the forecast dispersion and error by year and profitability. The figure provides a nice illustration of the eroding dichotomous forecasting ability of analysts. Clearly, analysts are narrowing the gap in their performance between profit and loss firms.

Table 2 also presents statistics for the mean raw error. Similar to the previous results, improvement in the raw error numbers occurs regardless of profitability, but the improvement is especially large for loss firms. For example, the raw error of loss firms decreases by more than half, from an average of US$0.48 in 1991 to US$0.21 in 2000 and US$0.16 in 2001.

The last columns of Table 2 show the percentage of optimistic forecasts. In the early sample period, analysts are overwhelmingly optimistic toward loss firms, more than 75% of time. The optimism remains above 70% until 1997 when it drops to 67.66%. From

Table 2
Forecast dispersion, error, raw error, and optimism by profitability

	Dispersion		Error		Raw error		Percent optimistic (negative surprise)	
	Profit	Loss	Profit	Loss	Profit	Loss	Profit	Loss
All quarters	0.15	0.53	0.35	0.78	0.06	0.23	33.63	64.48
1990–1995	0.18	0.88	0.37	1.02	0.07	0.33	40.32	75.93
1996–2001	0.13	0.43	0.33	0.70	0.05	0.20	29.76	60.70
Difference	0.05*	0.45*	0.04*	0.32*	0.02*	0.13*	10.56*	15.23*
1990	0.19	1.12	0.47	1.16	0.10	0.49	52.97	85.42
1991	0.24	1.11	0.48	1.09	0.08	0.48	48.40	78.44
1992	0.21	0.94	0.37	0.95	0.07	0.34	40.91	76.43
1993	0.17	0.91	0.37	0.96	0.08	0.34	41.67	74.80
1994	0.17	0.80	0.36	0.99	0.06	0.30	37.82	73.54
1995	0.16	0.81	0.35	1.11	0.06	0.28	37.54	76.75
1996	0.15	0.70	0.34	0.86	0.05	0.26	32.06	70.90
1997	0.12	0.50	0.32	0.78	0.05	0.22	31.58	67.66
1998	0.13	0.47	0.32	0.71	0.04	0.19	30.68	65.21
1999	0.14	0.39	0.33	0.70	0.05	0.20	26.84	58.42
2000	0.13	0.30	0.33	0.63	0.05	0.21	26.63	51.97
2001	0.15	0.33	0.35	0.55	0.05	0.16	29.44	53.12

This table reports mean analyst quarterly forecast properties sorted by profitability over the sample period 1990 through 2001. A profit occurs when actual quarterly earnings are greater than or equal to zero. A loss occurs when actual quarterly earnings are less than zero. See Table 1 for variable definitions.

* Difference is significantly different from zero with 99% confidence.

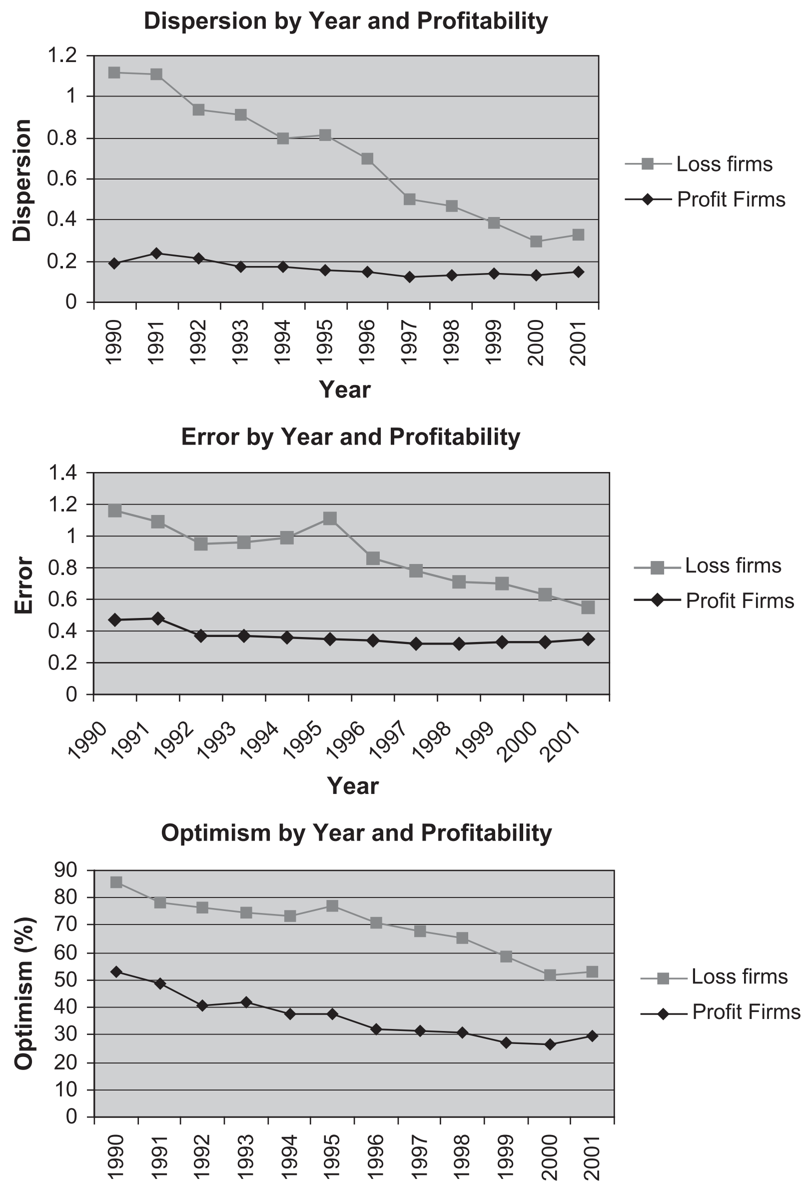


Fig. 1. Forecast properties by year and profitability.

there, the optimism continues to decrease, dropping to an almost unbiased 51.97% in 2000 and 53.12% in the 2001. For profit firms, optimism on average vanishes in 1991 and continues to decrease steadily throughout the sample period. By the end of the sample period, optimism is under 30%. The last chart in Fig. 1 illustrates this trend of decreasing optimism for both profit and loss firms.

Although the testing focuses on realized actual earnings to determine profitability, the results from Table 2 are repeated using expected earnings to determine profitability. Firms are resorted into profit and loss portfolios based on the mean forecast at fiscal year end. These results (not tabulated) are qualitatively similar to the Table 2 results, although average dispersion, error, and optimism are higher for expected profit firms (versus actual profit firms) and lower for expected loss firms. Optimism actually drops below 50% for expected loss firms during the last three sample years: 1999, 2000, and 2001. Related testing is performed on Table 6.

Regression models are utilized next to control for variables aside from profitability that influence forecasts. Previous studies have shown that size and growth prospects (growth indicated by book-to-market ratio) affect the information environment (e.g., Atiase, 1985; Ciccone, 2001).[7]

To test, two sets of regression models are used. The first set of regressions is employed to confirm the trend of lower dispersion and error during the sample period. These models use dispersion and error as the dependent variables and size, book-to-market ratio, a loss dummy variable, and year dummy variables as the independent variables. The Compustat database is used to gather the size and book-to-market ratio data. Size is defined as price times shares, computed at the beginning of the fiscal year. Book-to-market ratio is defined as beginning of fiscal year equity (Compustat item A216) divided by size. Logarithms of size and book-to-market ratio are used in the regressions. The loss dummy variable equals one if the actual First Call earnings are negative and zero otherwise. The year dummy variables equal one if the forecast is from the corresponding year and zero otherwise. The first year dummy variable corresponds to 1991, leaving 1990 as the base year. This specification is as follows for firm i during year t, quarter q.

$$\begin{aligned}\text{Forecast property}_{i,t,q} = {} & a + b_1 \ \log(\text{size})_{i,t} + b_2 \ \log(\text{b/m})_{i,t} \\ & + b_3 \ \text{loss dummy}_{i,t,q} + b_4 \ \text{year 1991 dummy}_{i,t} + \ldots \\ & + b_{14} \ \text{year 2001 dummy}_{i,t} + e_{i,t,q} \qquad (1)\end{aligned}$$

Table 3 presents the results of these regressions. Although size, book-to-market ratio, and especially losses affect the forecasts, the significant, negative values on the year dummy variables tend to increase in magnitude over the sample period. For example, using error as the dependent variable, the coefficient of the 1992 year dummy is -0.11 (indicating an average decrease of -0.11 relative to the 1990 base year), while that of the 2001 year dummy is -0.23 (indicating an average decrease of -0.23 relative to the 1990 base year). These results confirm the trends revealed in the portfolio results.

In the second set of regressions, models are employed annually from 1990 through 2001 to confirm the erosion of differences between profit and loss firm forecasts.

[7] The size of the analyst following is also included in separate regressions with no effect on the conclusions. Analyst following is not included in the presented results because of its strong correlation to size, thus blurring the relation between size and the forecast properties.

Table 3
Regression results using year dummy variables

	Dispersion		Error	
	Coefficient	*t* Value	Coefficient	*t* Value
Intercept	0.24	9.21	1.09	30.61
log (size)	0.01	2.17	– 0.04	– 22.61
log (book/market)	0.06	21.55	0.06	15.95
Loss dummy	0.42	82.48	0.43	61.21
1991	0.07	2.78	– 0.02	– 0.60
1992	0.00	0.21	– 0.11	– 3.71
1993	– 0.03	– 1.21	– 0.13	– 4.42
1994	– 0.04	– 1.99	– 0.13	– 4.47
1995	– 0.05	– 2.33	– 0.12	– 4.33
1996	– 0.05	– 2.45	– 0.15	– 5.34
1997	– 0.11	– 5.40	– 0.19	– 6.86
1998	– 0.11	– 5.44	– 0.19	– 6.82
1999	– 0.13	– 6.23	– 0.19	– 6.67
2000	– 0.15	– 7.61	– 0.20	– 7.31
2001	– 0.17	– 8.27	– 0.23	– 8.29
N	75,337		105,287	

This table reports the results of a regression model. Either forecast dispersion or error is the dependent variable. The independent variables are the logarithm of size (price times shares) in thousands, the logarithm of book-to-market value (equity/size), a loss dummy equal to one if the actual quarterly First Call earnings are below zero and equal to zero otherwise, and year dummy variables spanning 1991 through 2001 equal to one if the quarterly forecast is from the corresponding year. The regression model is below:

$$\text{Forecast property}_{i,t} = a + b_1 \log(\text{size})_{i,t} + b_2 \log(\text{b/m})_{i,t} + b_3 \text{ loss dummy}_{i,t} + b_4 \text{ year 1991 dummy}_{i,t} + \ldots + b_{14} \text{ year 2001 dummy}_{i,t} + e_{i,t}$$

Dispersion and error are the dependent variables, while size, book-to-market ratio, and a loss dummy variable are the independent variables. The annual model appears below:

$$\text{Forecast property}_{i,q} = a + b_1 \log(\text{size})_i + b_2 \log(\text{b/m})_i + b_3 \text{ loss dummy}_{i,q} + e_{i,q} \tag{2}$$

The results of these regressions appear on Table 4. Once again, the portfolio results are confirmed. For example, using dispersion as the dependent variable, the coefficient on the loss dummy variable decreases sharply over the sample period, dropping from 0.83 and 0.86 in 1990 and 1991, respectively, to 0.20 in 2001.

Table 5 shows the percentage of analysts forecasting the wrong sign. In the early sample period using the annual earnings, analysts forecast profits for firms with actual losses 33.95% of the time. This number is far greater than the reverse. In the early sample period, analysts forecast losses for firms with actual profits just a little over 1% of the time. Although over the sample period, there is no improvement in predicting profits for actual profit firms (profit prediction actually gets worse), the improvement for loss firms is rather extraordinary. At the end of the sample period, profits are forecasted for loss firms only 14.24% of the time in 2000 and 12.20% of the time in 2001, consistent with the increasing tendency of firms to warn of losses.

Table 4
Annual regression results using loss dummy variables

Year	Dispersion									
	Coefficient				t Value				F value	R^2 (adjusted)
	Intercept	Size	B/M	Loss dummy	Intercept	Size	B/M	Loss dummy		
1990	− 0.14	0.03	0.12	0.83	− 0.76	2.22	3.41	12.94	65.43	0.21
1991	0.14	0.01	0.12	0.86	0.88	1.11	4.97	17.19	115.18	0.18
1992	0.10	0.01	0.11	0.73	1.80	0.96	6.86	22.20	189.14	0.14
1993	0.20	0.00	0.06	0.73	2.61	0.10	4.29	27.04	258.12	0.14
1994	0.20	0.00	0.07	0.63	2.93	0.31	6.51	27.26	268.99	0.12
1995	0.15	0.00	0.04	0.66	2.39	0.65	4.10	31.80	354.31	0.13
1996	0.37	− 0.01	0.04	0.62	6.81	− 3.34	5.02	35.40	455.72	0.14
1997	0.25	− 0.01	0.04	0.38	5.85	− 2.05	5.95	29.54	324.43	0.09
1998	0.13	0.00	0.05	0.34	3.08	1.08	6.67	28.82	299.31	0.08
1999	0.08	0.01	0.06	0.29	1.73	2.43	10.13	23.20	218.10	0.07
2000	0.16	− 0.00	0.04	0.22	3.66	− 0.09	7.17	18.48	126.99	0.05
2001	− 0.08	0.02	0.04	0.20	− 1.77	5.29	6.51	16.95	103.18	0.05

Year	Error									
	Coefficient				t Value				F value	R^2 (adjusted)
	Intercept	Size	B/M	Loss dummy	Intercept	Size	B/M	Loss dummy		
1990	0.77	− 0.02	0.09	0.51	3.09	− 0.88	1.93	5.80	14.98	0.04
1991	1.16	− 0.05	0.09	0.50	6.97	− 3.71	3.12	8.96	45.28	0.05
1992	0.81	− 0.03	0.07	0.60	7.77	− 3.71	4.01	17.03	118.41	0.06
1993	1.02	− 0.05	0.09	0.54	10.88	− 6.21	5.40	17.58	146.80	0.06
1994	1.18	− 0.06	0.07	0.58	13.82	− 8.91	4.86	21.00	213.69	0.07
1995	1.06	− 0.05	0.04	0.68	12.83	− 8.18	2.41	25.27	285.53	0.08
1996	1.13	− 0.06	0.04	0.54	16.23	− 10.77	3.72	24.18	287.19	0.07
1997	0.95	− 0.05	0.03	0.41	14.56	− 9.22	3.10	21.17	228.30	0.05
1998	0.86	− 0.04	0.08	0.35	13.78	− 7.35	7.46	19.78	214.93	0.05
1999	0.78	− 0.03	0.07	0.37	11.79	− 5.87	6.69	19.09	192.21	0.05
2000	0.76	− 0.03	0.06	0.35	11.29	− 5.70	7.11	18.84	168.52	0.04
2001	0.70	− 0.02	0.06	0.19	8.91	− 3.94	4.90	9.36	58.84	0.02

This table reports the results of an annual regression model, run every sample year from 1990 through 2001. Either forecast dispersion or error is the dependent variable. The independent variables are the logarithm of size (price times shares) in thousands, the logarithm of book-to-market value (equity/size), and a loss dummy equal to one if the actual quarterly First Call earnings are negative and zero otherwise. The regression model is below:

$$\text{Forecast property}_i = a + b_1 \log(\text{size})_i + b_2 \log(\text{b/m})_i + b_3 \text{ loss dummy}_i + e_i$$

To directly examine forecast performance when actual profitability differs from forecasted profitability, firms are separated into four portfolios based on actual versus expected profits or losses. For example, one portfolio includes firms with expected profits that report actual losses, while another includes firms with expected losses reporting actual losses. Mean dispersion and error are computed for each of the four portfolios. The results are presented in Table 6.

In an unsurprising result, firms with expected and actual profits have the lowest dispersion and error. Interestingly, however, firms with expected and actual losses have the

Table 5
Percentage of firms with wrong sign mean forecasts

	Quarterly forecasts	
	Forecasted loss, actual profit (%)	Forecasted profit, actual loss (%)
All years	1.79	23.31
1990–1995	1.22	33.95
1996–2001	2.11	19.80
Difference	−0.89*	14.15*
1990	0.89	44.79
1991	1.58	35.11
1992	1.38	30.79
1993	1.04	31.85
1994	1.18	32.15
1995	1.27	37.08
1996	1.72	29.57
1997	1.73	24.28
1998	1.86	21.42
1999	2.52	19.59
2000	2.49	14.24
2001	2.89	12.20

This table reports the percentage of analysts forecasting the wrong sign (e.g., forecasting a profit when an actual loss is eventually reported) over the sample period 1990 through 2001. All numbers are in percent.

* Difference is significantly different from zero with 99% confidence.

second lowest dispersion and error, while the two portfolios containing firms with actual profitability different from expected profitability have the highest dispersion and error. In addition, although error does decrease in the portfolio of expected loss, actual loss firms throughout the sample period, the trend is not nearly as clear and the differences not nearly as large compared with the Table 2 results. These results, combined with the results from Table 5, suggest that a large portion of the decrease in loss firm error comes from two sources: (1) improvement in the error of expected profit, actual loss firms and (2) the higher percentage of losses being predicted (i.e., less expected profit, actual loss firms).

The final testing in this section examines the error and optimism of the mean analyst forecast versus the error and optimism of a "naïve" forecast, the actual First Call earnings in the prior fiscal period.[8] This test addresses several important issues. It provides a measure of the amount of value that analysts provide over and above a forecasting method simple enough to be employed by even the most unsophisticated of individual investors. The test also provides a standard by which to measure earnings predictive difficulty. Firms with accurate naïve forecasts can be thought of as having earnings that are relatively easy to predict. Related to prediction difficulty, the test also somewhat controls for earnings

[8] For the tabulated quarterly results, the naïve model compares the current quarter earnings with the prior quarter earnings (e.g., third quarter 1992 compared with second quarter 1992). To control for earnings seasonality, the prior year quarterly earnings are also used to compute naïve forecasts (e.g., second quarter 1993 compared with second quarter 1992). However, because these naïve forecasts are less accurate than the naïve forecasts using the prior quarter earnings, the results are presented using the more accurate prior quarter naïve forecasts. (Using all sample firms, the average naïve error is 0.82 using prior year quarterly earnings and 0.72 using prior quarter earnings.) The results using the prior year naïve forecasts are similar although analyst superiority is greater.

Table 6
Dispersion and error by expected and actual profitability

Expected	Quarterly forecasts							
	Dispersion				Error			
	Profit	Profit	Loss	Loss	Profit	Profit	Loss	Loss
Actual	Profit	Loss	Profit	Loss	Profit	Loss	Profit	Loss
All years	0.13	0.93	1.07	0.42	0.31	1.97	2.38	0.42
1990–1995	0.16	1.17	1.37	0.74	0.35	2.06	2.59	0.50
1996–2001	0.12	0.82	0.98	0.35	0.29	1.91	2.31	0.40
Difference	0.04*	0.35*	0.39*	0.39*	0.06*	0.15*	0.28*	0.10*
1990	0.19	1.31	0.67	0.98	0.47	2.01	2.09	0.49
1991	0.23	1.30	0.99	1.01	0.44	1.97	2.90	0.62
1992	0.19	1.38	2.00	0.76	0.34	2.06	2.76	0.46
1993	0.16	1.24	1.33	0.76	0.35	2.03	2.44	0.46
1994	0.15	1.08	1.30	0.68	0.33	2.07	2.57	0.49
1995	0.14	1.04	1.26	0.69	0.32	2.12	2.55	0.51
1996	0.13	1.04	1.22	0.57	0.30	1.89	2.25	0.43
1997	0.11	0.84	1.00	0.40	0.28	1.94	2.42	0.41
1998	0.11	0.75	1.08	0.40	0.28	1.88	2.11	0.39
1999	0.12	0.73	0.94	0.32	0.28	1.90	2.38	0.41
2000	0.11	0.68	0.84	0.24	0.28	1.98	2.18	0.41
2001	0.13	0.77	0.77	0.27	0.29	1.93	2.54	0.37

This table reports mean analyst quarterly forecast properties sorted by expected and actual profitability over the sample period 1990 through 2001. An actual profit occurs when actual quarterly earnings are greater than or equal to zero, while an actual loss occurs otherwise. A forecasted profit occurs when mean forecasted earnings are greater than or equal to zero, while a forecasted loss occurs otherwise. See Table 1 for variable definitions.

* Difference is significantly different from zero with 99% confidence.

volatility or earnings management (see also next section). Firms with managed or less volatile earnings would probably have more accurate naïve forecasts.

Error, raw error, and optimism are computed using both the analyst forecasts and the naïve forecasts for all sample firms having the required prior period actual earnings information. The sample size is 103,778 firm-quarter observations: 82,203 with profits and 21,575 (20.8%) with losses.

Table 7 reports the results for two forecast properties: error and raw error. For each sample firm, the analyst forecast error is subtracted from the naïve forecast error. For example, if the naïve forecast error is 0.90 and the analyst forecast error is 0.40, then the difference is 0.50. The mean of these differences is computed and reported in the table. Note that in the table, positive numbers indicate analyst superiority, and the larger the difference, the more accurate analyst forecasts are versus naïve forecasts.

Several findings are important. Analyst forecasts are considerably more accurate in every sample year indicating that analysts provide a great deal of value in forecasting earnings versus a simple naïve model. However, they provide more value when forecasting the earnings of loss firms. For example, for all years, the difference between the naïve and analyst error is on average 0.26 for profit firms and 0.45 for loss firms.

Analysts have also slightly increased the value of their forecasting during the sample period, particularly for loss firms. For example, in the early sample period, the analysts are

Table 7
Differences between naïve and analyst forecasts: error and raw error

	Quarterly forecasts					
	Error differences (naïve error – analyst error)			Raw error (RE) differences (naïve RE – analyst RE)		
	All	Profit	Loss	All	Profit	Loss
All years	0.30	0.26	0.45	0.08	0.07	0.08
1990–1995	0.26	0.24	0.39	0.07	0.07	0.07
1996–2001	0.32	0.27	0.47	0.08	0.08	0.08
Difference	– 0.06*	– 0.03*	– 0.08*	– 0.01*	– 0.01*	– 0.01
1990	0.27	0.23	0.48	0.07	0.05	0.18
1991	0.19	0.17	0.32	0.08	0.08	0.11
1992	0.29	0.26	0.45	0.08	0.08	0.06
1993	0.26	0.24	0.38	0.05	0.05	0.06
1994	0.27	0.25	0.35	0.07	0.07	0.06
1995	0.26	0.24	0.40	0.08	0.08	0.08
1996	0.32	0.28	0.55	0.08	0.08	0.07
1997	0.30	0.27	0.46	0.08	0.08	0.07
1998	0.36	0.29	0.59	0.09	0.09	0.10
1999	0.33	0.30	0.44	0.09	0.09	0.08
2000	0.31	0.29	0.39	0.08	0.09	0.07
2001	0.25	0.17	0.38	0.08	0.08	0.08

This table reports the difference between naïve forecast errors and analyst forecast errors over the sample period 1990 through 2001. Analyst forecast error and raw error are defined as in Table 1. Naïve forecast raw error is defined as the absolute value of actual quarterly earnings less the previous quarter's actual earnings. Naïve forecast error deflates this number by the absolute actual quarterly earnings. The reported differences are computed as the naïve error less the analyst error. Thus, positive differences indicate analyst superiority (i.e., lower errors): the higher the difference, the greater the analyst superiority.

*Difference is significantly different from zero with 99% confidence.

superior by 0.39 in predicting error. In the later sample period, this superiority increases to 0.47.

Although not tabulated, naïve forecasts for loss firms are markedly less accurate versus naïve forecasts for profit firms. The mean quarterly naïve forecast error is 0.60 for profit firms and 1.22 for loss firms. The differences remain fairly stable across the sample period. This suggests that loss firm earnings are much more difficult to predict. Thus, considering both the inherent difficulties and the trends of reduced error, analysts seem to be doing an adequate job when forecasting loss firm earnings.

Table 8 presents the results for differences in optimism. With respect to the percentage of optimism, it is assumed that the goal when forecasting is to achieve a systematically unbiased 50%. Therefore, the comparison of analyst forecast optimism versus naïve forecast optimism is computed using 50% as a reference. For example, if analysts are optimistic 45% of the time and naïve forecasts are optimistic 65% of the time, then analyst forecasts are superior by 10% with respect to the 50% goal [(65% – 50%) – (50% – 45%) = 10%]. A positive sign indicates better analyst performance; a negative sign indicates better naïve performance.

The results are fascinating. Naïve forecasts for loss firms are primarily optimistic (63.75%) while naïve forecasts for profit firms are primarily pessimistic (35.58%). Thus,

Table 8
Differences between naïve and analyst forecasts: optimism

	Quarterly forecasts					
	Profit			Loss		
	Percent optimistic, analysts	Percent optimistic, naïve	Analyst superiority versus unbiased 50%	Percent optimistic, analysts	Percent optimistic, naïve	Analyst superiority versus unbiased 50%
All years	33.42	35.58	− 2.16	64.43	63.75	− 0.68
1990–1995	40.29	35.63	4.66	76.70	68.10	− 8.60
1996–2001	29.78	35.56	− 5.78	60.69	62.43	1.74
Difference	10.51*	0.07	− 10.44	16.01*	5.67*	10.34
1990	53.13	35.78	11.09	84.07	69.91	− 14.16
1991	51.88	37.62	10.50	78.77	68.49	− 10.28
1992	41.32	35.84	5.48	77.97	65.85	− 12.12
1993	41.90	36.01	5.89	75.00	66.67	− 8.33
1994	37.95	35.23	2.72	74.69	68.19	− 6.50
1995	37.75	35.29	2.46	77.92	70.13	− 7.79
1996	32.50	33.78	− 1.28	72.67	69.16	− 3.51
1997	31.95	33.86	− 1.91	67.54	64.96	− 2.58
1998	30.53	37.15	− 6.62	64.97	65.22	0.25
1999	26.86	35.30	− 8.44	58.83	60.38	1.55
2000	26.18	34.90	− 8.72	52.21	60.58	8.37
2001	29.11	40.99	− 11.88	51.36	55.75	4.39

This table reports the difference between naïve forecast optimism and analyst forecast optimism over the sample period 1990 through 2001. Optimism is present if the mean forecast is greater than the corresponding actual earnings. As 50% is considered the unbiased target, analyst superiority is determined using 50% as the benchmark. Positive numbers in the "analyst superiority versus unbiased 50%" column indicate analyst superiority, while negative numbers indicate naïve forecast superiority. The analyst superiority column is computed as follows:

$$\text{Analyst superiority} = (|\%\text{ optimistic naïve} - 50\%|) - (|\%\text{ optimistic analysts} - 50\%|)$$

*Difference is significantly different from zero with 99% confidence.

the optimism analysts show toward loss firms and the pessimism analysts show toward profit firms is perhaps a natural reflection of an easy starting point. For profit firms, in the early sample period, analysts are nearly unbiased. However, as analyst pessimism increases during the sample period for profit firms, analyst superiority with regard to systematic biases steadily changes to inferiority. As an example, analysts are superior relative to the 50% reference for profit firms by 11.09% in 1990 and 10.50% in 1991. However, these numbers decrease to − 8.72% in 2000 and − 11.88% in 2001, indicating a decline in analyst performance. In contrast, for loss firms, analysts move steadily from inferior performance to superior performance. Fig. 2 shows the trends graphically. Like the corresponding table, positive numbers in the figure indicate superior analyst performance.

5. Earnings management, smoothing, and guidance issues

The increase in forecast pessimism (positive surprises) and decrease in forecast error seen in this and other studies is consistent with earnings management, guidance, and

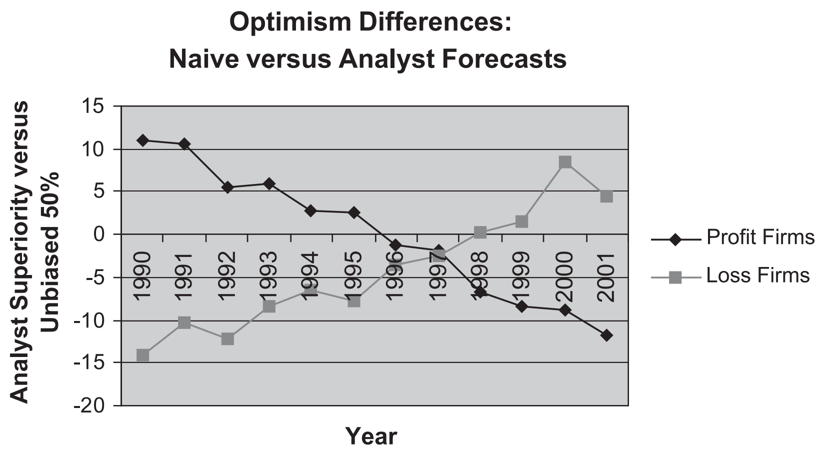


Fig. 2. Analyst versus naïve forecast differences in optimism by year. Note: positive numbers indicate analyst superiority; negative numbers indicate naïve superiority.

smoothing. Various tests are performed to see whether the trends are related to these issues and to differentiate among the potential explanations.

The first procedure examines the subset of firms that failed to meet all three incentives mentioned by Degeorge et al. (1999) when managing earnings: incentives of avoiding losses, avoiding earnings declines, and meeting analyst expectations. Thus, these firms are considered unlikely to be managing earnings as none of the incentives is reached.

Table 9 reports the results. Although the average dispersion, error, and raw error are all higher for this sample of firms versus the full loss firm subsample, similar degrees of improvement in each property are seen. As an example, the average error of these firms drops from 1.23 in the early sample period to 0.93 in the later sample period. This compares with the results for loss firms with either type of surprise from Table 2: 1.02 in the early sample period, decreasing to 0.70 in the later sample period.

To investigate smoothing, trends in earnings volatility are examined. If the decrease in forecasting performance is attributable to increased smoothing, earnings volatility should decrease as well. Earnings volatility is computed as the standard deviation of earnings from the eight most recent quarters. The sample of firms with eight quarters of earnings begins in 1992 and consists of 51,965 firms: 42,543 with profits and 9422 (18.1%) with losses. The trends in earnings volatility are reported in Table 10. Although loss firm earnings volatility decreases, profit firm volatility remains fairly stable across the sample period. Thus, earnings smoothing does not explain trends in profit firm forecasts. For loss firms, the magnitude of the decrease in earnings volatility is far less than the magnitude of the decrease in error and dispersion. Therefore, earnings volatility probably does not explain a large proportion of the trends in loss firm forecasts.

Related testing looks at forecasting trends in a set of firms considered unlikely candidates to smooth earnings, those firms with high earnings volatility. Thus, in each sample year, firms with high earnings volatility are separately analyzed. Both absolute and relative measures of high volatility are used. Absolute measures specify an arbitrary

Table 9
Forecast dispersion, error, and raw error: firms with optimistic forecasts (negative surprises), earnings declines, and losses

	Quarterly forecasts		
	Dispersion	Error	Raw error
All years	0.71	1.01	0.36
1990–1995	1.00	1.23	0.46
1996–2001	0.61	0.93	0.33
Difference	0.39*	0.30*	0.13*
1990	0.87	1.28	0.52
1991	1.20	1.27	0.65
1992	1.12	1.19	0.46
1993	1.03	1.14	0.52
1994	0.94	1.21	0.44
1995	0.93	1.31	0.39
1996	0.87	1.08	0.38
1997	0.66	0.99	0.34
1998	0.63	0.95	0.29
1999	0.54	0.94	0.33
2000	0.47	0.85	0.35
2001	0.50	0.74	0.25

This table reports mean analyst quarterly forecast properties for firms with optimistic forecasts, earnings declines, and losses over the sample period 1990 through 2001. An earnings decline is when actual quarterly earnings are less than the previous quarter's actual earnings. See Table 1 for the other variable definitions.

* Difference is significantly different from zero with 99% confidence.

earnings volatility number to which each firm's earnings volatility is compared, thus controlling for any changes in average volatility during the sample period. Quarterly earnings volatility is considered high if the standard deviation of the actual Street earnings is greater than US\$0.50 per share over the prior eight quarters.[9] Under the relative measures of volatility, a firm is considered to have high earnings volatility if its volatility is in the top 10% during the year. Although the results are not tabulated, the same trends of decreasing dispersion, error, and optimism throughout the sample period still exist for the high earnings volatility sample of firms using either the absolute or relative volatility measures.

The next test investigates earnings guidance by isolating firms with high dispersion. These firms are often considered to have a greater disparity of opinion (e.g., Krishnaswami & Subramaniam, 1999) and are, therefore, unlikely to be guiding analysts toward a specific earnings target.

Similar to the volatility tests, absolute and relative measures are used. Under the absolute method, firms are considered to have high dispersion if their dispersion measure is greater than or equal to 0.50.[10] This sample contains 8225 firms (9.7% of the full dispersion sample), 4028 with profits and 4197 (51.0%) with losses. Under the relative measure, firms are considered to have high dispersion if their dispersion measure is in the top 10% during the relevant year.

[9] Other arbitrary cutoff points are employed with similar results.

[10] Other arbitrary cutoff points are employed with similar results.

Table 10
Earnings volatility by year

	Eight quarter earnings volatility		
	All	Profit	Loss
All years	0.17	0.14	0.28
1992–1996	0.17	0.14	0.36
1997–2001	0.16	0.14	0.25
Difference	0.01*	0.00	0.11*
1992	0.18	0.16	0.32
1993	0.18	0.15	0.35
1994	0.18	0.16	0.35
1995	0.18	0.14	0.43
1996	0.16	0.13	0.33
1997	0.16	0.14	0.29
1998	0.15	0.13	0.23
1999	0.16	0.14	0.24
2000	0.16	0.14	0.26
2001	0.18	0.15	0.26

This table reports mean quarterly earnings volatility over the sample period 1992 through 2001. Quarterly earnings volatility is defined as the standard deviation of actual earnings from the eight previous quarters. As 2 years of earnings are needed before the volatility can be computed, the sample period does not include 1990 and 1991.

* Difference is significantly different from zero with 99% confidence.

Table 11
Forecast error, raw error, and optimism by profitability: firms with dispersion greater than 0.50

	Quarterly forecasts								
	Error			Raw error			Percent optimistic		
	All	Profit	Loss	All	Profit	Loss	All	Profit	Loss
All years	1.09	1.14	1.04	0.23	0.13	0.33	64.61	39.95	88.28
1990–1995	1.21	1.24	1.17	0.30	0.19	0.42	69.24	49.36	90.93
1996–2001	1.01	1.07	0.96	0.19	0.08	0.28	61.76	33.51	86.81
Difference	0.20*	0.17*	0.21*	0.11*	0.11*	0.14*	7.48*	15.85*	4.12*
1990	1.35	1.60	1.09	0.55	0.37	0.74	73.85	58.82	90.32
1991	1.15	1.18	1.13	0.38	0.17	0.60	68.05	48.77	88.74
1992	1.11	1.13	1.09	0.32	0.21	0.45	66.73	47.71	90.00
1993	1.20	1.27	1.12	0.26	0.19	0.34	69.06	49.37	91.43
1994	1.23	1.21	1.25	0.30	0.21	0.40	67.97	48.56	90.12
1995	1.26	1.30	1.22	0.24	0.12	0.35	71.90	50.00	92.65
1996	1.12	1.13	1.11	0.24	0.11	0.38	66.83	41.83	91.40
1997	1.01	1.06	0.97	0.20	0.08	0.31	63.19	36.77	87.94
1998	0.97	1.03	0.93	0.17	0.07	0.26	64.15	35.50	86.82
1999	0.98	1.08	0.90	0.18	0.08	0.27	56.75	25.67	85.02
2000	1.02	1.09	0.96	0.16	0.08	0.22	56.10	29.21	80.94
2001	0.90	0.95	0.87	0.16	0.08	0.22	60.13	25.95	86.47

This table reports mean analyst quarterly forecast properties for firms with forecast dispersion greater than 0.50 over the sample period 1990 through 2001. See Table 1 for variable definitions.

* Difference is significantly different from zero with 99% confidence.

Table 11 presents the results using the absolute measure. (The results using the relative measure are similar.) There is a clear reduction in forecast error and raw error during the sample period for both profit and loss firms. Optimism also decreases dramatically for profit firms, starting around 50% in the first few sample years, but reaching below 30% for the last three sample years. Loss firms, however, are dominated by overwhelming optimism throughout the sample period (an average of 88.28%), the lack of improvement indicating a problem area that analysts should address. Thus, although analysts have reduced the size of their errors for firms with high dispersion, they still tend to overestimate the earnings of high dispersion, loss firms. This testing suggests that systematic profit firm pessimism occurs regardless of whether the forecasts are guided. However, the reduction of loss firm optimism occurs when firms warn analysts of the impending loss.

Overall, the improved forecasting ability of analysts occurs regardless of increases in earnings management, guidance, or smoothing. The trends are consistent with concerns of legal liability as most of the reduction in dispersion and error is due to loss firms. The trends are also consistent with improved analyst forecasting abilities. The increase in pessimism for profit firms may be partly attributed to an overreliance on the previous period's earnings.

6. GAAP versus Street earnings and Regulation FD

Another issue is related to the Street versus GAAP earnings debate. Abarbanell and Lehavy (2000) suggest that using forecast provider databases, such as First Call, to obtain earnings data might impact conclusions reached in earnings-related studies. First Call collects data based on the earnings that firms publicize to the market, often known as Street earnings, which may be different from GAAP earnings. Therefore, following the procedure of Brown (2001), the sample of firms in which GAAP earnings from Compustat equal Street earnings from First Call are examined separately. The earnings are considered equal if the absolute value of the difference is less than US$0.02 to control for rounding differences and materiality. The results (not shown) are similar to the previous results for the reduced sample. Moreover, the difference in Street versus GAAP earnings has not increased over the sample period (not shown).

Finally, the passage of Regulation FD in August 2000 and its subsequent implementation on October 23, 2000 might affect forecasts made during the surrounding time periods. To investigate this issue, the quarterly forecast properties from the beginning of 1999 through the end of 2001 are computed for only firms that have fiscal quarters on a March, June, September, December cycle. This provides a sample with three distinct, easily identifiable subperiods: (1) a pre-Regulation FD period, from the first quarter of 1999 through the second quarter of 2000; (2) a period during the implementation of Regulation FD, the third and fourth quarters of 2000; and (3) a post-Regulation FD period, the first quarter of 2001 through the fourth quarter of 2001. The second period, during the implementation, includes the quarter in which the regulation was passed.

Table 12
Forecast dispersion, error, raw error, and optimism surrounding implementation of regulation FD

Year: month	Profit firms				Loss firms			
	Dispersion	Error	Raw error	Percent optimistic	Dispersion	Error	Raw error	Percent optimistic
Pre								
1999: 3	0.15	0.35	0.05	27.35	0.39	0.66	0.15	56.36
1999: 6	0.13	0.33	0.05	26.49	0.40	0.67	0.16	57.89
1999: 9	0.14	0.34	0.05	27.96	0.41	0.66	0.19	56.41
1999: 12	0.15	0.34	0.06	25.42	0.37	0.74	0.28	59.95
2000: 3	0.13	0.35	0.05	23.89	0.34	0.59	0.17	50.55
2000: 6	0.13	0.32	0.05	24.49	0.28	0.64	0.19	49.63
During								
2000: 9	0.13	0.31	0.06	28.71	0.23	0.60	0.19	47.68
2000: 12	0.14	0.32	0.06	29.63	0.30	0.64	0.26	56.54
Post								
2001: 3	0.14	0.33	0.05	30.90	0.33	0.51	0.17	52.74
2001: 6	0.16	0.35	0.05	27.40	0.30	0.53	0.14	51.75
2001: 9	0.16	0.37	0.06	34.47	0.34	0.56	0.18	54.89
2001: 12	0.15	0.33	0.05	22.41	0.32	0.54	0.13	47.02

This table reports mean analyst quarterly forecast properties for the quarters surrounding the implementation of Regulation Free Disclosure (Reg FD). Reg FD was passed in August 2000 and implemented in October 2000. See Table 1 for variable definitions. Only firms with fiscal quarters ending in March, June September, and December are included in the sample.

After evaluating the results, presented in Table 12 for profit and loss subsamples, there are no identifiable differences in the forecast property trends during the three periods surrounding Regulation FD implementation regardless of whether the sample includes all firms, profit firms, or loss firms.

7. Conclusions

This study documents almost continuous reductions in analyst forecast dispersion, error, and optimism during the time period 1990 through 2001. The reductions, however, primarily come about due to staggering advances in forecasting loss firm earnings. At the end of the sample period, differences in forecasting performance between profit and loss firms are relatively small. Attempts are made to control for various issues that might affect the conclusions, such as earnings management, guidance, and smoothing, Street versus GAAP earnings, or Regulation FD. None of those issues can wholly explain the trends.

In addition, it appears that loss firm earnings are more difficult to predict. Given the prediction difficulties, the value provided to the market by analysts appears to be greater for loss firms versus profit firms.

While this study does not contradict prior studies showing increases in earnings management or guidance, it does shed additional light on the issue. Analysts are undoubtedly not as optimistic, their incentives to get investment banking clients or private information perhaps no longer as important as the notoriety they receive when they mislead investors.

Future studies can examine trends in analyst buy, sell, or hold recommendations, another area in which the media and academic research (and also the Securities and Exchange Commission) have criticized analysts. Analysts are known to frequently make buy recommendations but rarely make sell recommendations, often preferring to drop coverage of a firm rather than issue a sell recommendation (e.g., Barber, Lehavy, McNichols, & Trueman, 2001; McNichols & O'Brien, 1997; Stickel, 1995).

Acknowledgements

This paper has benefited from the comments of Afshad Irani, Ahmad Etebari, Werner DeBondt, John Becker-Blease, participants at the 2003 Financial Management Conference (especially the discussant Larry Wall), participants at a University of New Hampshire Seminar, and the referee. The author would like to thank Pamela Grant and Chuck Hill of Thomson Financial Services, who have generously provided the First Call data.

References

Abarbanell, J., & Bernard, V. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance*, *47*, 1181–1208.

Abarbanell, J., & Lehavy, R. (2000). *Differences in commercial database reported earnings: Implications for inferences concerning analyst forecast rationality, the association between prices and earnings, and firm reporting discretion*. Working paper, University of North Carolina at Chapel Hill.

Ang, J., & Ciccone, S. (2001). *Analyst forecasts and stock returns*. Working paper. University of New Hampshire.

Asthana, S. (2003). Impact of information technology on post-earnings announcement drift. *Journal of Information Systems*, *17*, 1–17.

Atiase, R. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research*, *23*, 21–36.

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance*, *56*, 531–563.

Bartov, E. (1993). The timing of asset sales and earnings manipulation. *Accounting Review*, *68*, 840–855.

Bradshaw, M., & Sloan, R. (2002). GAAP versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research*, *40*, 41–65.

Brown, L. (1997). Analyst forecasting errors: Additional evidence. *Financial Analysts Journal*, *53*, 81–88.

Brown, L. (2001). A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research*, *39*, 221–241.

Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, *24*, 99–126.

Butler, K., & Saraoglu, H. (1999). Improving analysts' negative earnings forecasts. *Financial Analysts Journal*, *55*, 48–56.

Ciccone, S. (2001). *Analyst forecasts, financial distress, and business risk*. Working paper. University of New Hampshire.

Collingwood, H. (2001). The earnings game: Everybody plays, nobody wins. *Harvard Business Review*, *79*, 65–74.

DeBondt, W., & Thaler, R. (1990). Do security analysts overreact? *American Economic Review*, *80*, 52–57.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *Journal of Business*, *72*, 1–33.

Diether, K., Malloy, C., & Scherbina, A. (2002). Differences of opinion and the cross-section of stock returns. *Journal of Finance*, *57*, 2113–2141.

Dische, A. (2002). Dispersion in analyst forecasts and the profitability of earnings momentum strategies. *European Financial Management*, *8*, 211–228.

Dowen, R. (1996). Analyst reaction to negative earnings for large well-known firms. *Journal of Portfolio Management*, *23*, 49–55.

Dreman, D., & Berry, M. (1995). Analyst forecasting errors and their implications for security analysis. *Financial Analysts Journal*, *51*, 30–41.

Easterwood, J., & Nutt, S. (1999). Inefficiency in analyst forecasts: Systematic misreaction or systematic optimism? *Journal of Finance*, *54*, 1777–1797.

Fox, J. (1997, March 31). Learn to play the earnings game (and wall street will love you). *Fortune*. 76–80.

Fried, D., & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics*, *4*, 85–107.

Fudenberg, D., & Tirole, J. (1995). A theory of dividend smoothing based on incumbency rents. *Journal of Political Economy*, *103*, 75–93.

Gu, Z., & Wu, J. (2003). Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics*, *35*, 5–29.

Healy, P. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics*, *7*, 85–107.

Holthausen, R., Larcker, D., & Sloan, R. (1995). Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics*, *19*, 29–74.

Hong, H., Kubik, J., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *Rand Journal of Economics*, *31*, 121–144.

Imhoff, E., & Lobo, G. (1992). The effect of ex-ante earnings uncertainty on earnings response coefficients. *Accounting Review*, *67*, 427–439.

Kasznik, R., & Lev, B. (1995). To warn or not to warn: Management disclosures in the face of an earnings surprise. *Accounting Review*, *70*, 113–134.

Krishnaswami, S., & Subramaniam, V. (1999). Information asymmetry, valuation, and the corporate spin-off decision. *Journal of Financial Economics*, *53*, 73–112.

Matsumoto, D. (2002). Management's incentives to avoid negative earnings surprises. *Accounting Review*, *77*, 483–514.

McNichols, M., & O'Brien, P. (1997). Self-selection and analyst coverage. *Journal of Accounting Research*, *35*, 167–199.

O'Brien, P. (1988). Analysts' forecasts as earnings expectations. *Journal of Accounting and Economics*, *10*, 53–83.

Olsen, R. (1996). Implications of herding behavior for earnings estimation, risk assessment, and stock returns. *Financial Analysts Journal*, *52*, 37–41.

Richardson, S., Teoh, S., & Wysocki, P. (2001). *The walkdown to beatable analyst forecasts: The role of equity issuance and insider trading incentives*. Working paper. University of Michigan.

Skinner, D. (1994). Why firms voluntarily disclose bad news. *Journal of Accounting Research*, *32*, 38–60.

Skinner, D. (1997). Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics*, *23*, 249–282.

Skinner, D., & Sloan, R. (2002). Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies*, *7*, 289–312.

Stickel, S. (1990). Predicting individual analyst earnings forecasts. *Journal of Accounting Research*, *28*, 409–417.

Stickel, S. (1995). The anatomy of the performance of buy and sell recommendations. *Financial Analysts Journal*, *51*, 25–39.

Trueman, B. (1994). Analyst forecasts and herding behavior. *Review of Financial Studies*, *7*, 97–124.

Trueman, B., & Titman, S. (1988). An explanation of accounting income smoothing. *Journal of Accounting Research*, *26*, 127–139.

Vickers, M. (1999, May 29). Ho-Hum, another earnings surprise. *Business Week*. 83–84.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics*, *58*, 369–396.

THE JOURNAL OF FINANCE • VOL. LVI, NO. 5 • OCT. 2001

Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets

JAMES CLAUS and JACOB THOMAS*

ABSTRACT

The returns earned by U.S. equities since 1926 exceed estimates derived from theory, from other periods and markets, and from surveys of institutional investors. Rather than examine historic experience, we estimate the equity premium from the discount rate that equates market valuations with prevailing expectations of future flows. The accounting flows we project are isomorphic to projected dividends but use more available information and narrow the range of reasonable growth rates. For each year between 1985 and 1998, we find that the equity premium is around three percent (or less) in the United States and five other markets.

THE EQUITY RISK PREMIUM LIES at the core of financial economics. Representing the excess of the expected return on the stock market over the risk-free rate, the equity premium is unobservable and has been estimated using different approaches and samples. The estimates most commonly cited in the academic literature are from Ibbotson Associates' annual reviews of the performance of various portfolios of U.S. stocks and bonds since 1926. Those estimates lie in the region of seven to nine percent per year, depending on the specific series examined. This historic evidence is objective and easy to interpret and has convinced many, especially academic financial economists, that the Ibbotson estimates are the best available proxies for the equity premium (Welch (1999)).[1] For discussion purposes, we use "eight percent"

* Barclays Global Investors and Columbia Business School, respectively. We thank I/B/E/S Inc. for their database of earnings estimates and Enrique Arzac and René Stulz for many helpful suggestions and discussions. Useful comments were received from anonymous referees, Bala Dharan, Darin Clay, Ilia Dichev, Ben Esty, Bob Hodrick, Irene Karamanou, S.P. Kothari, Jimmy Liew, Jing Liu, Jim McKeown, Karl Muller, Jim Ohlson, Stephen Penman, Huai Zhang, and workshop participants at AAA annual meetings (San Diego), Columbia University, Copenhagen Business School, University of Michigan, Michigan State University, University of North Carolina-Chapel Hill, Northern Arizona University, Ohio State University, Penn State University, Prudential Securities Quantitative Conference, Syracuse University, and University of Texas-Austin.

[1] The annualized distribution of monthly common stock returns over the 30-day T-bill rate has a mean of 9.12 percent and a standard deviation of 20.06 percent (from data in Table A-16, Ibbotson Associates (1999)). If these 73 observations are independent and identically distributed, the sample mean is a reasonable estimate for the equity premium, and the standard error of 2.35 percent associated with the sample mean allows an evaluation of other hypothesized values of the equity premium.

and "the Ibbotson estimate" interchangeably to represent the historic mean of excess returns earned by U.S. equities since 1926. (Unless noted otherwise, all amounts and rates are stated in nominal, not real, terms.)

Our objective is to show empirically that eight percent is too high an estimate for the equity premium in recent years. Rather than examine observed returns, we estimate for each year since 1985 the discount rate that equates U.S. stock market valuations with the present value of prevailing forecasts of future flows. Subtracting 10-year risk-free rates from these estimated discount rates suggests that the equity premium is only about three percent.[2] An examination of five other large stock markets (Canada, France, Germany, Japan, and the United Kingdom) provides similar results. Despite substantial variation in the underlying fundamentals across markets and over time, observing that every one of our 69 country-year estimates lies well below eight percent suggests that the Ibbotson estimate is too high for our sample period. Examination of various diagnostics (such as implied future profitability) confirms that the projections required to support an eight percent equity premium are unreasonable and inconsistent with past experience.

Some features of our study should be emphasized at the outset. As we only seek to establish a reasonable upper bound for the equity premium, we select long-term growth assumptions that exceed past experience and do not adjust for optimism in the analyst forecasts used.[3] Also, we use the simplest structure necessary to conduct our analysis. Our estimates refer to a long-term premium expected to hold over all future years (whereas historical estimates measure one-period premia), and we assume that the premium is constant over those future years (we do incorporate anticipated variation in risk-free rates). Finally, each annual estimate is conditional on the information available in that year; we do not consider an unconditional equity premium toward which those conditional premia might gravitate in the long run.

We are not the first to question the validity of the Ibbotson estimate. Mehra and Prescott (1985) initiated a body of theoretical work that has examined the so-called "equity premium puzzle." Their model indicates that the variance–covariance matrix of aggregate consumption and returns on stocks and bonds, when combined with reasonable risk-aversion parameters, implies equity premium estimates that are less than one percent. Despite subsequent efforts to bridge this gap (e.g., Abel (1999)), concerns remain about the validity of the Ibbotson estimate (see Kocherlakota (1996), Cochrane (1997), and Siegel and Thaler (1997) for summaries).

[2] Gebhardt, Lee, and Swaminathan (forthcoming) find similar results when estimating firm-specific discount rates, rather than the market-level discount rates considered in this paper.

[3] As described later, analyst optimism has declined systematically over time and a simple adjustment for mean bias is inappropriate. Bayesian adjustments to control for observed analyst optimism are not considered because we focus on an upper bound. In general, we do not use more complex econometric techniques and data refinements that are available to get sharper point estimates (e.g., Mayfield (1999), Vuolteenaho (1999), and Ang and Liu (2000)).

Surveys of institutional investors also suggest an equity premium substantially below eight percent (e.g., Burr (1998)), and there are indications that this belief has been held for many years (e.g., Benore (1983)).[4] Also, the weighted average cost of capital used in discounted cash flow valuations provided in analysts' research reports usually implies an equity premium below five percent. Current share prices appear systematically overpriced if an eight percent equity premium is used on reasonable projections of future flows. This overpricing is more evident when examining mature firms, where there is less potential for disagreement about growth opportunities.

To identify possible reasons why the Ibbotson estimate might overstate the equity premium in recent years, apply the Campbell (1991) decomposition of observed returns (in excess of the expected risk-free rate) for the market portfolio. The four components are: (1) the expected equity premium for that period; (2) news about the equity premium for future periods; (3) news about current and future period real dividend growth; and (4) news about the real risk-free rate for current and future periods. Here, news represents changes in expectations between the beginning and end of the current period (for current period dividend growth and risk-free rates, it represents the unexpected portion of observed values). Summing up both sides of this relation for each year since 1926 indicates that the average excess return observed would exceed the equity premium today if: (1) conditional one-year-ahead equity premia have declined; (2) the conditional long-term equity premium anticipated for future years has declined; (3) news about real dividend growth was positive on average; or (4) the expected real risk-free rate has declined.

The first and second reasons for why the Ibbotson estimate overstates the current equity premium highlight the potential pitfalls of estimating equity premia from observed returns. Holding aside news about dividends and risk-free rates, valuations would exceed expectations if the equity premium has declined (since present values increase when expected rates of return decline). That is, unexpected changes in the equity premium cause historical equity premium estimates to move in the opposite direction. Blanchard (1993) concludes that the equity premium has declined since 1926 to two or three percent by the early 1990s, and speculates that this decline is caused by a simultaneous decline in expected real rates of return on stocks and an increase in expected real risk-free rates. (This increase in expected real risk-free rates is another puzzle, but that puzzle is beyond the scope of this paper.) The remarkable run-up in stock prices during the 1990s, both domestically as well as internationally, is also consistent with a recent decline

[4] While many argue for an equity premium between two and three percent (e.g., Bogle (1999, p. 76)), some suggest that the premium is currently close to zero (e.g., Glassman and Hassett (1998), and Wien (1998)). Surveys of *individual* investors, on the other hand, suggest equity premia even higher than the Ibbotson estimate. For example, the *New York Times* (October 10, 1997, page 1, "High hopes of mutual fund investors"), reported an equity premium in excess of 16 percent from a telephone survey conducted by Montgomery Asset Management.

in the equity premium. Stulz (1999) argues that increased globalization has caused equity premia to decline in all markets.

Examination of historic evidence over other periods and markets suggests that the U.S. experience since 1926 is unusual. Siegel (1992) finds that the excess of observed annual returns for NYSE stocks over short-term government bonds is 0.6, 3.5, and 5.9 percent over the periods 1802 to 1870, 1871 to 1925, and 1926 to 1990, respectively. Jorion and Goetzmann (1999) examine the evidence for 39 equity markets going back to the 1920s, and conclude that the high equity premium observed in the United States appears to be the exception rather than the rule. Perhaps some stock markets collapsed and those markets that survived, like the U.S. exchanges, exhibit better performance than expected (see Brown, Goetzmann, and Ross (1995)). This evidence is consistent with the third reason for the high Ibbotson premium: since 1926, news about real dividend growth for U.S. stocks has been positive on average.

Partially in response to these limitations of inferring equity premia from observed returns, financial economists have considered forward-looking approaches based on projected dividends.[5] Informally, expected rates of return on the market equal the forward dividend yield plus expected growth in dividends (this dividend growth model is discussed in Section I). While dividend yields are easily measured, expected dividend growth in perpetuity is harder to identify. Proxies used for expected dividend growth include observed growth in earnings, dividends, or economy-wide aggregates (e.g., Fama and French (2000)). Unfortunately, the dividend growth rate that can be sustained in perpetuity is a hypothetical rate that is not necessarily anchored in any observable series, leaving considerable room for disagreement (see the Appendix for explanation).

We use a different forward-looking approach, labeled the abnormal earnings (or residual income) model, to mitigate problems associated with the dividend growth model.[6] Recognizing that dividends equal earnings less changes in accounting (or book) values of equity allows the stream of projected dividends to be replaced by the current book value of equity plus a function of future accounting earnings (details follow in Section I). While book values feature prominently in the model, the inclusion of future abnormal earnings makes it isomorphic to the dividend discount model. Relative to the dividend growth model, this approach makes better use of currently

[5] A related approach is to run predictive regressions of market returns or equity premium on dividend yields and other variables (e.g., Campbell and Shiller (1988)). We do not consider that approach because the declining dividend yields in recent years have caused predicted equity premium to turn negative (e.g., Welch (1999)).

[6] The approach appears to have been discovered independently by a number of economists and accountants over the years. Preinreich (1938) and Edwards and Bell (1961) are two early cites. More recently, a large body of analytical and empirical work has utilized this insight (e.g., Penman (1999)). Examples of empirical investigations include market myopia (Abarbanell and Bernard (1999)), explaining cross-sectional variation in returns (Liu and Thomas (2000)), and stock picking (Frankel and Lee (1998a, 1998b)).

available information to reduce the importance of assumed growth rates, and it narrows the range of allowable growth rates by focusing on growth in rents, rather than dividend growth.

If the equity premium is as low as our estimates suggest, required rates of return (used for capital budgeting, regulated industries, and investment decisions) based on the Ibbotson estimate are severely overstated. Second, a smaller equity premium reduces the importance of estimating beta accurately (because required rates of return become less sensitive to variation in beta) and increases the magnitude of beta changes required to explain abnormal returns observed for certain market anomalies. Finally, reducing substantially the magnitude of the equity premium puzzle to be explained might reinvigorate theory-based studies.

In Section I we develop the abnormal earnings approach used in this paper and compare it with the dividend growth model. Section II contains a description of the sample and methodology. The equity premium estimates for the United States are reported in Section III, and those for the five other markets are provided in Section IV. To confirm that our estimates are robust, we conducted extensive sensitivity analyses, which we believe represent an important contribution of our research effort. A summary of that investigation is reported in Section V (details are provided in Claus and Thomas (1999a)) and Section VI concludes.

I. Dividend Growth and Abnormal Earnings Models

The Gordon (1962) dividend growth model is described in equation (1). This relation implies that the expected rate of return on the stock market (k^*) equals the forward dividend yield (d_1/p_0) plus the dividend growth rate in perpetuity (g) expected for the market.

$$p_0 = \frac{d_1}{k^* - g} \Rightarrow k^* = \frac{d_1}{p_0} + g \tag{1}$$

where

p_0 = current price, at the end of year 0,
d_t = dividends expected at the end of future year t,
k^* = expected rate of return on the market, derived from the dividend growth model, and
g = expected dividend growth rate, in perpetuity.

The Gordon growth model is a special case of the general Williams (1938) dividend discount model, detailed in equation (2), where dividend growth is constrained to equal g each year.

$$p_0 = \frac{d_1}{(1+k^*)} + \frac{d_2}{(1+k^*)^2} + \frac{d_3}{(1+k^*)^3} + \dots \tag{2}$$

Research using the dividend growth model has often assumed that g equals forecasted earnings growth rates obtained from sell-side equity analysts, who provide earnings forecasts along with their buy/sell recommendations. These forecasts refer to earnings growth over the next "cycle," which is commonly interpreted to represent the next five years. Consequently, we refer to this earnings growth forecast as g_5. While most studies using g_5 as a proxy for g have focused on the U.S. market alone (e.g., Brigham, Shome, and Vinson (1985)), some have examined other major equity markets also (e.g., Khorana, Moyer, and Patel (1997)). Estimates of the equity premium based on the assumption that g equals g_5 are similar in magnitude to the Ibbotson estimate derived from historical data. For example, Moyer and Patel (1997) estimate the equity premium each year over their 11-year sample period (1985 to 1995) and generate a mean estimate of 9.38 (6.96) percent relative to the 1-year (30-year) risk-free rate.

However, others have balked at using g_5 as a proxy for g (e.g., Malkiel (1996), Cornell (1999)) because it appears unreasonably high at an intuitive level, and have stepped down assumed growth rates. Forecasted values of g_5 for the United States over our sample period, which are close to 12 percent in all years, exceed nominal growth in S&P earnings, which has been only 6.6 percent since the 1920s (*Wall Street Journal*, June 16, 1997, "As stocks trample price measures, analysts stretch to justify buying"). Also, the real growth rate implied by the nominal 12 percent earnings growth rate exceeds both forecast and realized growth in GDP (since 1970, forecasts of expected real growth in GDP have averaged 2.71 percent, and realized real growth has averaged 2.81 percent).

While we show that g_5 is systematically optimistic relative to realized earnings, it is difficult to infer reliably the level of that optimism from the relatively short time-series of forecast errors available (reliable data on analyst forecasts go back only about 15 years). Moreover, the incentives for analysts to make optimistic forecasts vary across firms and over time. For example, the literature on U.S. analysts' forecasts suggests that while analysts tended to make optimistic forecasts early in our sample period (to curry favor with management), more recently, management has tended to guide near-term analyst forecasts downward to be able to meet or beat them when announcing earnings.[7] Even if unbiased estimates of near-term earnings growth (g_5) were available, the Appendix describes why those estimates as well as observed growth rates are conceptually different from g, the hypothetical dividend growth that can be sustained in perpetuity.

[7] Results reported in Table VI offer clear evidence of such a decline in optimism for all horizons. Bagnoli, Beneish, and Watts (1999) document how recent analyst forecasts are systematically below reported earnings for their sample, and also below "whisper" forecasts that are generally viewed as representing the market's true earnings expectations. Matsumoto (1999) offers evidence in support of management guiding analyst forecasts downward, and also investigates factors that explain cross-sectional variation in this propensity to guide analysts.

The abnormal earnings model is an alternative that mitigates many of the problems noted above. Expected dividends can be related to forecasted earnings using equation (3) below, and that relation allows a conversion of the discounted dividends relation in equation (2) to the abnormal earnings relation in equation (4).

$$d_t = e_t - (bv_t - bv_{t-1}) \tag{3}$$

$$p_0 = bv_0 + \frac{ae_1}{(1+k)} + \frac{ae_2}{(1+k)^2} + \frac{ae_3}{(1+k)^3} + \ldots, \tag{4}$$

where

e_t = earnings forecast for year t,

bv_t = expected book (or accounting) value of equity at the end of year t,

$ae_t = e_t - k(bv_{t-1})$ = expected abnormal earnings for year t, or forecast accounting earnings less a charge for the cost of equity, and

k = expected rate of return on the market portfolio, derived from the abnormal earnings model.

Equation (3), also known as the "clean surplus" relation, requires that all items affecting the book value of equity (other than transactions with shareholders, such as dividends and share repurchases/issues) be included in earnings. Under U.S. accounting rules, almost all transactions satisfy the clean-surplus assumption. An examination of the few transactions that do not satisfy this relation suggests that these violations occur ex post, and are not anticipated in analysts' earnings forecasts (e.g., Frankel and Lee (1998b)). Since we construct future book values using equation (3), by adding forecast income to and subtracting forecast dividends from beginning book values, clean surplus is maintained and the dividend and abnormal earnings relations in equations (2) and (4) are isomorphic.

Equation (4) shows that the current stock price equals the current book value of equity plus the present value of future expected abnormal earnings. Abnormal earnings, a proxy for economic profits or rents, adjusts reported earnings by deducting a charge for equity capital. Note that the market discount rates estimated from the abnormal earnings and dividend growth approaches are labeled differently: k and k^*. Also, the standard transversality conditions apply to both models: in the limit as t approaches infinity, the present value of future price, p_t (difference between price and book value, $p_t - bv_t$) must tend to zero in equation (2) (in equation (4)).

Financial economists have expressed concerns about accounting earnings deviating from "true" earnings (and book values of equity deviating from market values), in the sense that accounting numbers are noisy and easily manipulated. However, the equivalence between equations (2) and (4) is not impaired by differences between accounting and economic numbers, nor is it affected by the latitude available within accounting rules to report different accounting numbers. As long as forecasted earnings satisfy the clean surplus relation in equation (3) in terms of expectations, equation (4) is simply an algebraic restatement of equation (2), subject to the respective transversality conditions mentioned above.

Since the I/B/E/S database we use does not provide analysts' earnings forecasts beyond year +5, we assume that abnormal earnings grow at a constant rate (g_{ae}) after year +5, to incorporate dates past that horizon. Equation (4) is thus adapted as follows.

$$p_0 = bv_0 + \frac{ae_1}{(1+k)} + \frac{ae_2}{(1+k)^2} + \frac{ae_3}{(1+k)^3} + \frac{ae_4}{(1+k)^4} + \frac{ae_4}{(1+k)^5} + \left[\frac{ae_5(1+g_{ae})}{(k-g_{ae})(1+k)^5}\right]. \tag{5}$$

The last, bracketed term is a terminal value that captures the present value of abnormal earnings after year +5. The terms before are derived from accounting statements (bv_0) and analyst forecasts (e_1 to e_5). Note that there are three separate growth rates in this paper and the different growth rates refer to different streams and periods and arise from different sources. The rate g refers to dividend growth in perpetuity and is assumed by the researcher; g_5 refers to growth in accounting earnings over the first five years and is provided by financial analysts; and g_{ae} refers to abnormal earnings growth past year +5 and is assumed by the researcher.

Whereas expected rates of return are typically viewed as being stochastic (Samuelson (1965)), k^* and k in equations (1) and (5) are nonstochastic discount rates. Barring a few recent exceptions (e.g., Ang and Liu (2000) and Vuolteenaho (1999)), the literature has assumed that expected rates of return can be approximated by discount rates. We make that assumption too. While equation (1) is designed to only reflect a flat k^*, equation (2) can be restated to incorporate predictable variation over time in discount rates. Similarly, equation (5) can be restated to incorporate nonflat discount rates, as shown in Claus and Thomas (1999a). We consider the case when the equity premium is assumed to remain flat but discount rates vary over future periods based on the term-structure of risk-free rates. This restated version of equation (5) is

$$p_0 = bv_0 + \sum_{t=1}^{\infty}\left[\frac{ae_t}{\prod_{s=1}^{t}(1 + r_{fs} + rp)}\right], \tag{5a}$$

where

r_{fs} = forward one-year risk-free rate for year s,
rp = equity risk premium, assumed constant over all future years,
ae_t = expected abnormal earnings for year t, equals $e_t - bv_{t-1}(r_{ft} + rp)$ for years +1 through +5, and equals $ae_5(1 + g_{ae})^{t-5}$, from year +6 on.

While the abnormal earnings stream in equation (4) is equivalent to the corresponding dividend stream in equation (2), the abnormal earnings relation in equation (5) (and equation (5a)) offers the following advantages over the dividend growth model in equation (1). First, a substantial fraction of the "value profile" for the abnormal earnings model in equation (5) is fixed by numbers that are currently available and do not need to be assumed by the researcher (current book value and abnormal earnings for years +1 through +5). Value profile is a representation of the fraction of total value captured by each future year's flows. In contrast, the entire value profile for the dividend growth model is affected by the assumed growth rate, g. Since the fraction of value determined by assumed growth rates is lower for the abnormal earnings approach, those risk premium estimates are more reliable.

Second, in contrast to the potential for disagreement about a reasonable range for g, the rate at which rents can grow in perpetuity after year +5, g_{ae}, is less abstract and easier to gauge using economic intuition. For example, to obtain equity premia around 8 percent, rents at the market level would have to grow forever at about 15 percent, on average. It is unlikely that aggregate rents to U.S. equity holders would grow at such high rates in perpetuity because of factors such as antitrust actions, global competition, and pressure from other stakeholders. The historical evidence (e.g., Myers (1999)) is also at odds with such high growth rates in abnormal earnings.

Third, future streams for a number of value-relevant indicators, such as price-to-book ratios (P/B), price-to-earnings ratios (P/E), and accounting return on equity (*roe*), can also be projected under the abnormal earnings approach. This allows one to paint a more complete picture of the future for different assumed growth rates. Analysis of the levels of future P/B and profitability (excess of *roe* over k) implied by growth rates required to obtain equity premium estimates around eight percent are also inconsistent with past experience.

II. Data and Methodology

I/B/E/S provides the consensus of all available individual forecasts as of the middle (the Thursday following the second Friday) of each month. Forecasts and prices should be gathered soon after the prior year-end, as soon as equity book values (bv_0) are available. Rather than collect forecasts at different points in the year, depending on the fiscal year-end of each firm, we opted to collect data as of the same month each year for all firms to ensure that the risk-free rate is the same across each annual sample. Since most firms have December year-ends, and book values of equity can be obtained from the balance sheets that are required to be filed with the SEC within 90 days of the fiscal year-end, we collect forecasts as of April each year.[8] For

[8] For the few firm-years not filing within this 90-day deadline, the book value of equity can be inferred by the market by adding (subtracting) fourth quarter earnings (dividends) from the third quarter book value of equity.

firms with fiscal year-ends other than December, this procedure creates a slight upward bias in estimated equity premium, since the stock prices used (as of April) are on average higher than those near the prior year's fiscal year-end, when bv_0 was released. In addition to earnings forecasts, I/B/E/S also provides data for actual earnings per share, dividends per share, share prices, and the number of outstanding shares. Equity book values are collected from COMPUSTAT's Industrial Annual, Research, and Full Coverage Annual Files, for years up to and including 1997.

The sample includes firms with I/B/E/S earnings forecasts for years +1 and +2 (e_1 and e_2) and a five-year growth forecast (g_5) as well as share prices and shares outstanding as of the I/B/E/S cut off date each April. We also require nonmissing data for the prior year's book value, earnings, and dividends. Explicit forecasts for years +3, +4, and +5 are often unavailable, and are generated by projecting the growth rate g_5 on the prior year's earnings forecast: $e_t = e_{t-1}(1 + g_5)$.[9]

Earlier years in the I/B/E/S database, before 1985, were dropped because they provided too few firms with complete data to represent the overall market. From 1985 on, the number of firms with available data increases substantially. As shown in column 1 of Table I, the number of sample firms increases from 1,559 in 1985 to 3,673 in 1998. Comparison with the total number of firms and market capitalization of all firms on NYSE, AMEX, and Nasdaq each April indicates that, although our sample represents only about 30 percent of all such firms, it represents 90 percent or more of the total market capitalization. Overall, we believe our sample is fairly representative of the value-weighted market, and refer to it as "the market" hereafter.

Firm-level data are aggregated each year to generate market-level earnings, book values, dividends, and capitalization. Actual data for year 0 (the full fiscal year preceding each April when forecasts were collected) is provided in columns 2 through 6 of Table I. Forecasted and projected earnings for years +1 through +5 are reported in columns 7 through 11.

Table I reveals an interesting finding relating to dividend payouts: the ratio of market dividends to earnings is around 50 percent in most years (with a noticeable decline toward the end of the sample period).[10] We use this 50 percent payout ratio to project future dividends from earnings fore-

[9] If any of the explicit earnings forecasts for years +2, +3, +4, or +5 were negative, they were not used to project earnings for subsequent years. For about five percent of our sample, explicit earnings forecasts are available for all five years and do not need to be inferred using g_5. That subsample was investigated to confirm that projections based on five-year growth rates are unbiased proxies for the explicit forecasts for those years.

[10] Although this statistic is well known to macroeconomists, it is higher than average firm-level dividend payouts. Note, however, that aggregate earnings include many loss firms, especially in the early 1990s, when earnings were depressed because of write-offs and accounting changes. This results in a higher aggregate dividend payout than the average firm-level payout ratio, which is computed over profitable firms only (the payout ratio is meaningless for loss firms). Also, since the aggregate payout ratio is a value-weighted average dividend payout, it is more representative of large firms, which tend to have higher dividend payouts than small firms.

Table I

Market Capitalization, Book Values, Dividends, and Actual and Forecast Earnings for U.S. Stocks (1985 to 1998)

The market consists of firms on the I/B/E/S Summary files with forecasts for years +1, +2, and a five-year earnings growth estimate (g_5) as of April each year, and actual earnings per share, dividends per share, number of shares outstanding and share prices as of the end of the prior fiscal year (year 0). Book values of equity for year 0 are obtained from COMPUSTAT. When missing on the I/B/E/S files, forecasted earnings per share for years +3, +4, and +5 are determined by applying g_5, the forecasted five-year growth rate, to year +2 forecasted earnings. All per share numbers are multiplied by the number of shares outstanding to get amounts at the firm level, and these are added across firms to get amounts at the market level each year. All amounts, except for dividend payout, are in millions of dollars.

Forecast as of April	Number of Firms	Actual Values for Year 0					Forecast Earnings for Years +1 to +5				
		Earnings	Dividends	Payout	Book Value	Market Value	Year +1	Year +2	Year +3	Year +4	Year +5
	1	2	3	4	5	6	7	8	9	10	11
1985	1,559	154,858	71,134	46%	1,191,869	1,747,133	180,945	205,294	228,208	254,181	283,706
1986	1,613	155,201	73,857	48%	1,214,454	2,284,245	178,024	203,677	226,018	251,313	280,035
1987	1,774	146,277	81,250	56%	1,323,899	2,640,743	186,319	220,178	244,174	271,432	302,529
1988	1,735	167,676	86,237	51%	1,430,672	2,615,857	222,497	246,347	273,204	303,642	338,262
1989	1,809	229,070	97,814	43%	1,541,231	2,858,585	261,278	284,616	315,204	349,721	388,776
1990	1,889	228,216	107,316	47%	1,636,069	3,143,879	257,657	295,321	328,803	366,798	410,028
1991	1,939	218,699	108,786	50%	1,775,199	3,660,296	241,760	294,262	328,513	367,521	412,073
1992	2,106	202,275	113,962	56%	1,911,383	4,001,756	252,109	308,567	344,742	386,098	433,552
1993	2,386	247,988	127,440	51%	2,140,668	4,918,359	295,862	356,086	397,969	445,840	501,081
1994	2,784	290,081	129,186	45%	2,168,446	5,282,046	339,694	402,689	450,559	505,315	568,179
1995	2,965	365,079	147,575	40%	2,670,725	6,289,760	444,593	518,600	579,954	650,120	730,648
1996	3,360	446,663	175,623	39%	3,182,952	8,207,274	512,921	588,001	659,732	742,244	837,577
1997	3,797	547,395	201,017	37%	3,679,110	10,198,036	614,932	709,087	800,129	905,787	1,029,061
1998	3,673	526,080	178,896	34%	3,412,303	12,908,495	577,297	682,524	775,707	884,529	1,012,294

casts, as well as to project future book values (using equation (3)). The validity of this assumption is not critical; however, varying the payout ratio between 25 and 75 percent has little impact on the estimated discount rate (results available upon request).

Both short- and long-term risk-free rates have been used in studies that estimate discount rates from flows that extend over many future periods. While one-month or one-year rates are appropriate when inferring the equity premium from historic returns (observed return less risk-free yield for that period), for studies based on forecasted flows, the maturity of risk-free rates used should match that of the future flows (Ibbotson Associates (1999)). Although we allow for expected variation in risk-free rates when estimating the risk premium, using equation (5a), we find almost identical results using a constant risk-free rate in equation (5) equal to the long-term rate. In essence, the shape of the yield curves over our sample period is such that the forward rates settle rather quickly at the long-term rate, and the impact of discounting flows from earlier years in the profile at rates lower than the long-term rate is negligible. For the sensitivity analyses, we find it convenient to use the constant rate structure of equation (5), rather than the varying rate structure of equation (5a). We selected the 10-year risk-free rate for the constant risk-free rate because it is the longest maturity for which data could be obtained for all country-years in our sample. To allow comparisons with other studies that use 30-year risk-free rates, we note that the mean 30-year risk-free rate in April for each year of our U.S. sample period is 31 basis points higher than the mean 10-year risk-free rate we use.

For years beyond year +5, abnormal earnings are assumed to grow at the expected inflation rate, g_{ae}. As explained in the Appendix, the expected nominal inflation rate is higher than values of g_{ae} assumed in the literature, and is an upper bound for expected growth in abnormal earnings. We derive the expected inflation rate from the risk-free rate, based on the assumption that the real risk-free rate is approximately three percent.[11] Since we recognize that this assumption is only an educated guess, we consider in Section V.D other values of g_{ae} also. Fortunately, our estimated risk premium is relatively robust to variation in the assumed growth rate, g_{ae}, since a lower proportion of current market value is affected by g_{ae} in equations (5) and (5a), relative to the impact of g in equation (1).

III. Results

Since k appears in both the numerators (ae_t is a function of k) and denominators of the terms on the right-hand side of equation (5), the resulting

[11] The observed yields on recently issued inflation-indexed government bonds support this assumption. Although estimates of the real risk-free rate vary through time, and have historically been lower than three percent, more recently, the excess of the long-term risk-free rate over inflation forecasts has risen to three or four percent (e.g., Blanchard (1993), and discussion by Siegel).

equation is a polynomial in k with many possible roots. Empirically, however, only one root is real and positive (see Botosan (1997)). We search manually for the value of k that satisfies the relation each year, with the first iteration being close to the risk-free rate. The equity risk premium estimate (rp) that satisfies the valuation relation in equation (5a) is also estimated iteratively.

Table II provides the results of estimating rp, k, and k^*. The annual estimates for rp (in column 13) lie generally between three and four percent and are much lower than the historic Ibbotson estimate. Also, there is little variation over time: each annual estimate is remarkably close to the mean value of 3.39 percent. The annual estimates for k (in column 9) vary between a high of 14.38 percent in 1985 and a low of 8.15 percent in 1998. The corresponding risk-free rates (10-year Government T-bond yields) reported in column 8 vary with the estimated ks, between 11.43 percent in 1985 and 5.64 percent in 1998. As a result, the estimated equity premia (in column 11), equal to k less r_f, exhibit little variation around the time-series mean of 3.40 percent.

While the equation (5a) equity premium estimates (rp) derived from non-flat risk-free rates are in concept more accurate than those derived by subtracting 10-year risk-free rates from the flat k estimated from equation (5), the numbers reported in column 11 are very similar to those reported in column 13. We only consider the equation (5) estimates hereafter because (a) the magnitudes of the discount rates and their relation to risk-free rates are more transparent for the risk premium estimates based on constant risk-free rates, and (b) forward one-year rates for different maturities are not available for the other five markets,.

To understand better the relative magnitudes of the terms in equation (5), we report in the first seven columns of Table II the fraction of market values represented by each term. The fraction represented by book value (column 1) has generally declined over our sample period, from 68.2 percent in 1985 to 26.4 percent in 1998. To compensate, the fraction represented by terminal value (column 7) has increased from 26.6 percent in 1985 to 60 percent in 1998. The fraction represented by abnormal earnings for years +1 to +5 has also increased.

Column 10 of Table II contains our estimates for k^*, the market discount rate based on the dividend growth model described by equation (1), when dividends are assumed to grow in perpetuity at the five-year growth in earnings forecast (g_5). Since g_5 is not available at the aggregate level, we use the forecast growth in aggregate earnings from year +4 to +5 (see column 16 of Table V) to identify g_5 at the market level. To maintain consistency with prior research using the dividend growth model, we estimate d_1 by applying the earnings growth forecast for year 1 on prior year dividends ($d_1 = d_0 * e_1/e_0$). Our estimates for k^* are almost identical to those reported by Moyer and Patel (1997).[12] Note that these estimates of k^* are much larger than the

[12] Similar results are expected because the underlying data is taken from the same source, with minor differences in samples and procedures; for example, they use the S&P 500 index whereas we use all firms with available data.

Table II

Implied Expected Rate of Return on the Market (k and k^*) and Equity Risk Premium (rp and $k - r_f$) for U.S. Stocks (1985 to 1998)

The market is an aggregate of firms on the I/B/E/S Summary files with forecasts for years +1, +2, and a five-year earnings growth estimate (g_5) as of April each year, and actual earnings, dividends, number of shares outstanding and prices as of the end of the prior full fiscal year (year 0). Book values of equity for year 0 (bv_0) are obtained from COMPUSTAT. When missing, forecasted earnings for years +3, +4, and +5 are determined by applying g_5, the forecasted five-year growth rate, to year +2 forecasted earnings. The implied discount rate that satisfies the valuation relation in equation (5) below is k. Abnormal earnings (ae_t) equal reported earnings less a charge for the cost of equity (= beginning book value of equity $* k$). Assuming that 50 percent of earnings are retained allows the estimation of future book values from current book values and forecast earnings. The terminal value represents all abnormal earnings beyond year +5. Those abnormal earnings are assumed to grow at a constant rate, g_{ae}, which is assumed to equal the expected inflation rate, and is set equal to the current 10-year risk-free rate less 3 percent. The expected rate of return on the market is also estimated using equation (1), and is labeled k^*. Equation (1) is derived from the dividend growth model, and dividend growth in perpetuity, g, is assumed to equal the five-year earnings growth rate, g_5. Subtracting r_f from the discount rates k and k^* generates equity premium estimates. The equity premium (rp) is also estimated using equation (5a), which is based on the same information used in equation (5), except that the constant discount rate k is replaced by forward one-year risk-free rates at different maturities (r_{fs}) plus a constant risk premium (rp). All amounts, except for rates of return, are in millions of dollars.

$$k^* = \frac{d_1}{p_0} + g \tag{1}$$

$$p_0 = bv_0 + \frac{ae_1}{(1+k)} + \frac{ae_2}{(1+k)^2} + \frac{ae_3}{(1+k)^3} + \frac{ae_4}{(1+k)^4} + \frac{ae_5}{(1+k)^5} + \left[\frac{ae_5(1+g_{ae})}{(k-g_{ae})(1+k)^5}\right] \tag{5}$$

$$p_0 = bv_0 + \sum_{t=1}^{\infty}\left[\frac{ae_t}{\prod_{s=1}^{t}(1+r_{fs}+rp)}\right] \tag{5a}$$

Forecast as of April	Book Value as Percent of Market Value	Percent of Market Value Represented by Present Value of											
		ae_1	ae_2	ae_3	ae_4	ae_5	Terminal Value	10-year r_f	k from (5)	k^* from (1)	$k - r_f$	$k^* - r_f$	rp from (5a)
	1	2	3	4	5	6	7	8	9	10	11	12	13
1985	68.2%	0.5%	0.9%	1.1%	1.3%	1.5%	26.6%	11.43%	14.38%	16.14%	2.95%	4.71%	2.88%
1986	53.2%	1.6%	2.0%	2.1%	2.3%	2.4%	36.3%	7.30%	11.28%	14.90%	3.98%	7.60%	4.03%
1987	50.1%	1.3%	1.9%	2.1%	2.2%	2.3%	40.0%	8.02%	11.12%	15.08%	3.10%	7.06%	3.25%
1988	54.7%	1.7%	1.8%	1.9%	2.0%	2.2%	35.7%	8.72%	12.15%	15.52%	3.43%	6.80%	3.58%
1989	53.9%	2.0%	2.0%	2.0%	2.1%	2.2%	35.7%	9.18%	12.75%	14.85%	3.57%	5.67%	3.54%
1990	52.0%	1.6%	2.0%	2.1%	2.2%	2.3%	37.8%	8.79%	12.33%	15.41%	3.54%	6.62%	3.56%
1991	48.5%	1.1%	1.9%	2.0%	2.2%	2.4%	41.8%	8.04%	11.05%	15.16%	3.01%	7.12%	2.96%
1992	47.8%	1.1%	1.9%	2.1%	2.3%	2.5%	42.4%	7.48%	10.57%	15.55%	3.09%	8.07%	3.06%
1993	43.5%	1.7%	2.3%	2.5%	2.7%	2.9%	44.4%	5.97%	9.62%	15.12%	3.65%	9.15%	3.76%
1994	41.1%	2.1%	2.6%	2.8%	2.9%	3.1%	45.5%	5.97%	10.03%	15.02%	4.06%	9.05%	3.53%
1995	42.5%	2.1%	2.6%	2.7%	2.8%	3.0%	44.3%	7.06%	11.03%	14.96%	3.97%	7.90%	4.02%
1996	38.8%	2.2%	2.5%	2.6%	2.8%	3.0%	48.2%	6.51%	9.96%	14.96%	3.45%	8.45%	3.50%
1997	36.1%	2.2%	2.5%	2.6%	2.8%	3.0%	50.8%	6.89%	10.12%	13.88%	3.23%	6.99%	3.25%
1998	26.4%	2.1%	2.5%	2.7%	3.0%	3.2%	60.0%	5.64%	8.15%	13.21%	2.51%	7.57%	2.53%
Mean								7.64%	11.04%	14.98%	3.40%	7.34%	3.39%

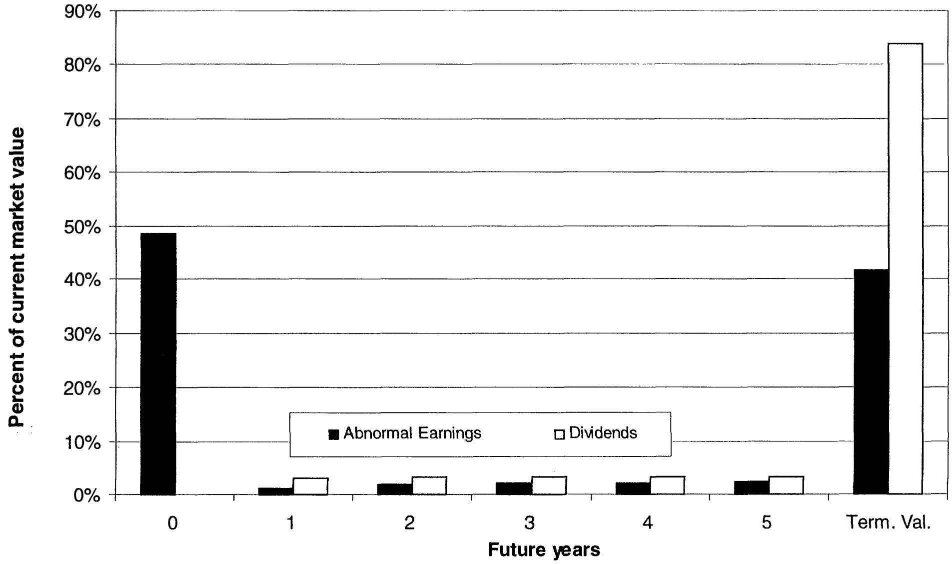


Figure 1. Comparison of value profile for abnormal earnings versus dividends, for abnormal earnings approach for U.S. stocks as of April, 1991. Based on the data in Table II, for the abnormal earnings approach described by equation (5), abnormal earnings are assumed to grow at 5.04 percent, the anticipated inflation rate, past year +5, and the resulting market discount rate (k) is 11.05 percent. For the abnormal earnings profile, the fractions represented by book value, abnormal earnings in years +1 through +5, and the terminal value are shown by the solid columns. For the dividend profile corresponding to those abnormal earnings projections, the fractions of current market capitalization that are represented by dividends in years +1 through +5 and the terminal value are shown by the hollow columns.

corresponding values of k, and the implied equity premium estimates reported in column 12 ($k^* - r_f$) are about twice those in column 11 ($k - r_f$). The mean equity premium of 7.34 percent in column 12 of Table II is approximately the same as the Ibbotson estimate. Note also the larger variation in column 12, around this mean, relative to the variation in columns 11 and 13.

The results in Table II can be used to illustrate two primary advantages of the abnormal earnings model over the dividend growth model. First, the abnormal earnings approach uses more available "hard" data (current book value and forecast abnormal earnings for years +1 to +5) to reduce the emphasis on "softer" growth assumptions (g_{ae}) used to build terminal values. Figure 1 contains a value profile for the terms in equation (5), using data for 1991. This year was selected because it represents a "median" profile: the terminal value is a smaller (larger) fraction of total value for years before (after) 1991. Recall from Table II that our estimate for k in 1991 is 11.05 percent. The terminal value is based on abnormal earnings growing at an anticipated inflation rate of 5.04 percent (g_{ae} is three percent less than the risk-free rate of 8.04 percent). The value profile for the abnormal earnings model, represented by the solid columns in Figure 1, shows that approximately 50 percent of the total value is captured by current book value, 10 percent is spread over the abnormal earnings for the next five years, and about 40 percent remains in the terminal value. This last term is the only one affected by our growth assumption. In contrast, for the dividend growth model in equation (1), the dividend growth rate (g), which is assumed to equal the five-year analyst forecast for earnings growth (g_5 = 12.12 percent), is the primary determinant of the estimated k^* (= 15.16 percent).

To offer a different perspective on why growth assumptions are more influential for projected dividends, relative to abnormal earnings, we converted the abnormal earnings profile in Figure 1 to an isomorphic value profile for dividends, represented by the hollow columns in Figure 1. (Note that these dividends refer to the flows underlying k, from the abnormal earnings model, and are different from the flows underlying k^*, the dividend growth model estimate.) The year +5 terminal value for the dividend profile in Figure 1 corresponds to a dividend growth in perpetuity of 6.8 percent.[13] Even though the abnormal earnings and dividend profiles in Figure 1 correspond to the same underlying projections, the terminal value for the dividend profile represents almost 85 percent of total value. As a result, assumed dividend growth rates have a larger impact on estimated discount rates, relative to abnormal earnings growth rate assumptions. For example, doubling the assumed value of g_{ae} to 10 percent increases the estimated discount rate by only about two percentage points. In contrast, increasing the dividend growth assumption by one percentage point raises the estimated discount rate by almost the same amount.[14]

The second major benefit of the abnormal earnings approach is that we can narrow the range of reasonable growth assumptions (g_{ae}), relative to the assumed growth rate for dividends (g). Since g is a hypothetical rate, it is not easy to determine whether 12.12 percent (the value of g underlying our 1991 estimate for k^*) is more or less reasonable than the 6.8 percent dividend growth in perpetuity (after year +5) implied by our abnormal earnings model projections. Fortunately, restating implied dividend growth rates in terms of terminal growth in abnormal earnings makes it easier to see why some dividend growth assumptions are unreasonable. The assumption that dividends grow at 12.12 percent implies that abnormal earnings past year +5 would need to grow in perpetuity at about 15 percent per year in equa-

[13] This dividend growth rate is obtained by using equation (1) on projected market value in year +5, rather than current market values (p_0) and the dividend in year six is the dividend in year +5 (= 50 percent of the earnings forecast for year +5) times the unknown growth rate. That is, solve for g in the relation $p_5 = d_5(1 + g)/(k - g)$.

[14] Note that in equation (1), changes in g increase k^* by exactly the same amount. For the dividend value profile in Figure 1, however, dividends for years +1 to +5 have been fixed by forecasted earnings and dividend payout assumptions. Therefore, increases in the dividend growth rate underlying the terminal value increase the estimated discount rate by a slightly smaller amount.

tion (5). This abnormal earnings growth rate corresponds to a real growth in rents of 10 percent (assumed long-term inflation rate is 5.04 percent), which is clearly an unreasonably optimistic assumption.

In sum, our estimates of the equity risk premium using the abnormal earnings approach are considerably lower than the Ibbotson rate, even though we believe the analyst forecasts we use, as well as the terminal growth assumptions we make, are optimistic. Adjusting for such optimism would lower our estimates further. While our estimates from the dividend growth approach are much closer to the Ibbotson rate, we believe they are biased upward because the assumed growth rate ($g = g_5$) is too high an estimate for dividend growth in perpetuity. The estimates from the abnormal earnings approach are more reliable because we use more available information to reduce the importance of assumed growth rates, and we are better able to reject growth rates as being infeasible by projecting rents rather than dividends. Additional benefits of using the abnormal earnings approach are illustrated in Section V.

IV. Equity Premium Estimates from Other Markets

Other equity markets offer a convenient opportunity to validate our domestic results. As long as the different markets are integrated with the United States and are of similar risk, those markets' estimates should proxy for the equity premium in the United States. We replicated the U.S. analysis on five other important equity markets with sufficient data to generate reasonably representative samples of those markets. Only a summary of our results is provided here; details of those analyses are in Claus and Thomas (1999b). The six markets exhibit considerable diversity in performance and underlying fundamentals over our sample period. This across-market variation increases the likelihood that the estimates we obtain from each market offer independent evidence.

As with the U.S. data, earnings forecasts, actual earnings per share, dividends per share, share prices, and the number of outstanding shares are obtained from I/B/E/S. Book values of equity as of the end of year 0 are collected from COMPUSTAT and Global Vantage for Canada and from Datastream for the remaining four countries. Unlike I/B/E/S and COMPUSTAT, Datastream drops firms that are no longer active. While such deletions are less frequent outside the United States, only surviving firms are included in our sample. Fortunately, no bias is created in this study since we equate market valuations with contemporaneous forecasts, and do not track performance.[15] Therefore, even if the surviving firms (included in our sample) performed systematically better or worse than firms that were dropped, our equity premium estimates are unbiased as long as market prices and earnings forecasts in each year are efficient and incorporate the same information.

[15] Note that there is no "backfilling" in our sample, where prior years' data for successful firms are entered subsequently.

All data are denominated in local currency. Currency risk is not an issue here, since it is present in the required rates of returns for both equities and government bonds. Thus the difference between the two rates should be comparable across countries.

We find that analysts' forecasts in these five markets exhibit an optimism bias, similar to that observed in the United States. We considered other potential sources of measurement error in the forecasts, but are confident that any biases created by these errors are unlikely to alter our equity premium estimates much. For example, in Germany, earnings could be computed in as many as four different ways: GAAP per International Accounting Standards, German GAAP, DVFA, and U.S. GAAP.[16] I/B/E/S employees indicated that they have been more successful at achieving consistency in recent years (all forecasts are on a DVFA basis), but they are not as certain about earlier years in their database. While differences in basis between forecast and actual items would affect analyst bias, they do not affect our estimates of market discount rates. Differences in basis across analysts contaminate the consensus numbers used, but the estimated market discount rates are relatively insensitive to changes in the near-term forecasts used.

To select the month of analysis for each country, we followed the same logic as that for the U.S. analysis. December was the most popular fiscal year-end for all countries except for Japan, where it was March. We then identified the period after the fiscal year-end by which annual earnings are required to be disclosed. This period differs across countries (see Table 1 in Alford et al. (1993)): it is three months for Japan and the United States, four months for France, six months for Canada and the United Kingdom, and eight months for Germany. We selected the month following the reporting deadline as the "sure to be disclosed" month to collect forecasts for any given year.

To include a country-year in our sample, we required that the total market value of all firms in our sample exceed 35 percent of the market value of "primary stock holdings" for that country, as defined by Datastream. Although we used a low hurdle to ensure that our sample contained contiguous years for all countries, a substantially greater proportion of the Datastream Market Index than our minimum hurdle is represented for most country-years.

The equity-premium estimates using the abnormal earnings and dividend growth approaches as well as the prevailing risk-free rates for different country-year combinations with sufficient data are reported in Table III. The number of years with sufficient firms to represent the overall market was highest for Canada (all 14 years between 1985 and 1998), and lowest for Japan (8 years). As with the U.S. sample, we use a 50 percent aggregate

[16] The German financial analyst society, Deutsche Vereinigung für Finanzanalyse (DVFA), has developed a system used by analysts (and often by firms) to adjust reported earnings data to provide a measure that is closer to permanent or core earnings. The adjustment process uses both reported financial information as well as firms' internal records. GAAP refers to Generally Accepted Accounting Principles or the accounting rules under which financial statements are prepared in different domiciles.

Table III

Implied Equity Premium Using Abnormal Earnings and Dividend Growth Approaches ($k - r_f$ and $k^* - r_f$) for International Stocks (1985 to 1998)

The market is an aggregate of firms on the I/B/E/S Summary files with forecasts for years +1, +2, and a five-year earnings growth estimate (g_5) as of April each year, and actual earnings, dividends, number of shares outstanding, and prices as of the end of the prior full fiscal year (year 0). Book values of equity for year 0 (bv_0) are obtained from COMPUSTAT, Global Vantage, and Datastream. Forecasted earnings for years +3, +4, and +5 are determined by applying g_5, the forecasted 5-year growth rate, to year +2 forecasted earnings. All amounts are measured in local currencies. r_f is the 10-year government bond yield. The implied discount rate that satisfies the valuation relation in equation (5) below is k. Abnormal earnings (ae_t) equal reported earnings less a charge for the cost of equity (= beginning book value of equity $* k$). Assuming that 50% of earnings are retained allows the estimation of future book values from current book values and forecast earnings. The terminal value represents all abnormal earnings beyond year +5. Those abnormal earnings are assumed to grow at a constant rate, g_{ae}, which is assumed to equal the expected inflation rate, and is set equal to r_f less 3 percent. The expected rate of return on the market is also estimated using equation (1), and is labeled k^*. Equation (1) is derived from the dividend growth model, and dividend growth in perpetuity, g, is assumed to equal the five-year earnings growth rate, g_5.

$$p_0 = bv_0 + \frac{ae_1}{(1+k)} + \frac{ae_2}{(1+k)^2} + \frac{ae_3}{(1+k)^3} + \frac{ae_4}{(1+k)^4} + \frac{ae_5}{(1+k)^5} + \left[\frac{ae_5(1+g_{ae})}{(k-g_{ae})(1+k)^5}\right] \tag{5}$$

$$k^* = \frac{d_1}{p_0} + g \tag{1}$$

	Canada			France			Germany			Japan			U.K.		
Year	r_f	$k - r_f$	$k^* - r_f$	r_f	$k - r_f$	$k^* - r_f$	r_f	$k - r_f$	$k^* - r_f$	r_f	$k - r_f$	$k^* - r_f$	r_f	$k - r_f$	$k^* - r_f$
1985	10.50%	4.41%	7.45%												
1986	8.82%	2.93%	6.64%												
1987	9.16%	1.56%	4.53%	8.72%	2.06%	6.06%									
1988	9.66%	2.83%	4.67%	9.35%	4.00%	3.90%	6.78%	3.43%	4.59%						
1989	9.29%	3.08%	3.66%	8.76%	3.64%	6.11%	6.83%	3.87%	5.48%				10.16%	3.17%	7.24%
1990	10.69%	1.51%	2.97%	9.66%	3.04%	4.23%	8.99%	1.10%	3.23%				11.39%	2.57%	5.06%
1991	10.08%	0.75%	3.71%	8.81%	2.94%	4.41%	8.42%	1.03%	4.72%	6.72%	−0.95%	0.38%	10.49%	2.47%	7.27%
1992	8.18%	0.42%	6.36%	8.74%	2.26%	5.81%	7.89%	2.16%	5.03%	5.38%	−0.86%	−0.34%	9.12%	2.77%	8.69%
1993	7.32%	1.69%	6.59%	7.18%	2.31%	10.57%	6.14%	0.70%	4.19%	4.45%	−1.05%	4.36%	7.64%	3.29%	10.75%
1994	9.29%	1.65%	7.67%	6.82%	1.70%	8.24%	7.26%	1.30%	8.77%	4.24%	−1.04%	4.56%	8.63%	2.87%	8.50%
1995	7.93%	2.71%	6.77%	7.80%	2.06%	10.04%	6.70%	2.22%	9.84%	2.80%	1.12%	9.50%	8.44%	3.02%	8.59%
1996	7.69%	2.69%	6.89%	6.39%	2.38%	12.26%	6.41%	2.14%	8.40%	3.17%	0.79%	7.82%	7.92%	3.34%	8.43%
1997	6.35%	2.28%	7.10%	5.66%	2.28%	9.69%	5.68%	2.28%	11.56%	2.47%	1.65%	9.46%	7.02%	2.53%	7.81%
1998	5.36%	2.68%	7.44%	5.02%	2.53%	13.44%				1.65%	1.99%	10.89%	5.84%	2.09%	6.77%
Mean	8.59%	2.23%	5.89%	7.74%	2.60%	7.90%	7.11%	2.02%	6.58%	3.86%	0.21%	5.83%	8.66%	2.81%	7.91%
S.D.	1.55%	1.04%	1.62%	1.51%	0.68%	3.27%	1.04%	1.03%	2.82%	1.67%	1.31%	4.27%	1.68%	0.40%	1.49%

dividend payout ratio to generate future dividends and book values, and assume that abnormal earnings grow at the expected inflation rate, which is assumed to be three percent less than the prevailing risk-free rate. For the few years when r_f in Japan is below three percent, we set $g_{ae} = 0$.

The equity premium values based on the abnormal earnings approach $(k - r_f)$ generally lie between two and three percent, except for Japan, where the estimates are considerably lower (and even negative in the early 1990s). Finding that none of the almost 70 estimates of $k - r_f$ reported in Tables II and III are close to the Ibbotson estimate suggests strongly that that historical estimate is too high. In contrast, the equity premium estimates based on the dividend growth approach with dividends growing in perpetuity at the five-year earnings growth forecast (g_5) are considerably higher, similar to the pattern observed in the United States. The dividend growth estimates are very close to those reported in Khorana et al. (1997), which uses a similar approach and a similar sample.

Repeating the sensitivity analyses conducted on the United States (described in Section V) on these five markets produced similar conclusions. The abnormal earnings estimates generate projections that are consistent with experience, but the dividend growth estimates are biased upward and generate projections that are too optimistic because the five-year earnings growth forecast (g_5) is too high an estimate for dividend growth in perpetuity. The values of g_5 suggest mean real dividend growth rates in perpetuity that range between 6.09 percent for Canada and 8.25 percent for Japan. These real rates exceed historic real earnings growth rates, and are at least twice as high as the real GDP growth rates forecast for these countries.

The results observed for Japan are unusual and invite speculation. While our results suggest that the equity premium in Japan increased during the sample period, from about -1 percent in the early 1990s to 2 percent in the late 1990s, these results are also consistent with a stock market bubble that has gradually burst. That is, early in our sample period, prices were systematically higher than the fundamentals (represented by analysts' forecasts) would suggest, and have gradually declined to a level that is supported by analysts' forecasts. Note that our sample excludes the peak valuations in the late 1980s before the crash. Perhaps the implied equity premium in that period would be even more negative than the numbers we estimate for the early 1990s. Regardless of whether the poor performance of Japanese equities in the 1990s is due to correction of an earlier mispricing, it is useful to contrast the inferences from a historic approach with those from a forward-looking approach such as ours: the former would conclude that equity premia have fallen in Japan during the 1990s, whereas our approach suggests the opposite.

V. Sensitivity Analyses

This section summarizes our analysis of U.S. equity data designed to gauge the robustness of our conclusion that the equity premium is much lower than historic estimates. We begin by considering two relations for P/B and P/E ratios that allow us to check whether our projections under the dividend growth and abnormal earnings models are reasonable. Next, we document the extent of analyst optimism in our data. Finally, we consider the sensitivity of our risk premium estimates to the assumed abnormal earnings growth rate (g_{ae}).[17]

A. P/B Ratios and the Level of Future Profitability

The first relation we examine is that between the P/B ratio and future levels of profitability (e.g., Penman (1999)), where future profitability is the excess of the forecast market accounting rate of return (roe_t) over the required rate of return, k.

$$\frac{p_0}{bv_0} = 1 + \frac{roe_1 - k}{(1+k)} + \frac{roe_2 - k}{(1+k)^2}\left(\frac{bv_1}{bv_0}\right) + \frac{roe_3 - k}{(1+k)^3}\left(\frac{bv_2}{bv_0}\right) + \ldots, \tag{6}$$

where $roe_t = e_t/bv_{t-1}$ is the accounting return on equity in year t.

This relation indicates that the P/B ratio is explained by expected future profitability ($roe_t - k$).[18] Firms expected to earn an accounting rate of return on equity equal to the cost of capital should trade currently at book values ($p_0/bv_0 = 1$). Similarly, the P/B ratio expected in year +5 (p_5/bv_5), which is determined by the assumed growth in abnormal earnings after year +5 (g_{ae}), should be related to profitability beyond year +5. To investigate the validity of our assumed growth rates, we examine the profiles of future P/B ratios and profitability levels to check if they are reasonable and related to each other as predicted by equation (6). Future book values are generated by adding projected earnings and subtracting projected dividends (assuming a 50 percent payout) to the prior year's book value. Similarly, projected market values are obtained by growing the prior year's market value at the discount rate (k) less projected dividends.

Table IV provides data on current and projected values of P/B ratios and profitability. Current market and book values are reported in columns 1 and 2, and projected market and book values in year +5 are reported in columns

[17] We also examined Value Line data for the DOW 30 firms for two years: 1985 and 1995 (details in Claus and Thomas (1999a)). Value Line provides both dividend forecasts (over a four- or five-year horizon) and a projected price. This price is, in effect, a terminal value estimate, which obviates the need to assume dividend growth in perpetuity. Unfortunately, those risk premium estimates appear to be unreliable: The estimated discount rate is 20 percent (8.5 percent) for 1985 (1995). These results are consistent with Value Line believing that the DOW 30 firms are undervalued (overvalued) in 1985 (1995); that is, current price does not equal the present value of forecast dividends and projected prices. This view is supported by their recommendations for the proportion to be invested in equity: it was 100 percent through the 1980s, and declined through the 1990s (it is currently at 40 percent).

[18] The growth in book value terms in equation (6), bv_t/bv_0, which add a multiplicative effect, have been ignored in the discussion because of the built-in correlation with $roe_t - k$. Higher roe_t results in higher e_t, which in turn causes higher growth in bv_t because dividend payouts are held constant at 50 percent for all years.

Table IV

Price-to-Book Ratios (p_t/bv_t), Forecast Accounting Return on Equity (roe_t) and Expected Rates of Return (k) for U.S. Stocks (1985 to 1998)

To examine the validity of assumptions underlying k, which is the implied discount rate that satisfies the valuation relation in equation (5), current price-to-book ratios are compared with estimated future returns on equity (roe_t) to examine fit with equation (6) below. The market is an aggregate of firms on the I/B/E/S Summary files with forecasts for years +1, +2, and a five-year earnings growth estimate (g_5) as of April each year, and actual earnings, dividends, number of shares outstanding, and prices as of the end of the prior full fiscal year (year 0). Book values of equity for year 0 (bv_0) are obtained from COMPUSTAT. When missing, forecasted earnings for years +3, +4, and +5 are determined by applying g_5 to year +2 forecasted earnings. Assuming that 50 percent of earnings are retained allows the estimation of future book values from current book values and forecast earnings. Return on equity (roe_t) equals forecast earnings scaled by beginning book value of equity (bv_{t-1}). Market and book value amounts are in millions of dollars.

$$p_0 = bv_0 + \frac{ae_1}{(1+k)} + \frac{ae_2}{(1+k)^2} + \frac{ae_3}{(1+k)^3} + \frac{ae_4}{(1+k)^4} + \frac{ae_5}{(1+k)^5} + \left[\frac{ae_5(1+g_{ae})}{(k-g_{ae})(1+k)^5}\right] \quad (5)$$

$$\frac{p_0}{bv_0} = 1 + \frac{roe_1 - k}{(1+k)} + \frac{roe_2 - k}{(1+k)^2}\left(\frac{bv_1}{bv_0}\right) + \ldots \quad (6)$$

	Year 0 Equity Values		Year +5 Equity Values		Price/Book Ratio		Forecast Accounting Return on Equity						
Forecasts as of April	Market Value (p_0)	Book Value (bv_0)	Market Value (p_5)	Book Value (bv_5)	In Year 0 (p_0/bv_0)	In Year 5 (p_5/bv_5)	In Year 1 (roe_1)	In Year 2 (roe_2)	In Year 3 (roe_3)	In Year 4 (roe_4)	In Year 5 (roe_5)	In Year 6 (roe_6)	k from Eq. (5)
	1	2	3	4	5	6	7	8	9	10	11	12	13
1985	1,747,133	1,191,869	2,676,683	1,768,036	1.5	1.5	15%	16%	16%	17%	17%	17%	14.38%
1986	2,284,245	1,214,454	3,197,490	1,783,987	1.9	1.8	15%	16%	16%	17%	17%	17%	11.28%
1987	2,640,743	1,323,899	3,727,459	1,936,215	2.0	1.9	14%	16%	16%	16%	17%	17%	11.12%
1988	2,615,857	1,430,672	3,779,033	2,122,648	1.8	1.8	16%	16%	16%	17%	17%	17%	12.15%
1989	2,858,585	1,541,231	4,200,867	2,341,029	1.9	1.8	17%	17%	17%	18%	18%	18%	12.75%
1990	3,143,879	1,636,069	4,589,685	2,465,373	1.9	1.9	16%	17%	17%	18%	18%	18%	12.33%
1991	3,660,296	1,775,199	5,181,184	2,597,264	2.1	2.0	14%	16%	16%	17%	17%	17%	11.05%
1992	4,001,756	1,911,383	5,574,848	2,773,918	2.1	2.0	13%	15%	16%	16%	17%	17%	10.57%
1993	4,918,359	2,140,668	6,595,210	3,139,088	2.3	2.1	14%	16%	16%	17%	17%	17%	9.62%
1994	5,282,046	2,168,446	7,336,322	3,301,664	2.4	2.2	16%	17%	18%	18%	19%	18%	10.47%
1995	6,289,760	2,670,725	8,837,148	4,132,682	2.4	2.1	17%	18%	18%	19%	19%	19%	11.03%
1996	8,207,274	3,182,952	11,206,787	4,853,189	2.6	2.3	16%	17%	18%	18%	19%	18%	9.96%
1997	10,198,036	3,679,110	14,103,523	5,708,609	2.8	2.5	17%	18%	18%	19%	20%	19%	10.12%
1998	12,908,495	3,412,303	16,838,377	5,378,478	3.8	3.1	17%	18%	19%	20%	21%	20%	8.15%
Mean					2.2	2.1	15%	17%	17%	18%	18%	18%	11.04%

3 and 4. These values are used to generate current and year +5 P/B ratios, reported in columns 5 and 6. Columns 7 through 12 contain the forecasted accounting rate of return on equity for years 1 to 6, which can be compared with the estimated market discount rate, k, reported in column 13, to obtain forecasted profitability.

The current P/B ratio has been greater than 1 in every year in the sample period, and has increased steadily over time, from 1.5 in 1985 to 3.8 in 1998. Consistent with equation (6), all forecasted *roe* values for years 1 through 6 in Table IV exceed the corresponding values of k. Increases in the P/B ratio over the sample period are mirrored by corresponding increases in forecast profitability ($roe_t - k$) in years +1 through +5 as well as forecast profitability in the posthorizon period (after year +5), as measured by the implied price-to-book ratio in year +5. Finally, the tendency for P/B ratios to revert gradually over the horizon toward one (indicated by the year +5 values in column 6 being smaller than the year 0 values in column 5) is consistent with intuition (e.g., Nissim and Penman (1999)).

We also extended our investigation to years beyond year +5 for the assumptions underlying the abnormal earnings estimates, and find that the pattern of projections for P/B and *roe* remain reasonable. In contrast, those projections for the assumptions underlying the dividend growth model estimates suggest that the underlying growth rates are unreasonably high. To provide an illustrative example of those results, we contrast in Figure 2 the patterns for future *roe* and P/B that are projected for the dividend growth and abnormal earnings approaches for 1991. The *roe* levels are marked off on the left scale, and P/B ratios are shown on the right scale. Recall that the market discount rates estimated for the abnormal earnings and dividend growth approaches are 11.05 percent (k) and 15.16 percent (k^*) and the corresponding terminal growth rates for abnormal earnings and dividends are 5.04 percent and 12.12 percent.

The projections for the abnormal earnings method (indicated by bold lines) continue to remain reasonable. The P/B ratio always exceeds one, but it trends down over time. Consistent with P/B exceeding one, the *roe* is always above the 11.05 percent cost of capital, and trends toward it after year +5. Note that the optimistic analyst forecasts cause *roe* projections to climb for years +1 through +5, but the subsequent decline in *roe* is because the profitability growth implied by g_{ae} (our assumed growth in abnormal earnings past year +5) is lower than that implied by g_5.

The results for the dividend growth approach illustrate the benefits of using projected accounting ratios to validate assumed growth rates. The profitability (*roe*) is actually below the cost of equity of 15.16 percent (k^*), for the first three years, even though the P/B ratio is greater than one. Thereafter, the profitability keeps increasing, to a level above 20 percent by year +15. Both the high level of profitability and its increasing trend are not easily justified, especially when they are observed repeatedly for every year in our sample. Similarly, the increasing pattern for P/B, which is projected to increase from about two to about three by year +15, is hard to justify.

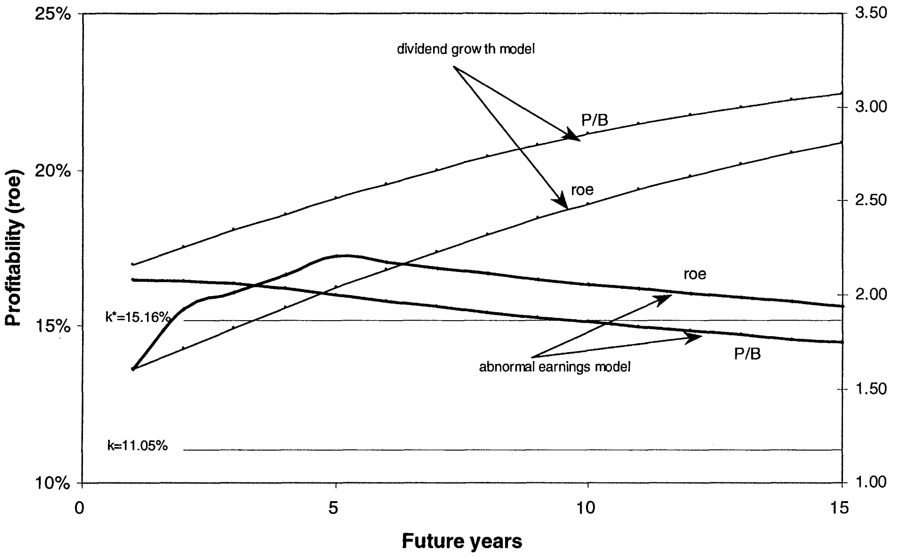


Figure 2. Pattern of future price-to-book (P/B) ratios and profitability, measured as excess of accounting return on equity (*roe*) over estimated discount rates (k^* and k), for dividend growth and abnormal earnings approaches for U.S. stocks as of April, 1991. For the dividend growth model described by equation (1) in Table II, dividends are assumed to grow at the consensus five-year earnings growth rate of 12.12 percent, and future *roe* is compared with the estimated market discount rate of 15.16 percent (k^*). For the abnormal earnings model described by equation (5) in Table II, abnormal earnings are assumed to grow at an anticipated inflation rate of 5.04 percent, and *roe* is compared with the estimated market discount rate of 11.05 percent (k). Projected P/B ratios are shown for both models.

These projections are, however, consistent with an estimated discount rate that is too high. Since near-term analysts' forecasts of profitability are below this discount rate, future levels of profitability have to be unreasonably high to compensate.

B. P/E Ratios and Forecast Growth in Profitability

The second relation we use to check the validity of our assumptions regarding g_{ae} is the price–earnings ratio, described by equation (7) (see derivation in Claus and Thomas, 1999a). Price–earnings ratios are a function of the present value of future changes in abnormal earnings, multiplied by a capitalization factor ($= 1/k$).

$$\frac{p_0}{e_1} = \frac{1}{k}\left[1 + \frac{\Delta ae_2}{e_1(1+k)} + \frac{\Delta ae_3}{e_1(1+k)} + \dots\right], \tag{7}$$

where $\Delta ae_t = ae_t - ae_{t-1}$ is the change in expected abnormal earnings over the prior year.

The price–earnings ratio on the left-hand side deviates slightly from the traditional representation in the sense that it is a "forward" price–earnings ratio, based on expected earnings for the upcoming year, rather than a "trailing" price–earnings ratio (p_0/e_0), which is based on earnings over the year just concluded. The relation between future earnings growth and forward price–earnings ratios is simpler than that for trailing price–earnings ratios.[19] Therefore, we use only the forward price–earnings ratio here and refer to it simply as the P/E ratio.

The results reported in Table V describe P/E ratios and growth in abnormal earnings derived from analysts' forecasts for the market. The first four columns provide market values and the corresponding upcoming expected earnings for year 0 and year +5. These numbers are used to generate the current and year +5 P/E ratios reported in columns 5 and 6, which can be compared to the values of $1/k$ reported in column 18.[20] According to equation (7), absent growth in abnormal earnings, the P/E ratio should be equal to $1/k$, and the P/E ratio should be greater (less) than $1/k$ for positive (negative) expected growth in abnormal earnings. Forecast growth rates in abnormal earnings for years +2 through +6 are reported in columns 7 through 11. To maintain equivalence with the terms in equation (7), growth in abnormal earnings is scaled by earnings expected for year +1 (e_1) and then discounted.

To understand the relations among the numbers in the different columns, consider the row corresponding to 1991. The market P/E ratio of 15.1 is higher than the inverse of the discount rate ($1/k = 9.0$). That difference of 6.1 is represented by the sum of the present value of the abnormal earnings growth terms in future years, scaled by e_1 (this sum needs to be multiplied by $1/k$ as shown in equation (7)). These growth terms decline from 13 percent in year 2 to 2 percent in year 6, and continue to decline thereafter. By year +5, the market P/E is expected to fall (to 11.7), since some of the growth in abnormal earnings (represented by the amounts in columns 7 through 11) is expected to have already occurred by then. Turning to the other sample years, the P/E ratios in year 0 (column 5) have generally increased through the sample period, and so have the values of $1/k$. Consistent with P/E ratios exceeding $1/k$ in every year, abnormal earnings are forecast to exhibit positive growth for all cells in columns 7 to 11. Also, the P/E ratios in year +5 are forecast to decline, relative to the corresponding year 0 P/E values, because of the value represented by the amounts in columns 7 to 11.

[19] Since the numerator of the P/E ratio is an ex-dividend price (p_0), the payment of a large dividend (d_0) would reduce p_0 without affecting trailing earnings (e_0), thereby destroying the relation between p_0 and e_0. This complication does not arise when expected earnings for the upcoming period (e_1) is used instead of e_0.

[20] If the numbers in Table V appear to be not as high as the trailing P/E ratios commonly reported in the popular press, note that forward P/E ratios are generally smaller than trailing P/E ratios for the following reasons. First, next year's earnings are greater than current earnings because of earnings growth. Second, current earnings contain one-time or transitory components that are on average negative, whereas forecast earnings focus on core or continuing earnings.

For purposes of comparison with other work, we also report in columns 12 through 17 of Table V the growth in forecast earnings (as opposed to growth in abnormal earnings) for years +1 through +6. Forecasted growth in earnings declines over the horizon, similar to the pattern exhibited by growth in abnormal earnings. Note the similarity in the pattern of earnings growth for all years in the sample period: the magnitudes of earnings growth estimates appear to settle at around 12 percent by year +5, before dropping sharply to values around 7 percent in the posthorizon period (year +6). Again, this decline occurs because the earnings growth implied by g_{ae} (our assumed growth in abnormal earnings past year +5) is lower than g_5.

The results in Table V confirm the predictions derived from equation (7) as well as the intuitive links drawn in the literature. As with the results for P/B ratios, the trends for P/E ratios and growth in abnormal earnings exhibit no apparent discrepancies that might suggest that the assumptions underlying our abnormal earnings model are unreasonable.

C. Bias in Analyst Forecasts

We considered a variety of biases that may exist in the I/B/E/S forecasts, but found only the well-known optimism bias to be noteworthy (details provided in Claus and Thomas (1999a)).[21] We compute the forecast error for each firm in our sample, representing the median consensus forecast as of April less actual earnings, for different forecast horizons (year +1, +2, . . . +5) for each year between 1985 and 1997. Table VI contains the median forecast errors (across all firms in the sample for each year), scaled by share price. In general, forecasted earnings exceed actual earnings, and the extent of optimism increases with the horizon.[22] There is, however, a gradual reduction in optimism toward the end of the sample period.

Since the forecast errors in Table VI are scaled by price, comparing the magnitudes of the median forecast errors with the inverse of the trailing P/E ratios (or E/P ratios) is similar to a comparison of forecast errors with earnings levels. While the trailing E/P ratios for our sample vary between 5 and 9 percent, the forecast errors in Table VI vary between values that are in the neighborhood of 0.5 percent for year +1 to around 3 percent in year +5. Comparing the magnitudes of year +5 forecast errors with the implied E/P ratios indicates that forecasted earnings exceed actual earnings by as

[21] I/B/E/S removes one-time items (typically negative) from reported earnings. That is, the level of optimism would have been even higher if we had used reported numbers instead of actual earnings according to I/B/E/S.

[22] In addition to increasing with forecast horizon, the optimism bias is greater for certain years where earnings were depressed temporarily. The higher than average dividend payouts observed in Table I for 1987 and 1992 indicate temporarily depressed earnings in those years, and the forecast errors are also higher than average for those years. For example, the two largest median year +2 forecast errors are 1.86 and 1.81 percent, and they correspond to two-year out forecasts made in 1985 and 1990.

Table V
Forward Price-to-Earnings Ratios (p_t/e_{t+1}) and Growth in Forecast Abnormal Earnings and Earnings for U.S. Stocks (1985 to 1998)

To examine the validity of assumptions underlying k, which is the implied discount rate that satisfies the valuation relation in equation (5), current and forecast forward price-to-earnings ratios are compared with growth in forecast abnormal earnings to examine fit with equation (7) below. The market is an aggregate of firms on the I/B/E/S Summary files with forecasts for years +1, +2, and a five-year earnings growth estimate (g_5) as of April each year, and actual earnings, dividends, number of shares outstanding, and prices as of the end of the prior full fiscal year (year 0). Book values of equity for year 0 (bv_0) are obtained from COMPUSTAT. Abnormal earnings (ae_t) equal reported earnings less a charge for the cost of equity (= beginning book value of equity $*$ k). Future market values are projected for each year by multiplying beginning market values by $(1 + k)$ and subtracting dividends. When missing, forecasted earnings for years +3, +4, and +5 are determined by applying g_5 to year +2 forecasted earnings. Assuming that 50 percent of earnings are retained allows the estimation of future book values from current book values and forecast earnings. Market equity values and earnings amounts are in millions of dollars.

$$p_0 = bv_0 + \frac{ae_1}{(1+k)} + \frac{ae_2}{(1+k)^2} + \frac{ae_3}{(1+k)^3} + \frac{ae_4}{(1+k)^4} + \frac{ae_5}{(1+k)^5} + \left[\frac{ae_5(1+g_{ae})}{(k-g_{ae})(1+k)^5}\right] \tag{5}$$

$$\frac{p_0}{e_1} = \frac{1}{k}\left[1 + \frac{\Delta ae_2}{e_1(1+k)} + \frac{\Delta ae_3}{e_1(1+k)^2} + \dots\right] \tag{7}$$

Forecasts as of April	Year 0 Values: Market Value (p_0)	Year 0 Values: Earnings (e_1)	Year +5 Values: Market Value (p_5)	Year +5 Values: Earnings (e_6)	Forward P/E Ratio: In Year 0 (p_0/e_1)	Forward P/E Ratio: In Year 5 (p_5/e_6)	PV of ae Growth (Δae_t), Scaled by e_1: +2	+3	+4	+5	+6	Growth in Forecast Earnings: +1	+2	+3	+4	+5	+6	$1/k$ from Eq. (5)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1985	1,747,133	180,945	2,676,683	308,308	9.7	8.7	5%	3%	3%	3%	1%	17%	13%	11%	11%	12%	9%	7.0
1986	2,284,245	178,024	3,197,490	299,896	12.8	10.7	7%	4%	5%	5%	1%	15%	14%	11%	11%	11%	7%	8.9
1987	2,640,743	186,319	3,727,459	324,573	14.2	11.5	10%	5%	5%	5%	1%	27%	18%	11%	11%	11%	7%	9.0
1988	2,615,857	222,497	3,781,766	364,583	11.8	10.4	4%	4%	4%	4%	1%	33%	11%	11%	11%	11%	8%	8.2
1989	2,858,585	261,278	4,200,867	420,673	10.9	10.0	2%	3%	3%	4%	1%	14%	9%	11%	11%	11%	8%	7.8
1990	3,143,879	257,657	4,589,685	442,911	12.2	10.4	7%	4%	4%	4%	1%	13%	15%	11%	12%	12%	8%	8.1
1991	3,660,296	241,760	5,181,184	442,291	15.1	11.7	13%	5%	6%	6%	2%	11%	22%	12%	12%	12%	7%	9.0
1992	4,001,756	252,109	5,574,848	463,780	15.9	12.0	14%	6%	6%	6%	2%	25%	22%	12%	12%	12%	7%	9.5
1993	4,918,359	295,862	6,595,210	531,812	16.6	12.4	13%	6%	7%	7%	1%	19%	20%	12%	12%	12%	6%	10.4
1994	5,282,046	339,694	7,174,214	604,559	15.5	11.9	11%	6%	6%	7%	1%	17%	19%	12%	12%	12%	6%	10.0
1995	6,289,760	444,593	8,837,148	783,736	14.1	11.3	9%	5%	6%	6%	2%	22%	17%	12%	12%	12%	7%	9.1
1996	8,207,274	512,921	11,206,787	893,185	16.0	12.5	8%	6%	7%	7%	2%	15%	15%	12%	13%	13%	7%	10.0
1997	10,198,036	614,932	14,103,523	1,100,714	16.6	12.8	8%	7%	7%	8%	2%	19%	16%	11%	12%	12%	7%	9.9
1998	12,908,495	577,297	16,838,377	1,069,786	22.4	15.7	12%	9%	10%	11%	2%	19%	16%	11%	12%	12%	7%	12.3
Mean					14.6	11.6	9%	5%	6%	6%	1%	19%	16%	11%	12%	12%	7%	9.2

Table VI

Optimism Bias in I/B/E/S Forecasts for U.S. Stocks: Median Forecast Errors for Forecasts Made Between 1985 and 1997

The following table represents the median of all forecast errors scaled by share price for each year examined. The forecast error is calculated for each firm as of April each year, and equals the median consensus forecasted earnings per share minus the actual earnings per share, scaled by price. The year when the forecasts were made is listed in the first row, while the first column lists the horizon of that forecast. For each year and horizon combination, we report the median forecast error and the number of firms in the sample. To interpret the Table, consider the values of 0.78 percent and 1,680 reported for the +1/1985 combination., in the top left-hand corner of the table. This means that the median value of the difference between the forecasted and actual earnings for 1986 was 0.78 percent of price, and that sample consisted of 1,680 firms with available forecast errors. The results confirm that analyst forecasts are systematically positively biased and that this bias increases with the forecast horizon; however, the extent of any such bias has been declining steadily over time.

		Year Forecast Was Made													
		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	Mean
Forecast	Median	0.78%	0.65%	0.37%	0.07%	0.44%	0.58%	0.39%	0.17%	0.15%	0.03%	0.04%	0.00%	0.00%	0.28%
Year +1	Obs.	1,680	1,707	1,878	1,815	1,868	1,932	1,959	2,176	2,492	2,710	2,895	3,261	3,462	
Forecast	Median	2.05%	1.40%	0.79%	0.99%	1.74%	1.88%	1.21%	0.87%	0.58%	0.34%	0.32%	0.27%	—	1.04%
Year +2	Obs.	1,545	1,572	1,732	1,701	1,757	1,815	1,896	2,084	2,287	2,594	2,694	2,852	—	
Forecast	Median	2.84%	0.99%	1.44%	2.22%	2.78%	2.39%	1.50%	0.95%	0.63%	0.54%	0.45%	—	—	1.52%
Year +3	Obs.	1,406	1,449	1,596	1,576	1,634	1,744	1,826	1,936	2,159	2,396	2,346	—	—	
Forecast	Median	2.63%	2.04%	2.80%	3.19%	3.17%	2.83%	1.54%	0.91%	0.77%	0.60%	—	—	—	2.05%
Year +4	Obs.	1,285	1,344	1,492	1,474	1,586	1,696	1,724	1,825	2,024	2,132	—	—	—	
Forecast	Median	3.54%	3.44%	3.86%	3.59%	3.43%	2.91%	1.36%	0.94%	0.74%	—	—	—	—	2.65%
Year +5	Obs.	1,201	1,260	1,411	1,432	1,528	1,621	1,618	1,704	1,815	—	—	—	—	

much as 50 percent at that horizon. These results suggest that our equity premium estimates are biased upward because we do not adjust for the considerable optimism in earnings forecasts for years +1 to +5. They also suggest that we are justified in dropping assumed growth rates for earnings past year +5 (column 17 versus column 16 in Table V).

D. Impact of Variation in the Assumed Growth Rate in Abnormal Earnings Beyond Year +5 (g_{ae})

We begin by considering two alternative cases for g_{ae}: three percent less and three percent more than our base case, where g_{ae} is assumed to equal the expected inflation rate. As mentioned in the Appendix, our base growth rate of $g_{ae} = r_f - 3\%$ is higher than any rate assumed in the prior abnormal earnings literature. Adding another three percent to the growth rate, which would require rents to grow at a three percent real rate in perpetuity, raises the level of optimism further. Dropping three percent from the base case, in the lower growth scenario, would be equivalent to assuming a very low nominal growth rate in abnormal earnings, and would be only slightly more optimistic than the assumptions in much of the prior abnormal earnings literature.

For the higher (lower) growth rate scenario, corresponding to $g_{ae} = r_f$ ($g_{ae} = r_f - 6\%$), the average risk premium over the 14-year sample period increases (decreases) to a mean of 4.66 (2.18), from a mean of 3.40 percent for the base case. Even for the high growth rate in abnormal earnings, the increase in the estimated risk premium is modest, and leaves it substantially below the traditional estimates of the risk premium. While increasing (decreasing) the growth rate increases (decreases) the terminal value, it also reduces (increases) the present value of that terminal value because of the higher (lower) discount rate it engenders.

We also considered a synthetic market portfolio each year constructed to have no expected future abnormal earnings, to avoid the need for an assumed abnormal earnings growth rate beyond year +5. As described in equation (6), portfolios with P/B = 1 should exhibit no abnormal earnings; that is, the roe_t should on average equal k for this synthetic market. The last term in equation (5), representing the terminal value of abnormal earnings beyond year +5, is set to zero and the estimates for k obtained iteratively each year. The mean estimate for $k - r_f$ from this synthetic market is 2.20 percent, which is slightly lower than the mean risk premium of 3.40 percent in Table II. Note that a lower discount rate is not expected for the synthetic market, since it has a beta close to one each year and has a lower P/B than the market. (Low P/B firms are expected to generate higher returns (e.g., Gebhardt, Lee, and Swaminathan (forthcoming).) The higher discount rates observed for the assumptions underlying our abnormal earnings model support our view that the analyst forecasts we use and our assumption that the terminal growth in abnormal earnings equals expected inflation ($g_{ae} = r_f - 3\%$) are both optimistic.

VI. Conclusion

Barring some notable exceptions (e.g, Siegel (1992 and 1998), Blanchard (1993), Malkiel (1996), and Cornell (1999)), academic financial economists generally accept that the equity premium is around eight percent, based on the performance of the U.S. market since 1926. We claim that these estimates are too high for the post-1985 period that we examine, and the equity premium is probably no more than three percent. Our claim is based on estimates of the equity premium obtained for the six largest equity markets, derived by subtracting the 10-year risk-free rate from the discount rate that equates current prices to forecasted future flows (derived from I/B/E/S earnings forecasts). Growth rates in perpetuity for dividends and abnormal earnings need to be much higher than is plausible to justify equity premium estimates of about eight percent. Not only are such growth rates substantially in excess of any reasonable forecasts of aggregate growth (e.g., GDP), the projected streams for various indicators, such as price-to-book and price-to-earnings ratios, are also internally contradictory and inconsistent with intuition and past experience.

We agree that the weight of the evidence provided by the historical performance of U.S. stock markets since 1926 is considerable. Yet there are reasons to believe that this performance exceeded expectations, because of potential declines in the equity premium, good luck, and survivor bias. While projecting dividends to grow at earnings growth rates forecast by analysts provides equity premium estimates as high as eight percent, we show that those growth forecasts exhibit substantial optimism bias and need to be adjusted downward. In addition to our results, theory-based work, historical evidence from other periods and other markets, and surveys of institutional investors all suggest that the equity premium is much lower than eight percent. Overall, we believe that an eight percent equity premium is not supported by an analysis that compares current market prices with reasonable expectations of future flows for the markets and years that we examine.

Appendix: Assumed Growth Rates in Perpetuity for Dividends (g) and Abnormal Earnings (g_{ae})

While the conceptual definition of g is clear—it is the dividend growth rate that can be sustained in perpetuity, given current capital and future earnings[23]—determining this rate from fundamentals is not easy. To illustrate, take two firms that are similar in every way, except that they have announced different dividend policies in the current period, which results in a higher expected forward dividend yield (d_1/p_0) for one firm than the other, say 7 percent and 1 percent. What can be said about g for the two firms?

[23] Assuming too high a rate would cause the capital to be depleted in some future period, and assuming too low a rate would cause the capital to grow "too fast."

Examination of equation (1) indicates that g for the low dividend yield firm must be 6 percent higher than g for the higher dividend yield firm, assuming they both have the same discount rate (k^*). If k^* equals 10 percent, for example, the value of g for the two firms must be 3 percent and 9 percent. These two values of g are substantially different from each other, even though the two firms are not.

In addition to being a hypothetical rate, g need not be related to historic or forecasted near-term growth rates for earnings or dividends. Dividend payout ratios can change over time because of changes in the investment opportunity set available and the relative attractiveness of cash dividends versus stock buybacks. Since changes in dividend payout affect the dividend yield, which in turn affects g, historic growth rates may not be relevant for g. Also, if dividend policies are likely to change over time, g need not be related to g_5 (the growth rate forecast for earnings over the next five years), a rate that is frequently used to proxy for g. Various scenarios can be constructed for the two firms in the example above to obtain similar historic and/or near-term forecast growth rates and yet have substantially different values for g.

Despite the difficulties noted above, both historic and forecast rates for aggregate dividends, earnings, and other macroeconomic measures (such as GDP) have been used as proxies for g. We note that these proxies create additional error. First, it is important to hold the unit of investment constant through the period where growth is measured. In particular, any growth created at the aggregate level by the issuance/retirement of equity since the beginning of the period should be ignored. Second, profits from all activities conducted outside the publicly traded corporate sector that are included in the macroeconomic measures should be deleted, and all overseas profits relating to this sector that are excluded from some macroeconomic measures should be included.

To control for the unit of investment problem, we use forecasted growth in per-share earnings rather than aggregate earnings, and to mitigate the problems associated with identifying g, we focus on growth in rents (abnormal earnings), g_{ae}, rather than dividends. To understand the benefits of switching to g_{ae}, it is important to describe some features of abnormal earnings. Expected abnormal earnings would equal zero if book values of equity reflected market values.[24] If book values measure input costs fairly, but do not include the portion of market values that represent economic rents (not yet earned), abnormal earnings would reflect those rents. However, the magnitude of such rents at the aggregate market level is likely to be small, and any rents that emerge are likely to be dissipated over time for the usual reasons (antitrust actions, global competition, etc.). As a result, much of the

[24] That is, if market prices are efficient and book values are marked to market values each period, market (book) values are expected to adjust each period so that no future abnormal returns (abnormal earnings) are expected.

earlier literature using the abnormal earnings approach has assumed zero growth in abnormal earnings past the "horizon" date.[25]

Returning to the two-firm example, shifting the focus from growth in dividends to growth in rents removes much of the confusion caused by transitory changes in dividend payouts and dividend yields: these factors should have no impact on growth in rents, since the level of and growth in rents are determined by economic factors such as monopoly power. That is, even though the two firms have different forecasted earnings and dividends, the forecasted abnormal earnings and growth in abnormal earnings should be identical.

We believe, however, that the popular assumption of zero growth in abnormal earnings may be too pessimistic because accounting statements are conservative and understate input costs: assets (liabilities) tend to be understated (overstated) on average. For example, many investments (such as research and development, advertising, and purchased intangibles) are written off too rapidly in many domiciles. As a result, abnormal earnings tend to be positive, even in the absence of economic rents. Growth in abnormal earnings under conservative accounting is best understood by examining the behavior of the excess of *roe* (the accounting rate of return on the book value of equity) over k (the discount rate). Simulations and theoretical analyses (e.g., Zhang (2000)) of the steady-state behavior of the accounting rate of return under conservative accounting suggest two important determinants: the long-term growth in investment and the degree of accounting conservatism. These analyses also suggest that *roe* approaches k, but remains above it in the long-term.

Even though a decline in the excess of *roe* over k should cause the magnitude of abnormal earnings to fall over time, a countervailing factor is the growth in investment, which increases the base on which abnormal earnings are generated. We assume as a first approximation that the latter effect is greater than the former, and that abnormal earnings increase in perpetuity at the expected inflation rate. Since we recognize that this assumption is an approximation, we elected to err on the side of choosing too high a growth rate to ensure that our equity premium estimates are not biased downward. Also, we conduct sensitivity analyses to identify the impact on our equity premium estimates of varying the assumed growth rate within a reasonable range.

REFERENCES

Abarbanell, Jeffrey and Victor Bernard, 1999, Is the U.S. stock market myopic? *Journal of Accounting Research* 38, 221–242.

Abel, Andrew B., 1999, Risk premia and term premia in general equilibrium, *Journal of Monetary Economics* 43, 3–33.

Alford, Andrew, Jennifer Jones, Richard Leftwich, and Mark Zmijewski, 1993, The relative informativeness of accounting disclosures in different countries, *Journal of Accounting Research* 31, 183–223.

[25] That is, abnormal earnings persist, but show no growth. Some papers are even more conservative, and have assumed that abnormal earnings drop to zero past the horizon date.

Ang, Andrew and Jun Liu, 2000. A generalized affine earnings model, Working paper, Columbia University.

Bagnoli, Mark, Messod D. Beneish, and Susan G. Watts, 1999, Whisper forecasts of quarterly earnings per share, *Journal of Accounting and Economics* 28, 27–50.

Benore, Charles, 1983, A survey of investor attitudes toward the electrical power industry, Paine Webber Mitchell Hutchins, Inc., New York.

Blanchard, Oliver J., 1993, Movements in the equity premium, *Brookings Papers on Economic Activity* 2, 75–138.

Bogle, John C., 1999, *Common Sense on Mutual Funds* (John Wiley & Sons, New York).

Botosan, Christine A., 1997, Disclosure level and the cost of equity capital, *Accounting Review* 72, 323–350.

Brigham, Eugene F., Dilip K. Shome, and Steve R. Vinson, 1985, Risk premium approach to measuring a utility's cost of equity, *Financial Management* 14, 33–45.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance* 50, 853–873.

Burr, Barry B., 1998, 20%+ again? Keep dreaming, *Pensions and Investments* 26, 33–34.

Campbell, John Y., 1991, A variance decomposition for stock returns, *The Economic Journal* 101 (405), 157–179.

Campbell, John Y. and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Claus, James J. and Jacob K. Thomas, 1999a, The equity risk premium is lower than you think it is: Empirical estimates from a new approach, Working paper, Columbia University.

Claus, James J. and Jacob K. Thomas, 1999b, Measuring the equity premium using earnings forecasts: An international analysis, Working paper, Columbia University.

Cochrane, John H., 1997, Where is the market going? Uncertain facts and novel theories, *Economic Perspectives* 21, 3–37.

Cornell, Bradford, 1999, *The Equity Risk Premium: The Long-Run Future of the Stock Market* (John Wiley & Sons, New York).

Edwards, Edgar O. and Philip W. Bell, 1961, *The Theory and Measurement of Business Income* (University of California Press, Berkeley, CA).

Fama, Eugene F. and Kenneth R. French, 2000, The equity premium, Working paper, University of Chicago.

Frankel, Richard and Charles M. C. Lee, 1998a, Accounting valuation, market expectation, and cross-sectional stock returns, *Journal of Accounting and Economics* 25, 283–319.

Frankel, Richard and Charles M. C. Lee, 1998b, Accounting diversity and international valuation, Working paper, University of Michigan.

Gebhardt, William, Charles M. C. Lee and Bhaskaran Swaminathan, Toward an ex ante cost of capital, *Journal of Accounting Research* forthcoming.

Glassman, James K. and Kevin A. Hassett, 1998, Are stocks overvalued? Not a chance, *Wall Street Journal*, March 30.

Gordon, Myron, 1962, *The investment, financing and valuation of the corporation* (Irwin, Homewood, IL).

Ibbotson Associates, Inc., 1999, Stocks, Bonds, Bills and Inflation, *Yearbook*, Chicago, IL.

Jorion, Philipe and William N. Goetzmann, 1999, Global stock markets in the twentieth century, *Journal of Finance* 54, 953–980.

Khorana, Ajay, R. Charles Moyer, and Ajay Patel, 1997, An examination of equity risk premium forecasts in G-6 countries, Working paper, Babcock Graduate School of Management, Wake Forest University.

Kocherlakota, Naryana, 1996, The equity premium: It's still a puzzle, *Journal of Economic Literature* 34, 42–71.

Liu, Jing and Jacob K. Thomas, 2000, Stock returns and accounting earnings, *Journal of Accounting Research* 38, 71–101.

Malkiel, Burton G., 1996, The equity premium puzzle, Keynote speech at 1996 Ibbotson Associates equity risk premium conference, Chicago, IL.

Matsumoto, Dawn A., 1999, Management's Incentives to Guide Analysts' Forecasts, Working paper, Harvard Business School.

Mayfield, E. Scott, 1999, Estimating the market risk premium, Working paper, Harvard Business School.

Mehra, Rajnish and Edward C. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics* 15, 145–161.

Moyer, R. Charles, and Ajay Patel, 1997, The equity market risk premium: a critical look at alternative ex ante estimates, Working paper, Babcock Graduate School of Management, Wake Forest University.

Myers, James N., 1999, Conservative accounting and finite firm life: Why residual income valuation estimates understate stock price, Working paper, University of Washington.

Nissim, Doron and Stephen H. Penman, 1999, Ratio analysis and equity valuation, Working paper, Columbia Business School.

Penman, Stephen H., 1999, *Financial Statement Analysis and Security Valuation* (Irwin/McGraw Hill, New York).

Preinreich, Gabriel A. D., 1938, Annual survey of economic theory: the theory of depreciation, *Econometrica* 6, 219–241.

Samuelson, Paul, 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review* 6, 41–49.

Siegel, Jeremy J., 1992, The equity premium: Stock and bond returns since 1802, *Financial Analysts Journal* January/February, 28–38.

Siegel, Jeremy J., 1998, *Stocks for the long run,* 2nd edition (McGraw-Hill, New York).

Siegel, Jeremy J., and Richard H. Thaler, 1997, The equity risk premium puzzle, *Journal of Economic Perspectives* 11, 191–200.

Stulz, René M., 1999, Globalization, corporate finance, and the cost of capital, *Journal of Applied Corporate Finance* 12(3), 8–25.

Vuolteenaho, Tuomo, 1999, Understanding the aggregate book-to-market ratio, Working paper, University of Chicago.

Welch, Ivo, 1999, Views of financial economists on the equity premium and professional controversies, Working paper, UCLA.

Wien, Byron R., 1998, U.S. strategy: Risk premium—R.I.P. Morgan Stanley Dean Witter.

Williams, John. B., 1938, *The Theory of Investment Value* (Harvard University Press, Cambridge, MA).

Zhang, Xiao-jun, 2000, Conservative accounting and equity valuation, *Journal of Accounting and Economics* 29, 125–149.

Rational Asset Prices

George M. Constantinides

The Journal of Finance, Vol. 57, No. 4, Papers and Proceedings of the Sixty-Second Annual Meeting of the American Finance Association, Atlanta, Georgia, January 4-6, 2002. (Aug., 2002), pp. vi+1567-1591.

Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28200208%2957%3A4%3Cvi%3ARAP%3E2.0.CO%3B2-E

The Journal of Finance is currently published by American Finance Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/afina.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Fri Oct 6 12:47:57 2006

THE JOURNAL OF FINANCE • VOL. LVII, NO. 4 • AUGUST 2002

Rational Asset Prices

GEORGE M. CONSTANTINIDES*

ABSTRACT

The mean, covariability, and predictability of the return of different classes of financial assets challenge the rational economic model for an explanation. The unconditional mean aggregate equity premium is almost seven percent per year and remains high after adjusting downwards the sample mean premium by introducing prior beliefs about the stationarity of the price–dividend ratio and the (non)-forecastability of the long-term dividend growth and price–dividend ratio. Recognition that idiosyncratic income shocks are uninsurable and concentrated in recessions contributes toward an explanation. Also borrowing constraints over the investors' life cycle that shift the stock market risk to the saving middle-aged consumers contribute toward an explanation.

A central theme in finance and economics is the pursuit of a *unified* theory of the *rate of return* across different classes of financial assets. In particular, we are interested in the *mean*, *covariability*, and *predictability* of the return of financial assets. At the macro level, we study the short-term risk-free rate, the term premium of long-term bonds over the risk-free rate, and the aggregate equity premium of the stock market over the risk-free rate. At the micro level, we study the premium of individual stock returns and of classes of stocks, such as the small-capitalization versus large-capitalization stocks, the "value" versus "growth" stocks, and the past losing versus winning stocks.

The neoclassical rational economic model is a *unified* model that views these premia as the reward to risk-averse investors that *process information rationally* and *have unambiguously defined preferences over consumption* that typically (but not necessarily) belong to the von Neumann–Morgenstern class. Naturally, the theory allows for market incompleteness, market imperfections, informational asymmetries, and learning. The theory also allows for differences among assets for liquidity, transaction costs, tax status, and other institutional factors.

The cause of much anxiety over the last quarter of a century is evidence interpreted as failure of the rational economic paradigm to explain the price level and the rate of return of financial assets both at the macro and micro

*University of Chicago and NBER. I thank John Campbell, Gene Fama, Chris Geczy, Lars Hansen, John Heaton, Rajnish Mehra, L'uboš Pástor, Dick Thaler, and particularly Alon Brav and John Cochrane, for their insightful comments and constructive criticism. Finally, I thank Lior Menzly for his excellent research assistance and insightful comments throughout this project. Naturally, I remain responsible for errors.

George M. Constantinides
President of the American Finance Association
2001

levels. A celebrated example of such evidence, although by no means the only one, is the failure of the *representative-agent* rational economic paradigm to account for the large average premium of the aggregate return of stocks over short-term bonds and the small average return of short-term bonds from the last quarter of the 19th century to the present. Dubbed the "Equity Premium Puzzle" by Mehra and Prescott (1985), it has generated a cottage industry of rational and behavioral explanations of the level of asset prices and their rate of return.

Another example is the large increase in stock prices in the early and middle 1990s, which Federal Reserve Chairman Alan Greenspan decried as "Irrational Exuberance" even before the unprecedented further increase in stock prices and price–dividend ratios in the late 1990s.

My objective is to revisit some of this evidence and explore the extent to which the rational economic paradigm explains the price level and the rate of return of financial assets over the past 100+ years, both at the macro and micro levels.

In Section I, I reexamine the statistical evidence on the size of the unconditional mean of the aggregate equity return and premium. First, I draw a sharp distinction between *conditional*, *short-term forecasts* of the mean equity return and premium and *estimates of the unconditional mean*. I argue that the currently low conditional short-term forecasts of the return and premium do not lessen the burden on economic theory to explain the large unconditional mean equity return and premium, as measured by their sample average over the past 130 years. Second, I argue that even though one may introduce one's own strong prior beliefs and adjust downwards the sample-average estimate of the premium, the unconditional mean equity premium is at least 6 percent per year and the annual Sharpe ratio is at least 32 percent. These numbers are large and call for an economic explanation.

In Section II, I discuss limitations of the current theory to explain empirical regularities. I argue that per capita consumption growth covaries too little with the return of most classes of financial assets and this implies that the observed aggregate equity return, the long-term bond return, and the observed returns of various subclasses of financial assets are *too large*, *too variable*, and *too predictable*.

In the remaining sections, I revisit and examine the extent to which we can explain the asset returns by relaxing the assumptions of complete consumption insurance, perfect markets, and time-separable preferences. As the reader will readily observe—and I offer my apologies—my choice of issues is eclectic and mirrors in part my own research interests.

In Section III, I show that idiosyncratic income shocks concentrated in periods of economic recession play a key role in generating the mean equity premium, the low risk-free rate, and the predictability of returns. I argue that insufficient attention has been paid to the fact that the annual aggregate labor income exceeds annual dividends by a factor of over 20. Labor income is by far the single most important source of household savings and consumption. The shocks to labor income are uninsurable and persistent and arrive with greater frequency during economic contractions. Idiosyncratic income shocks go a long way toward explaining the unconditional moments of asset returns and the predictability of returns. The construct of per capita consumption is largely irrelevant in explaining the behavior of asset returns because idiosyncratic income shocks are averaged out in per capita consumption.

In Section IV, I show that borrowing constraints over the life cycle play an important role in simultaneously addressing the above issues *and* the demand for bonds. I argue that insufficient attention has been paid to the consumers' life cycle consumption and savings decisions in a market with borrowing constraints. These considerations are important in addressing the limited participation of consumers in the capital markets, the irrelevance of the construct of per capita consumption, and the demand for short-term bonds by consumers with moderate risk aversion, given that equities earn on average a large premium over short-term bonds.

In Section V, I discuss the role of limited market participation. In Section VI, I discuss the role of habit persistence in addressing the same class of issues. In Section VII, I conclude that the observed asset returns do not support the case for abandoning the rational economic theory as our null hypothesis. Much more remains to be done to fully exploit the ramifications of the rational asset-pricing paradigm.

I. How Large Is the Equity Premium?

The average premium of the arithmetic rate of return of the S&P Composite Index over the risk-free rate, measured over the last 130 years, is almost *7 percent* and the annual Sharpe ratio is 36 percent. If the equity premium is a stationary process, then the average premium is an unbiased estimate of the *unconditional* mean equity premium. One may introduce one's own prior beliefs and shave about 1 percent off the premium. The premium and the Sharpe ratio are still large and challenge economic theory for an explanation.

In Table I, I report the sample mean of the annual arithmetic aggregate equity return and of the equity premium. I proxy the aggregate equity return with the S&P Composite Index return. I proxy the annual risk-free rate with the rolled-over return on three-month Treasury bills and certificates. The reported real return is CPI-adjusted for inflation. Over the period 1872 to 2000, the sample mean of the real equity return is 8.9 percent and of the premium is 6.9 percent. Over the period 1926 to 2000, the sample mean of the equity return is 9.7 percent and that of the premium is 9.3 percent. Over the postwar period 1951 to 2000, the sample mean of the equity return is 9.9 percent and that of the premium is 8.7 percent. These sample means are large. Siegel (1998, 1999), Ibbotson Associates (2001), Ibbotson and Chen (2001), Dimson, Marsh, and Staunton (2002), Fama and French (2002), Mehra and Prescott (2002), and several others report the sample means of the equity return and premium in the United States and other countries and conclude that they are large. Some differences arise based on the proxy used for the risk-free rate.

Table I

The Equity Return and Premium

This table shows the sample mean and standard deviation of the annualized real arithmetic return on the S&P Composite Index total return series, the sample mean of the real risk-free rate, and the sample mean of the equity premium. The arithmetic rate of return on equity from the beginning to the end of year t is defined as $R_{t+1} = (P_{t+1} + D_{t+1} - P_t)/P_t$, where P_t is the real price of the aggregate equity at the beginning of year t and D_{t+1} is the aggregate real dividend from the beginning to the end of year t. All returns and premia are in percent. Real returns are CPI adjusted. The table also displays the mean annual growth, $(100/T)\{\ln(P_{T+1}/X_{T+1}) - \ln(P_1/X_1)\}$, of the price/$X$ ratio, where X is the dividends, earnings, book equity, or National Income. The pre-1926 S&P Index price series, the CPI series, the earnings series, and the dividends series are obtained from Shiller's database. The S&P Composite Index returns series post-1926 is obtained from the Ibbotson database. For years prior to 1926, the returns are calculated from the S&P 500 Index and dividend series, assuming no dividend reinvestment. The book equity series is obtained from Davis, Fama, and French (2000) and Vuolteenaho (2000). The National Income is obtained from the Bureau of Labor Statistics. The risk-free rate series is the one constructed by Mehra and Prescott (2002) and is based on an annual average nominal return on three-month Treasury certificates and bills.

	1872–2000	1872–1950	1951–2000	1926–2000
Sample mean S&P return	8.87	8.24	9.87	9.70
Std of return	18.49	19.28	17.32	20.33
Sample mean risk-free rate	2.00	2.54	1.15	0.40
Sample mean premium	6.87	5.69	8.72	9.30
Std of premium	19.19	20.23	17.45	20.50
Sharpe ratio	0.36	0.28	0.50	0.45
Mean annual growth of				
Price/dividends	1.18	−0.22	3.39	1.81
Price/earnings	0.71	−0.57	2.73	1.28
Price/book equity	1.18	−0.11	3.18	2.26
Price/national income	NA	NA	1.27	NA

I draw a sharp distinction between *conditional, short-term forecasts of the mean equity return and premium* and *estimates of the unconditional mean*. The conditional forecasts of the mean equity return and premium at the end of the 20th century and the beginning of the 21st are substantially lower than the estimates of the unconditional mean by at least three measures. First, based on evidence that price–dividend and price–earnings ratios forecast aggregate equity returns and that the values of these ratios prevailing at the beginning of the 21st century are well above their historic averages, Campbell and Shiller (1998) and Shiller (2000) forecast a *conditional* equity premium well below its sample average.[1] Second, Claus and Thomas (2001)

[1] Shiller (1984), Campbell and Shiller (1988a, 1988b), and Fama and French (1988) provide early evidence that the aggregate price–dividend and price–earnings ratios forecast aggregate equity returns. Goyal and Welch (1999) argue that the out-of-sample evidence is less convincing. I do not review here the debates and extensions relating to this literature. In the following paragraphs and in Appendix A, I argue that the forecastability results provide little, if any, guidance to my primary goal in this section, the estimation of the unconditional mean equity return.

calculate the expected aggregate equity premium to be a little above 3 percent in the period 1985 to 1998, based on analysts' earnings forecasts. Third, Welch (2001) reports that the mean forecast among finance and economics professors for the one-year conditional equity premium is 3.5 percent in 2001, down from 6 percent in 1997. These findings are important in their own right and relevant in asset allocation.

However, the currently low conditional, short-term forecasts of the equity premium do not necessarily imply that the unconditional estimate of the mean premium is lower than the sample average. Therefore, the low conditional forecasts do not necessarily lessen the burden on economic theory to explain the large sample average of the equity return and premium over the past 130 years.

The predictability of aggregate equity returns by the price–dividend and price–earnings ratios raises the possibility that use of these financial ratios may improve upon the estimates of the *unconditional* mean equity return (and premium) that are based on the sample mean, an approach pursued earlier by Fama and French (2002).[2] Over the period 1872 to 2000, the price–dividend ratio increased by a factor of 4.6 and the price–earnings ratio by a factor of 2.5. Over the period 1926 to 2000, the price–dividend ratio increased by a factor of 3.9 and the price–earnings ratio increased by a factor of 2.6.[3] One may consider adjusting downwards the sample-mean estimate of the unconditional mean return on equity, but it is unclear by how much.

The size of the adjustment ought to relate to the perceived cause of the increase of these financial ratios. In the year 1998, 52 percent of the U.S. adult population held equity either directly or indirectly, compared to 36 percent of the adult population in 1989. This equitization has been brought about by the increased accessibility of information on the stock market, electronic trading, the growth of mutual funds, the growth of defined-contribution pension plans, and demographic changes. Other regime shifts include the advent of the technology/media/telecoms "new economy" and changes in the taxation of dividends and capital gains. Explanations of the price increase that rely on economic models that are less than fully rational include cultural and psychological factors and tap into the rich and burgeoning literature on behavioral economics and finance.[4]

How does one process this information and adjust the sample mean estimate of the unconditional mean return and premium? To address this issue, I denote by $v_t \equiv \ln(P_t/X_t)$ the logarithm of the ratio of the price to the

[2] The estimators employed in Fama and French (2002) and in this section are discussed in Appendix A.

[3] The increase in these financial ratios should be interpreted with caution. The increase in the price–dividend ratio is due in part to an increase in share repurchases and a decrease in the fraction of dividend-paying firms.

[4] I do not provide a systematic review of the offered explanations. Heaton and Lucas (1999), Shiller (2000), and McGrattan and Prescott (2001) provide lucid accounts of a number of these explanations in the context of both rational economic models and models that deviate from full rationality.

normalizing variable X_t, where the normalizing variable stands for the aggregate dividends, earnings, book equity, National Income, or some combination of these and other economic variables.[5] I choose the normalizing variable X_t in a way that I can plausibly assert that the log financial ratio is stationary. Over the sample period of length T years, the mean annual (geometric) growth of the financial ratio P_t/X_t is given by $(v_{T+1} - v_1)/T$. I define the *adjusted* estimator of the unconditional mean of the annual aggregate real equity return as the sample mean return, less some fraction beta of the sample mean annual growth of the financial ratio, $\hat{R}_{SAMPLE} - \beta(v_{T+1} - v_1)/T$. If the equity return and the log financial ratio are stationary processes, then the adjusted estimator is unbiased for any value of beta.[6] However, the assumption of stationarity alone is insufficient to determine the value of beta.

The beta of the most efficient (mean squared error) adjusted estimator is equal to the slope coefficient of the regression of the sample mean return on the sample mean growth of the financial ratio, $(v_{T+1} - v_1)/T$. Since I have only one sample (of length T), I cannot run such a regression and must rely on information *outside the sample* and/or prior beliefs about the underlying economic model. In Appendix A, I present a set of sufficient conditions that imply that the beta of the most efficient estimator within this class of adjusted estimators is equal to one, when the adjustment is based on the price–dividend ratio. In addition to stationarity, the other main conditions are that the price–dividend ratio does not forecast the *long-run* growth in dividends *and* the *long-run* dividend growth does not forecast the price–dividend ratio. Adoption of the stationarity and (non)forecastability conditions requires strong prior beliefs.

In Table I, I report the mean annual growth of various financial ratios. Over the period 1951 to 2000, the mean annual growth of the price–dividend ratio is 3.4, the price–earnings ratio is 2.7, the price–book equity ratio is 3.2, and the price–National Income ratio is 1.3. Even if I subtract the entire mean annual growth of the price–earnings ratio from the sample mean, the adjusted estimate of the unconditional mean premium is 6.0 percent and is large. The corresponding estimate over the 1926 to 2000 period is 8.0 percent.

An alternative approach is to consider the longer sample period 1872 to 2000. Over this period, the mean annual growth of the price–dividend ratio and price–earnings ratio is 1.2 percent and 0.7 percent, respectively. Thus, this type of adjustment is largely a nonissue over the full sample. Essentially, the change in the financial ratios is "amortized" over 129 years and makes little difference in the estimate. Over the full period 1872 to 2000, the sample mean equity premium is 6.9 percent and the annual Sharpe ratio is

[5] The ratio of the stock market value to the National Income is discussed in Mehra (1998).

[6] A caveat is in order: Without additional assumptions, it is unclear what optimality properties (beyond unbiasedness) are associated with this class of estimators. Neither least squares, nor maximum likelihood, nor Bayesian methods motivate this class of estimators without further assumptions.

36 percent. Any adjustment with the average growth of the financial ratios still leaves the unconditional mean premium large and in need of an economic explanation.

II. Limitations of the Current Theory

The neoclassical rational-expectations economic model parsimoniously links the returns of all assets to the per capita consumption growth through the Euler equations of consumption (see Merton (1973), Rubinstein (1976), Lucas (1978), and Breeden (1979)). According to the theory, the risk premia of financial assets are explained by their covariance with per capita consumption growth. However, per capita consumption growth covaries too little with the returns of most classes of financial assets and this creates a whole class of asset-pricing puzzles: the aggregate equity return, the long-term bond return, and the returns of various subclasses of financial assets are too large, too variable, and too predictable. Attempts to leverage the low covariability typically backfire, implying that the observed risk-free rate is too low and has too low variance. I discuss in some depth the aggregate equity puzzle because it exemplifies many of the problems that arise in attempting to explain the premium of any subclass of financial assets.

The covariance of the per capita consumption growth with the aggregate equity return is *positive*. The rational model explains why the aggregate equity premium is positive. However, the covariance is typically one order of magnitude lower than what is needed to explain the premium. Thus, the equity premium is a *quantitative* puzzle.[7]

The equity premium puzzle is *robust*. One may address the problem by testing the Euler equations of consumption or by calibrating the economy. Either way, it is a puzzle. In calibrating an exchange economy, the model cannot generate the first and second unconditional moments of the equity returns. In testing and rejecting the Euler equations of consumption, one abstracts from the market clearing conditions. The rejections tell us that variations in the assumptions on the supply side of the economy do not resolve the puzzle.

The challenge is a *dual puzzle* of the equity premium that is too high and the risk-free rate that is too low relative to the predictions of the model. In calibrating an economy, the strategy of increasing the risk aversion coefficient in order to lever the effect of the problematic low covariance of consumption growth with equity returns increases the predicted risk-free rate

[7] Grossman and Shiller (1981), Hansen and Singleton (1982), Ferson and Constantinides (1991), Hansen and Jagannathan (1991), and many others test and reject the Euler equations of consumption. Mehra and Prescott (1985) calibrate an economy to match the process of consumption growth. They demonstrate that the unconditional mean annual premium of the aggregate equity return over the risk-free rate is, at most, 0.35 percent. This is too low, no matter how one estimates the unconditional mean equity premium. Weil (1989) stresses that the puzzle is a dual puzzle of the observed too high equity return and too low risk-free rate.

and aggravates the risk-free-rate puzzle. In testing the Euler equations of consumption, the rejections are strongest when the risk-free rate is included in the set of test assets.

Several generalizations of essential features of the model have been proposed to mitigate its poor performance. They include alternative assumptions on preferences,[8] modified probability distributions to admit rare but disastrous market-wide events,[9] incomplete markets,[10] and market imperfections.[11] They also include a better understanding of data problems such as limited participation of consumers in the stock market,[12] temporal aggregation,[13] and the survival bias of the U.S. capital market.[14] Many of these generalizations contribute in part toward our better understanding of the economic mechanism that determines the pricing of assets. I refer the reader to the excellent reviews in the textbooks by Campbell, Lo, and MacKinlay (1997) and Cochrane (2001), and in the articles by Cochrane and Hansen (1992), Kocherlakota (1996), Cochrane (1997), Campbell (2001, 2002), and Mehra and Prescott (2002).

III. Idiosyncratic Income Shocks and Incomplete Markets

A. *The Role of Idiosyncratic Income Shocks*

In economic recessions, investors are exposed to the double hazard of stock market losses and job loss. Investment in equities not only fails to hedge the risk of job loss but also accentuates its implications. Investors require a hefty equity premium in order to be induced to hold equities. In sum, this is the argument that I formalize below and address the predictability of asset returns and their unconditional moments.

The observed correlation of per capita consumption growth with stock returns is low. Over the years, I have grown skeptical of how meaningful an economic construct *aggregate* (as opposed to *disaggregate*) consumption is,

[8] For example, Abel (1990), Constantinides (1990), Epstein and Zin (1991), Ferson and Constantinides (1991), Benartzi and Thaler (1995), Campbell and Cochrane (1999), Anderson, Hansen, and Sargent (2000), Bansal and Yaron (2000), and Boldrin, Christiano, and Fisher (2001).

[9] The merits of this explanation are discussed in Mehra and Prescott (1988) and Rietz (1988).

[10] For example, Bewley (1982), Mehra and Prescott (1985), Mankiw (1986), Constantinides and Duffie (1996), Heaton and Lucas (1996), Storesletten, Telmer, and Yaron (2001), Brav, Constantinides, and Geczy (2002), and Krebs (2002).

[11] For example, Aiyagari and Gertler (1991), Danthine, Donaldson, and Mehra (1992), He and Modest (1995), Bansal and Coleman (1996), Heaton and Lucas (1996), Daniel and Marshall (1997), and Constantinides, Donaldson, and Mehra (2002a).

[12] Mankiw and Zeldes (1991), Brav and Geczy (1995), Attanasio, Banks, and Tanner (2002), Brav et al. (2002), and Vissing-Jorgensen (2002).

[13] Heaton (1995), Lynch (1996), and Gabaix and Laibson (2001).

[14] See Brown, Goetzmann, and Ross (1995). However, Jorion and Goetzmann (1999, Table 6) find that the average real capital gain rate of a U.S. equities index exceeds the average rate of a global equities index that includes both markets that have and have not survived by merely one percent per year.

and how hard we should push aggregate or per capita consumption to explain returns. At a theoretical level, aggregate consumption is a meaningful economic construct if the market is complete or effectively so.[15] In a complete market, heterogeneous households are able to equalize, state by state, their marginal rate of substitution. The equilibrium in a heterogeneous-household, full-information economy is isomorphic in its pricing implications to the equilibrium in a representative-household, full-information economy, if households have von Neumann–Morgenstern preferences.[16] The strong assumption of market completeness is indirectly built into asset pricing models in finance and neoclassical macroeconomic models through the assumption of the existence of a representative household.

Bewley (1982), Mehra and Prescott (1985), and Mankiw (1986) suggest the potential of enriching the asset-pricing implications of the representative-household paradigm, by relaxing the assumption of complete markets.[17] Constantinides and Duffie (1996) find that incomplete markets substantially enrich the implications of the representative-household model. Their main result is a proposition demonstrating, by construction, the existence of household income processes, consistent with given aggregate income and dividend processes, such that equilibrium equity and bond price processes match the given equity and bond price processes.

The theory requires that the idiosyncratic income shocks must have three properties in order to explain the returns on financial assets. First, they must be *uninsurable*. If the income shocks can be insured, then the household consumption growth is equal, state by state, to the aggregate consumption growth, and household consumption growth cannot do better than aggregate consumption growth in explaining the returns. Second, the income shocks must be *persistent*. If the shocks are transient, then households can smooth their consumption by borrowing or by drawing down their savings.[18] Third, the income shocks must be *heteroscedastic*, with *countercyclical* conditional variance.

A good example of a major uninsurable income shock is job loss. Job loss is *uninsurable* because unemployment compensation is inadequate. Layoffs have *persistent* implications on household income, even though the laid-off

[15] The market is effectively complete when all households have preferences that imply one-fund or two-fund separation.

[16] See Negishi (1960), Constantinides (1982), and Mehra and Prescott (1985, an unpublished earlier draft).

[17] There is an extensive literature on the hypothesis of complete consumption insurance. See Cochrane (1991), Mace (1991), Altonji, Hayashi, and Kotlikoff (1992), and Attanasio and Davis (1997).

[18] Aiyagari and Gertler (1991) and Heaton and Lucas (1996) find that consumers facing *transient* shocks come close to the complete-markets rule of complete risk sharing even with transaction costs and/or borrowing costs, *provided that the supply of bonds is not restricted to an unrealistically low level.*

workers typically find another job quickly.[19] Layoffs are *countercyclical* as they are more likely to occur in recessions.

The first implication of the theory is an explanation of the countercyclical behavior of the equity risk premium: The risk premium is highest in a recession because the stock is a poor hedge against the uninsurable income shocks, such as job loss, that are more likely to arrive during a recession.

The second implication is an explanation of the unconditional equity premium puzzle: Even though per capita consumption growth is poorly correlated with stocks returns, investors require a hefty premium to hold stocks over short-term bonds because stocks perform poorly in recessions, when the investor is most likely to be laid off.

Since the proposition demonstrates the existence of equilibrium in frictionless markets, it implies that the Euler equations of household (but not necessarily of per capita) consumption must hold. Furthermore, *since the given price processes have embedded in them whatever predictability of returns by the price–dividend ratios, dividend growth rates, and other instruments that the researcher cares to ascribe to returns, the equilibrium price processes have this predictability built into them by construction.*

B. Empirical Evidence and Generalizations

Brav et al. (2002) provide empirical evidence of the importance of uninsurable idiosyncratic income risk on pricing. They estimate the RRA coefficient and test the set of Euler equations of *household* consumption on the premium of the value-weighted and the equally weighted market portfolio return over the risk-free rate, and on the premium of value stocks over growth stocks.[20] They do not reject the Euler equations of *household* consumption with RRA coefficient between two and four, although they reject the Euler equations of per capita consumption with any value of the RRA coefficient. A RRA coefficient between two and four is economically plausible.

Open questions remain that warrant further investigation. According to the theory in Constantinides and Duffie (1996), periods with frequent and large uninsurable idiosyncratic income shocks are associated with both dispersed cross-sectional distribution of the household consumption growth and low stock returns. An interesting empirical question is *which moments* of the

[19] The empirical evidence is sensitive to the model specification. Heaton and Lucas (1996) model the income process as *univariate* and provide empirical evidence from the Panel Study on Income Dynamics (PSID) that the idiosyncratic income shocks are transitory. Storesletten et al. (2001) model the income process as *bivariate* and provide empirical evidence from the PSID that the idiosyncratic income shocks have a highly persistent component that becomes more volatile during economic contractions. Storesletten, Telmer, and Yaron (2000) corroborate the latter evidence by studying household consumption over the life cycle.

[20] In related studies, Jacobs (1999) studies the PSID database on food consumption; Cogley (2002) and Vissing-Jorgensen (2002) study the CEX database on broad measures of consumption; Jacobs and Wang (2001) study the CEX database by constructing synthetic cohorts; and Ait-Sahalia, Parker, and Yogo (2001) instrument the household consumption with the purchases of certain luxury goods.

cross-sectional distribution of the household consumption growth capture the dispersion. Brav et al. (2002) find that, in addition to the mean and variance, the *skewness* of the cross-sectional distribution is important in explaining the equity premium.

Krebs (2002) provides a theoretical justification as to why it is possible that neither the variance nor the skewness, but higher moments of the cross-sectional distribution are important in explaining the equity premium. He extends the Constantinides and Duffie (1996) model that has only *lognormal* idiosyncratic income shocks by introducing *rare* idiosyncratic income shocks that drive consumption close to *zero*. In his model, the conditional variance and skewness of the idiosyncratic income shocks are nearly constant over time. Despite this, Krebs demonstrates that the original proposition of Constantinides and Duffie remains valid, that is, there exist household income processes, consistent with given aggregate income and dividend processes, such that equilibrium equity and bond price processes match the given equity and bond price processes. Essentially, he provides a theoretical justification as to why it may be hard to empirically detect the rare but catastrophic shocks in the low-order cross-sectional moments of household consumption growth. In Appendix B, I present an example based on Krebs (2002).

A promising direction for future research is to address the relation between the equity return and the higher-order cross-sectional moments of household consumption with Monte Carlo methods. Another promising direction is to instrument the hard-to-observe time-series changes in the cross-sectional distribution with Labor Bureau statistics.

IV. The Life Cycle and Borrowing Constraints

A. Borrowing Constraints over the Life Cycle

Borrowing constraints provide an endogenous partial explanation for the limited participation of young consumers in the stock market. Constantinides et al. (2002a) construct an overlapping-generations exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period. The key feature is that the bulk of the *future* income of the young consumers is derived from their wages forthcoming in their middle age, while the *future* income of the middle-aged consumers is derived primarily from their savings in equity and bonds.

The young would like to invest in equity, given the observed large equity premium. However, they are unwilling to decrease their current consumption in order to save by investing in equity, because the bulk of their lifetime income is derived from their wages forthcoming in their middle age. They would like to borrow, but the borrowing constraint prevents them from doing so. Human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection. The model explains why many consumers do not participate in the stock market in the early phase of their life cycle.

The future income of the middle-aged consumers is derived from their current savings in equity and bonds. Therefore, the risk of holding equity and bonds is concentrated in the hands of the middle-aged saving consumers. This concentration of risk generates the high equity premium and the demand for bonds, in addition to the demand for equity, by the middle-aged.[21] The model recognizes and addresses *simultaneously*, at least in part, the *equity premium*, the *limited participation* in the stock market, and the *demand for bonds*.

The model serves as a useful laboratory to address a range of economic issues. Campbell et al. (2001), and Constantinides, Donaldson, and Mehra (2001) address the cost of Social Security reform. Storesletten et al. (2001) explore the interaction of life-cycle effects and the uninsurable wage income shocks and find that the interaction plays an important role in explaining asset returns. Heaton and Lucas (1999) explore whether changes in market participation patterns account for the recent rise in stock prices and find that they do not.

B. Utility of Wealth—An Old Folks' Tale

The low covariance of the growth rate of aggregate consumption with equity returns is a major stumbling block in explaining the mean aggregate equity premium and the cross section of the asset returns, in the context of a representative-consumer economy with time separable preferences. Mankiw and Shapiro (1986) find that the market beta often explains asset returns better than the consumption beta does. Over the years, a number of different economic models have been proposed that effectively increase the covariance of equity returns with the growth rate of aggregate consumption, by proxying the growth rate of aggregate consumption with the aggregate stock market return in the Euler equations of consumption.[22]

I present an old folks' tale, introduced in Constantinides, Donaldson, and Mehra (2002a, 2002b), that accomplishes this goal without introducing Epstein–Zin (1991) preferences or preferences defined directly over wealth.

[21] See also the discussion in the related papers by Bodie, Merton, and Samuelson (1992), Jagannathan and Kocherlakota (1996), Bertaut and Haliassos (1997), Cocco, Gomes, and Maenhout (1999), and Storesletten et al. (2001).

[22] Friend and Blume (1975) explain the mean equity premium with low RRA coefficient by assuming a single-period economy in which the end-of-period consumption inevitably equals the end-of-period wealth. Epstein and Zin (1991) introduce a recursive preference structure that emphasizes the timing of the resolution of uncertainty. Even though the preferences are defined over consumption alone, the stock market return enters directly in the Euler equations of consumption. Bakshi and Chen (1996) introduce a set of preferences defined over consumption and wealth—the spirit of capitalism—that also have the effect of introducing the stock market return in the Euler equations of consumption.

Old folks who are rich enough to be nontrivial investors in the capital markets care about their wealth just as much as younger folks do, even though the state of their health and their medical expenses account for their consumption patterns better than fluctuations of their wealth do. This simple observation takes us a long way toward understanding why the stock market return does a better job than the growth of aggregate consumption does in explaining asset returns.

In the context of an overlapping-generations economy, the major investors in the market are the middle-aged households at the saving phase of their life cycle. These households save with the objective to maximize the utility of their "consumption" in their middle and old age. The insight here is that "consumption" of the old consists of two components, direct consumption, c_D; and the "joy of giving," c_B, in the form of *inter vivos* gifts and *post mortem* bequests. Since the old households' direct consumption is constrained by the state of their health, the correlation between the direct consumption of the old and the stock market return is *low*, a prediction that is borne out empirically. Therefore, the balance of the old households' wealth, c_B, is *a fortiori* highly correlated with the stock market return. In terms of a utility function of consumption at the old age, $u(c_D) + v(c_B)$, that is separable over direct consumption and bequests, the model predicts an Euler equation of consumption with marginal utility at the old age given by $v'(c_B)$ and not by $u'(c_D)$, where c_B is proxied by the stock market value.

This model remains to be tested. Nevertheless, it reinforces the general point that per capita consumption measures neither the total consumption of the marginal investor in the stock market nor that part of the marginal investor's consumption that is unconstrained by health and medical considerations.

V. Limited Stock Market Participation

Limited stock market participation is another potential culprit in understanding why models of per capita consumption do a poor job in explaining returns. Whereas we understood all along that many households whose consumption is counted in the measure of per capita consumption do not hold stocks, it took a paper by Mankiw and Zeldes (1991) to point out that the emperor has no clothes.[23] Even though 52 percent of the U.S. adult population held stock either directly or indirectly in 1998, compared to 36 percent in 1989, stockholdings remain extremely concentrated in the hands of the wealthiest few. Furthermore, wealthy entrepreneurs may be inframarginal in the stock market if their wealth is tied up in private equity.

[23] Since then, several papers have studied the savings and portfolio composition of households, stratified by income, wealth, age, education, and nationality. See Blume and Zeldes (1993), Haliassos and Bertaut (1995), Heaton and Lucas (1999, 2000), Poterba (2001), and the collected essays in Guiso, Haliassos, and Jappelli (2001).

Mankiw and Zeldes (1991) calculate the per capita food consumption of a subset of households, designated as *asset holders* according to a criterion of asset holdings above some threshold. They find that the implied RRA coefficient decreases as the threshold is raised. Brav and Geczy (1995) confirm their result by using the nondurables and services per capita consumption, reconstructed from the Consumer Expenditure Survey (CEX) database. Attanasio et al. (2002), Brav et al. (2002), and Vissing-Jorgensen (2002) find some evidence that per capita consumption growth can explain the equity premium with a relatively high value of the RRA coefficient, once we account for limited stock market participation. However, Brav et al. point out that the statistical evidence is weak and the results are sensitive to experimental design.

Limited stock market participation is a fact of life and empirical tests of the Euler equations of consumption should account for it. However, my interpretation of the empirical results is that recognition of limited stock market participation alone is insufficient to explain the returns on assets. Essentially, *the subset of households that are marginal in the stock market are still subject to uninsurable idiosyncratic income risk and we should take that into account also in attempting to explain asset returns.*

VI. Habit Persistence

Habit persistence has a long tradition in economic theory, dating back to Marshall (1920) and Duesenberry (1949). It is the property of preferences that an increase in consumption increases the marginal utility of consumption at adjacent dates relative to the marginal utility of consumption at distant ones. Building on earlier work by Ryder and Heal (1973) and Sundaresan (1989), I demonstrate in Constantinides (1990) that habit persistence can, in principle, reconcile the high mean equity premium with the low variance of consumption growth and with the low covariance of consumption growth with equity returns. Habit persistence lowers the intertemporal elasticity of substitution in consumption, given the risk aversion. The mean equity premium is equal to the covariance of consumption growth with equity returns, divided by this elasticity. Therefore, given the risk aversion, habit persistence lowers the elasticity and raises the mean equity premium.[24]

There are several interesting variations of the above class of preferences. Pollak (1970) discusses a model of *external* habit persistence in which the consumer does not take into account the effect of current consumption on future preferences. Abel (1990) and Campbell and Cochrane (1999) address

[24] Ferson and Constantinides (1991) test the special case of the linear habit model in which the habit depends only on the *first* lag of own consumption and report that the habit model performs better than the time-separable model and that the habit persistence parameter is economically and statistically significant. See also Hansen and Jagannathan (1991) and Heaton (1995).

the equity premium in the context of models with external habit persistence. In particular, the latter introduce a *nonlinear* specification of habit, reverse-engineered to keep the variability of the interest rate low. The large average equity premium, the predictability of long-horizon returns, and the behavior of equity prices along the business cycle are induced by a volatile RRA coefficient that has the value of 80 in the steady state and much higher still in economic recessions. Calibrated with the actual history of aggregate consumption, the model hits the aggregate price–dividend ratio in a number of periods but misses it in the 1950s and 1990s.

A promising direction for future research is to endogenize the currently ad hoc specification of the nonlinear habit. Another direction is to address the predictability of asset returns and their behavior along the business cycle in a model that benefits from the added flexibility of the nonlinear specification of habit but keeps risk aversion low and credible with the specification of habit to be *internal*.

Empirical tests of consumption-based models that incorporate habit persistence and aimed at explaining asset returns produce mixed results.[25] It is hardly surprising that the results on both the habit and the external habit persistence models are mixed. The National Income and Product Accounts (NIPA) per capita consumption series is an imperfect proxy of the consumption of investors that are marginal in the capital markets, given the earlier-identified problems of incomplete consumption insurance, limited participation of households in the capital markets, borrowing constraints, and the exclusion of bequests from the definition of consumption. Both NIPA per capita consumption and consumption surplus over habit have low covariance with asset returns. Nonlinear refinements in the definition of habit do not remedy the problem of low covariance with asset returns. Habit persistence may well gain in empirical relevance in explaining asset returns, once we correctly measure the consumption of the unconstrained marginal investors in the capital markets.

Habit persistence is already gaining ground as an ingredient of economic models addressing a diverse set of economic problems beyond asset pricing, including the consumption–saving behavior and the home-equity puzzle. Habit persistence is a sensible property of preferences. It is also a property that allows for the separate specification of the RRA coefficient and the intertemporal elasticity of substitution within the class of von Neuman–Morgenstern preferences.

[25] Ferson and Harvey (1992) report positive results for the linear external habit model. Wachter (2001) reports that long lags of consumption growth predict the short-term interest rate, as implied by the nonlinear external habit model. Li (2001) reports that in both the linear and the nonlinear external habit models, the surplus consumption over habit has limited success in explaining the time series of the premia of stock and bond portfolios. Menzly, Santos, and Veronesi (2001) develop an external habit model and report that it helps explain the cross section of asset returns.

VII. Concluding Remarks

I examine the observed asset returns and conclude that the evidence does not support the case for abandoning the rational economic model. I argue that the standard model is greatly enhanced by relaxing some of its assumptions. In particular, I argue that we go a long way toward addressing market behavior by recognizing that consumers face uninsurable and idiosyncratic income shocks, for example, the loss of employment. The prospect of such events is higher in economic downturns and this observation takes us a long way toward understanding both the unconditional moments of asset returns and their variation along the business cycle.

I also argue that life-cycle considerations are important and often overlooked in finance. Borrowing constraints become important when placed in the context of the life cycle. The fictitious representative consumer that holds all the stock market and bond market wealth does not face credible borrowing constraints. Young consumers, however, do face credible borrowing constraints. I trace their impact on the equity premium, the demand for bonds—Who holds bonds if the equity premium is so high?—and on the limited participation of consumers in the capital markets.

Finally, I argue that relaxing the assumption of convenience that preferences are time separable drives a wedge between the preference properties of risk aversion and intertemporal elasticity of substitution, within the class of von Neumann–Morgenstern preferences. Further work along these lines may enhance our understanding of the price behavior along the business cycle with credibly low risk-aversion coefficient.

I believe that the integration of the notions of *incomplete markets*, the *life cycle*, *borrowing constraints*, and other sources of *limited stock market participation* is a promising vantage point from which to study the prices of asset and their returns both theoretically and empirically within the class of rational asset-pricing models.

At the same time, I believe that specific deviations from rationality in the agents' choices and in the agents' processing of information potentially enhance the realism and economic analysis of certain phenomena on a case-by-case basis.[26] However, several examples of apparent deviation from rationality may be reconciled with the rational economic paradigm, once we recognize that rational investors have incomplete knowledge of the fundamental structure of the economy and engage in learning.[27] In any case, the collection of these deviations from rationality does not yet amount to a new economic paradigm that challenges the rational economic model.

It has been more than 60 years since Keynes (1936) wrote about animal spirits, and 15 since Shiller (1984) wrote about noise traders and DeBondt and Thaler (1985) wrote about stock market overreaction. I have yet to see an *unambiguously articulated* set of principles that emerges from the kalei-

[26] Barberis and Thaler (2002) and Hirshleifer (2001) provide excellent reviews of this literature.

[27] Brav and Heaton (2002) provide excellent discussion of these issues.

doscope of these clinical investigations and that is put forth as an alternative to the rational economic paradigm. Serious scholars are keenly aware of this criticism and hard at work to address it. Until such a paradigm is put forth *and* is empirically vindicated, the rational economic paradigm remains our principal guide to economic behavior.

Appendix A. Estimation of the Unconditional Mean Return on Equity

I define the adjusted estimator of the unconditional mean of the annual aggregate arithmetic real return on equity as

$$\hat{R}_x \equiv T^{-1}\sum_{t=1}^{T} R_{t+1} - \beta T^{-1}(v_{T+1} - v_1) = \hat{R}_{SAMPLE} - \beta T^{-1}(v_{T+1} - v_1). \quad \text{(A1)}$$

The term $v_t \equiv \ln(P_t/X_t)$ is the logarithm of the price of aggregate equity, normalized with the variable X_t, where X_t stands for the aggregate dividends, earnings, book equity, National Income, or some other economic variable.

I assume that R_t and v_t are stationary processes. Then $E[v_{T+1} - v_1] = 0$ and $\hat{R}_x$ is an *unbiased* estimator of the unconditional mean equity return. Note that the assumption of stationarity alone does not determine the value of the parameter beta that provides the most efficient estimator of the unconditional mean equity return. The variance of the estimator $\hat{R}_x$ is

$$\begin{aligned}\operatorname{var}(\hat{R}_x) = \operatorname{var}(\hat{R}_{SAMPLE}) &- 2\beta \operatorname{cov}(\hat{R}_{SAMPLE}, T^{-1}(v_{T+1} - v_1)) \\ &+ \beta^2 \operatorname{var}(T^{-1}(v_{T+1} - v_1))\end{aligned} \quad \text{(A2)}$$

and is minimized when beta is set equal to

$$\beta^* = \frac{\operatorname{cov}(\hat{R}_{SAMPLE}, T^{-1}(v_{T+1} - v_1))}{\operatorname{var}(T^{-1}(v_{T+1} - v_1))}. \quad \text{(A3)}$$

The beta of the most efficient (mean squared error) estimator is equal to the slope coefficient of the regression of $\hat{R}_{SAMPLE}$ on $T^{-1}(v_{T+1} - v_1)$.

Since I have only one sample of length T, I cannot run such a regression and must rely on information *outside the sample* and/or prior beliefs about the underlying economic model. Essentially, within the sample of length T, I can examine the high-frequency behavior of the joint time series R_t and v_t, but I need to assert my prior beliefs on how these findings relate to the behavior of the joint time series at the T-year frequency.

For example, consider the case in which v_t stands for the log price–dividend ratio. Since a high price–dividend ratio forecasts *in-sample* low long-horizon returns, it is a plausible prior belief that it also forecasts low T-horizon returns, $\operatorname{cov}(\hat{R}_{SAMPLE}, v_1) < 0$, for $T = 50$ years (1951 to 2000) or

$T = 129$ years (1872 to 2000). It is also a plausible prior belief that periods of high returns are not followed by low price–dividend ratios, that is, it is plausible to believe that $\mathrm{cov}(\hat{R}_{SAMPLE}, v_{T+1}) \gtrsim 0$. Then equation (A3) implies that the beta of the most efficient estimator is positive.

I present a set of sufficient (but not necessary) conditions that imply that the beta of the most efficient estimator in the class $\hat{R}_x$ equals one. Let v_t stand for the log price–dividend ratio and assume the following: (1) the returns and the price–dividend ratio are stationary, (2) the price–dividend ratio does not forecast the growth in dividends, (3) dividend growth does not forecast the price–dividend ratio, (4) the price–dividend ratio does not forecast the difference in the conditional variance of the capital gain rate and the dividend growth rate, and (5) the difference in the conditional variance of the capital gain rate and the dividend growth rate does not forecast the price–dividend ratio. To prove the claim, I use a Taylor-series expansion:

$$\Delta v_{t+1} = \Delta P_{t+1}/P_t - \Delta D_{t+1}/D_t - k_{t+1} \tag{A4}$$

where

$$k_{t+1} \equiv (\Delta P_{t+1}/P_t)^2/2 - (\Delta D_{t+1}/D_t)^2/2$$

and write the sample mean of the arithmetic return as

$$\begin{aligned}
\hat{R}_{SAMPLE} &\equiv T^{-1}\sum_{t=1}^{T}\{D_{t+1}/P_t + \Delta P_{t+1}/P_t\} \\
&= T^{-1}\sum_{t=1}^{T}\{D_{t+1}/P_t + \Delta D_{t+1}/D_t + k_{t+1} + \Delta v_{t+1}\} \qquad \text{(A5)}\\
&= T^{-1}\sum_{t=1}^{T}\{D_{t+1}/P_t + \Delta D_{t+1}/D_t + k_{t+1}\} + T^{-1}(v_{T+1} - v_1).
\end{aligned}$$

I substitute the value of $\hat{R}_{SAMPLE}$ from equation (A5) into equation (A3) and obtain the result that the variance of the estimator is minimized when the value of beta is one:

$$\begin{aligned}
\beta^* &= \frac{\mathrm{cov}\left(\sum_{t=1}^{T} D_{t+1}/P_t, (v_{T+1} - v_1)\right)}{\mathrm{var}(v_{T+1} - v_1)} + \frac{\mathrm{cov}\left(\sum_{t=1}^{T} \Delta D_{t+1}/D_t, v_{T+1}\right)}{\mathrm{var}(v_{T+1} - v_1)} \\
&\quad - \frac{\mathrm{cov}\left(\sum_{t=1}^{T} \Delta D_{t+1}/D_t, v_1\right)}{\mathrm{var}(v_{T+1} - v_1)} + \frac{\mathrm{cov}\left(\sum_{t=1}^{T} k_{t+1}, (v_{T+1} - v_1)\right)}{\mathrm{var}(v_{T+1} - v_1)} + 1 \qquad \text{(A6)}\\
&= 1.
\end{aligned}$$

The first term in equation (A6) is approximately zero because the stationarity of the price–dividend ratio implies

$$\operatorname{cov}\left(\sum_{t=1}^{T} D_{t+1}/P_t, v_{T+1}\right) \approx \operatorname{cov}\left(\sum_{t=1}^{T} D_{t+1}/P_t, v_1\right). \tag{A7}$$

The second term in equation (A6) is zero because, by assumption, the dividend growth rate does not forecast the price–dividend ratio. The third term is zero because, by assumption, the price–dividend ratio does not forecast the dividend growth. Finally, the fourth term is zero because, by assumption, the price–dividend ratio does not forecast and is not forecasted by the difference of the conditional variance of the capital gain rate and the dividend growth rate.

Thus, when X_t stands for the dividends and conditions (1)–(5) hold, the minimum variance estimator in the class of estimators given by equation (A1) is

$$\begin{aligned}\hat{R}_D &= \hat{R}_{SAMPLE} - T^{-1}(v_{T+1} - v_1) \\ &= T^{-1}\sum_{t=1}^{T}\{D_{t+1}/P_t + \Delta D_{t+1}/D_t\} + T^{-1}\sum_{t=1}^{T} k_{t+1}.\end{aligned} \tag{A8}$$

Fama and French (2002) report adjusted estimates of the unconditional mean return (and premium) based on the fundamentals dividends and earnings. Specifically, their estimate of the expected stock return based on the dividend growth model is equivalent to $T^{-1}\sum_{t=1}^{T}\{D_{t+1}/P_t + \Delta D_{t+1}/D_t\}$ and their biased-adjusted estimate is equivalent to $T^{-1}\sum_{t=1}^{T}\{D_{t+1}/P_t + \Delta D_{t+1}/D_t\} + T^{-1}\sum_{t=1}^{T} k_{t+1}$. Ibbotson and Chen (2001) also report adjusted estimates of the unconditional mean return (and premium) based on dividends, income, earnings, payout ratio, book equity, and National Income.

Appendix B. Extension of the Constantinides and Duffie (1996) Model

I illustrate an extension of the Constantinides and Duffie (1996) model along the lines of Krebs (2002). The extension provides theoretical justification as to why it may be hard to detect empirically in the low-order cross-sectional moments of household consumption growth the rare but catastrophic shocks that play a major role in driving asset prices.

The ith household's consumption, $c_{i,t}$, follows the process

$$\frac{c_{it}}{c_{i,t-1}} = \frac{c_t}{c_{t-1}} X_{i,t}\eta_{i,t}. \tag{B1}$$

The random variables $\{\eta_{i,t}\}$ have the following properties: Distinct subsets of $\{\eta_{i,t}\}$ are independent; for all i and t; $\eta_{i,t}$ is independent of $c_{t-1}, c_t, c_{i,t-1}, X_{i,t}$, and the asset prices; and $E[\eta_{i,t}] = 1$. Since the random variables $\{\eta_{i,t}\}$ are independent of the asset prices, they do not contribute to the equity premium. One may choose to view them as observation error, but does not have to.

In the Constantinides and Duffie (1996) model, the idiosyncratic income shocks are lognormal: $X_{i,t} = e^{b_t \epsilon_{i,t} - b_t^2/2}$ with $\epsilon_{i,t}$ normal and $\eta_{i,t} \equiv 1$. The conditional variance, b_t^2, explains the risk premia because it is modeled as countercyclical and correlated with the stock returns. Whereas Brav et al. (2002) find that the pricing kernel $I^{-1}\sum_{i=1}^{I}(c_{i,t}/c_{i,t-1})^{-\alpha}$ goes a long way toward explaining the equity premium and the value-versus-growth premium, they also find little evidence that the conditional variance, b_t^2, is correlated with stock returns, or indeed whether the time series of this variance has any discernible pattern relative to the business cycle. I build this feature in the model by choosing a binomial distribution for $X_{i,t}$.

I assume that the random variables $\{X_{it}\}$ have the following properties: Distinct subsets of $\{X_{it}\}$ are independent; for all i and t, X_{it} is independent of $c_{t-1}, c_t, c_{i,t-1}$ and $X_{i,t-1}$; and X_{it} has the following binomial distribution:

$$
\begin{aligned}
X_{i,t} &= \frac{1 - y_t^{-\alpha^{-1}} \pi^{1+\alpha^{-1}}}{1-\pi}, \quad \text{with probability } 1-\pi \\
&= y_t^{-\alpha^{-1}} \pi^{\alpha^{-1}}, \quad \text{with probability } \pi,
\end{aligned}
\tag{B2}
$$

where $0 < \pi \ll 1$, and α is the constant RRA coefficient. The variable $y_t, y_t > 0$ is defined shortly. Since

$$
E\left[\frac{c_{it}}{c_{i,t-1}} \middle| y_t, \frac{c_t}{c_{t-1}}\right] = \frac{c_t}{c_{t-1}},
\tag{B3}
$$

arguments along the lines in Constantinides and Duffie (1996) identify c_t as the per capita consumption.

The time-t expectation of the ith household's marginal rate of substitution, conditional on $\{c_t/c_{t-1}, y_t\}$, is

$$
\begin{aligned}
&E\left[e^{-\rho}\left(\frac{c_{it}}{c_{i,t-1}}\right)^{-\alpha} \middle| \frac{c_t}{c_{t-1}}, y_t\right] \\
&\quad = e^{-\rho}\left(\frac{c_t}{c_{t-1}}\right)^{-\alpha}\{(1-\pi)^{1+\alpha}(1 - y_t^{-\alpha^{-1}}\pi^{1+\alpha^{-1}})^{-\alpha} + y_t\} E[\eta_{i,t}^{-\alpha}] \\
&\quad \approx e^{-\rho}\left(\frac{c_t}{c_{t-1}}\right)^{-\alpha}(1 + y_t) E[\eta_{i,t}^{-\alpha}], \quad \text{for } \pi \ll 1.
\end{aligned}
\tag{B4}
$$

I define the variable y_t implicitly with the equation

$$e^{-\rho}\left(\frac{c_t}{c_{t-1}}\right)^{-\alpha}(1+y_t)E[\eta_{i,t}^{-\alpha}] = M_t, \tag{B5}$$

where M_t is the pricing kernel that supports the given joint process of aggregate income, asset prices, and dividends. By construction, it follows that any individual household's marginal rate of substitution, $e^{-\rho}(c_{it}/c_{i,t-1})^{-\alpha}$, supports the given joint process of aggregate income, asset prices, and dividends.

Finally, I demonstrate that the variance, skewness, and higher moments of the cross-sectional distribution of the households' consumption growth need not bear any relationship to asset returns and the business cycle. This is despite the fact that each individual household's marginal rate of substitution supports the given joint process of aggregate income, asset prices, and dividends.

The Nth central moment, $N \geq 1$, of the households' logarithmic consumption growth is the sum of the Nth central moments of $\ln(c_t/c_{t-1})$, $\ln(X_{i,t})$, and $\ln(\eta_{i,t})$, given the assumed independence of $c_t/c_{t-1}, X_{i,t}$, and $\eta_{i,t}$. It is easily shown that

$$\lim_{\pi\to 0} E[(\ln X_{i,t})^N] = 0, \quad N \geq 1. \tag{B6}$$

If the probability of the idiosyncratic consumption shocks is sufficiently low, $\pi \ll 1$, the central moments of the households' consumption growth are driven by the corresponding central moments of the per capita consumption growth and $\eta_{i,t}$. These moments need not bear any pattern relating to the business cycle and need not be correlated in any particular way with the asset returns. Despite this, each individual household's marginal rate of substitution supports the given joint process of aggregate income, asset prices, and dividends. The illustration explains why it may be empirically difficult or infeasible to detect the idiosyncratic consumption shocks in the cross-sectional moments of household consumption growth.

REFERENCES

Abel, Andrew B., 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review Papers and Proceedings* 80, 38–42.

Ait-Sahalia, Yacine, Jonathan A. Parker, and Motohiro Yogo, 2001, Luxury goods and the equity premium, Working paper 8417, NBER.

Aiyagari, Rao S., and Mark Gertler, 1991, Asset returns with transactions costs and uninsured individual risk, *Journal of Monetary Economics* 27, 311–331.

Altonji, Joseph G., Fumio Hayashi, and Laurence J. Kotlikoff, 1992, Is the extended family altruistically linked? *American Economic Review* 82, 1177–1198.

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent, 2000, Robustness, detection and the price of risk, Working paper, University of Chicago.

Attanasio, Orazio P., James Banks, and Sarah Tanner, 2002, Asset holding and consumption volatility, *Journal of Political Economy* 110, forthcoming.

Attanasio, Orazio P., and Steven J. Davis, 1997, Relative wage movements and the distribution of consumption, *Journal of Political Economy* 104, 1227–1262.

Bakshi, Gurdip, and Zhiwu Chen, 1996, The spirit of capitalism and stock market prices, *American Economic Review* 86, 133–157.

Bansal, Ravi, and John W. Coleman, 1996, A monetary explanation of the equity premium, term premium, and risk-free rate puzzles, *Journal of Political Economy* 104, 1135–1171.

Bansal, Ravi, and Amir Yaron, 2000, Risks for the long run: A potential resolution of asset pricing puzzles, Working paper 8059, NBER.

Barberis, Nicholas, and Richard Thaler, 2002, A survey of behavioral finance, in George M. Constantinides, Milton Harris, and Rene Stulz, eds.: *Handbook of the Economics of Finance* (North-Holland Elsevier, Amsterdam), forthcoming.

Benartzi, Shlomo, and Richard H. Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics* 110, 73–92.

Bertaut, Carol C., and Michael Haliassos, 1997, Precautionary portfolio behavior from a life-cycle perspective, *Journal of Economic Dynamics and Control* 21, 1511–1542.

Bewley, Truman F., 1982, Thoughts on tests of the intertemporal asset pricing model, Working paper, Northwestern University.

Blume, Marshall E., and Stephen P. Zeldes, 1993, The structure of stock ownership in the U.S., Working paper, University of Pennsylvania.

Bodie, Zvi, Robert C. Merton, and William F. Samuelson, 1992, Labor supply flexibility and portfolio choice in a life cycle model, *Journal of Economic Dynamics and Control* 16, 427–449.

Boldrin, Michel, Lawrence J. Christiano, and Jonas D. M. Fisher, 2001, Habit persistence, asset returns, and the business cycle, *American Economic Review* 91, 149–166.

Brav, Alon, George M. Constantinides, and Christopher C. Geczy, 2002, Asset pricing with heterogeneous consumers and limited participation: Empirical evidence, *Journal of Political Economy* 110, forthcoming.

Brav, Alon, and Christopher C. Geczy, 1995, An empirical resurrection of the simple consumption CAPM with power utility, Working paper, University of Chicago.

Brav, Alon, and J. B. Heaton, 2002, Competing theories of financial anomalies, *Review of Financial Studies* 15, 475–506.

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Brown, Stephen J., William N. Goetzmann, and Stephen Ross, 1995, Survival, *Journal of Finance* 50, 853–873.

Campbell, John Y., 2001, Asset pricing at the millennium, *Journal of Finance* 55, 1515–1567.

Campbell, John Y., 2002, Consumption-based asset pricing, in George M. Constantinides, Milton Harris, and Rene Stulz, eds.: *Handbook of the Economics of Finance* (North-Holland Elsevier, Amsterdam), forthcoming.

Campbell, John Y., Joao F. Cocco, Francisco J. Gomes, and Pascal J. Maenhout, 2001, Investing retirement wealth: A life-cycle model, in John Y. Campbell and Martin Feldstein, eds.: *Risk Aspects of Social Security Reform* (University of Chicago Press, Chicago).

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

Campbell, John Y., and Robert J. Shiller, 1988a, The dividend–price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Robert J. Shiller, 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance*, 43, 661–676.

Campbell, John Y., and Robert J. Shiller, 1998, Valuation ratios and the long-run stock market outlook, *Journal of Portfolio Management* 24, 11–26.

Claus, James, and Jacob Thomas, 2001, Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets, *Journal of Finance* 56, 1629–1666.

Cocco, Joao F., Francisco J. Gomes, and Pascal J. Maenhout, 1999, Portfolio choice over the life cycle, Working paper, Harvard University.

Cochrane, John H., 1991, A simple test of consumption insurance, *Journal of Political Economy* 99, 957–976.

Cochrane, John H., 1997, Where is the market going? Uncertain facts and novel theories, *Economic Perspectives* (Federal Reserve Bank of Chicago) 21, 3–37.

Cochrane, John H., 2001, *Asset Pricing* (Princeton University Press, Princeton, NJ).

Cochrane, John H., and Lars Peter Hansen, 1992, Asset pricing explorations for macroeconomics, in Olivier J. Blanchard and Stanley Fischer, eds.: *NBER Macroeconomics Annual* (MIT Press, Cambridge, MA).

Cogley, Timothy, 2002, Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey, *Journal of Monetary Economics* 49, 309–334.

Constantinides, George M., 1982, Intertemporal asset pricing with heterogeneous consumers and without demand aggregation, *Journal of Business* 55, 253–267.

Constantinides, George M., 1990, Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.

Constantinides, George M., John B. Donaldson, and Rajnish Mehra, 2001, Junior must pay: Pricing the implicit put in privatizing social security, Working paper, University of Chicago.

Constantinides, George M., John B. Donaldson, and Rajnish Mehra, 2002a, Junior can't borrow: A new perspective on the equity premium puzzle, *Quarterly Journal of Economics* 117, 269–296.

Constantinides, George M., John B. Donaldson, and Rajnish Mehra, 2002b, Junior is rich: Bequests as consumption, Working paper, University of Chicago.

Constantinides, George M., and Darrell Duffie, 1996, Asset pricing with heterogeneous consumers, *Journal of Political Economy* 104, 219–240.

Daniel, Kent, and David Marshall, 1997, The equity premium puzzle and the risk-free rate puzzle at long horizons, *Macroeconomic Dynamics* 1, 452–484.

Danthine, Jean-Pierre, John B. Donaldson, and Rajnish Mehra, 1992, The equity premium and the allocation of income risk, *Journal of Economic Dynamics and Control* 16, 509–532.

Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns, *Journal of Finance* 55, 389–406.

DeBondt, Werner F. M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.

Dimson, Elroy, Paul Marsh, and Michael Staunton, 2002, *Triumph of the Optimists: 101 Years of Global Investment Returns* (Princeton University Press, Princeton, NJ).

Duesenberry, James S., 1949, *Income, Saving, and the Theory of Consumer Behavior* (Harvard University Press, Cambridge, MA).

Epstein, Larry G., and Stanley E. Zin, 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis, *Journal of Political Economy* 99, 263–286.

Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Fama, Eugene F., and Kenneth R. French, 2002, The equity premium, *Journal of Finance* 57, 637–660.

Ferson, Wayne E., and George M. Constantinides, 1991, Habit persistence and durability in aggregate consumption, *Journal of Financial Economics* 29, 199–240.

Ferson, Wayne E., and Campbell R. Harvey, 1992, Seasonality and consumption-based asset pricing, *Journal of Finance* 47, 511–552.

Friend, Irwin, and Marshall E. Blume, 1975, The demand for risky assets, *The American Economic Review* 65, 900–922.

Gabaix, Xavier, and David Laibson, 2001, The 6D bias and the equity premium puzzle, in Ben Bernanke and Ken Rogoff, eds.: *NBER Macroeconomics Annual 2001* (MIT Press, Cambridge, MA).

Goyal, Amit, and Ivo Welch, 1999, Predicting the equity premium, Working paper, UCLA.

Grossman, Sanford J., and Robert J. Shiller, 1981, The determinants of the variability of stock market prices, *American Economic Review Papers and Proceedings* 71, 222–227.

Guiso, Luigi, Michael Haliassos, and Tullio Jappelli, 2001, *Household Portfolios* (MIT Press, Cambridge, MA).

Haliassos, Michael, and Carol C. Bertaut, 1995, Why do so few hold stocks? *The Economic Journal* 105, 1110–1129.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1288.

He, Hua, and David M. Modest, 1995, Market frictions and consumption-based asset pricing, *Journal of Political Economy* 103, 94–117.

Heaton, John C., 1995, An empirical investigation of asset pricing with temporally dependent preference specifications, *Econometrica* 63, 681–717.

Heaton, John C., and Deborah J. Lucas, 1996, Evaluating the effects of incomplete markets on risk sharing and asset pricing, *Journal of Political Economy* 104, 443–487.

Heaton, John C., and Deborah J. Lucas, 1999, Stock prices and fundamentals, in Ben Bernanke and Julio Rotemberg, eds.: *NBER Macroeconomics Annual 1999* (MIT Press, Cambridge, MA).

Heaton, John, C., and Deborah J. Lucas, 2000, Portfolio choice and asset prices: The importance of entrepreneurial risk, *Journal of Finance* 55, 1163–1198.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533–1597.

Ibbotson Associates, 2001, *Stocks, Bonds, Bills, and Inflation Yearbook* (Ibbotson Associates, Chicago).

Ibbotson, Roger G., and Peng Chen, 2001, Stock market returns in the long run: Participating in the real economy, Working paper, Ibbotson Associates.

Jacobs, Kris, 1999, Incomplete markets and security prices: Do asset-pricing puzzles result from aggregation problems? *Journal of Finance* 54, 123–163.

Jacobs, Kris, and Kevin Q. Wang, 2001, Idiosyncratic consumption risk and the cross section of asset returns, Working paper, McGill University and University of Toronto.

Jagannathan, Ravi, and Narayana R. Kocherlakota, 1996, Why should older people invest less in stocks than younger people? *Federal Bank of Minneapolis Quarterly Review* 20, 11–23.

Jorion, Philippe, and William N. Goetzmann, 1999, Global stock markets in the twentieth century, *Journal of Finance* 54, 953–980.

Keynes, John M., 1936, *The General Theory of Employment, Interest, and Money* (Hartcourt, Brace and Company, New York).

Kocherlakota, Narayana R., 1996, The equity premium: It's still a puzzle, *Journal of Economic Literature* 34, 42–71.

Krebs, Tom, 2002, Testable implications of consumption-based asset pricing with incomplete markets, *Journal of Mathematical Economics*, forthcoming.

Li, Yuming, 2001, Expected returns and habit persistence, *Review of Financial Studies* 14, 861–899.

Lucas, Robert, Jr., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1446.

Lynch, Anthony W., 1996, Decision frequency and synchronization across agents: Implications for aggregate consumption and equity returns, *Journal of Finance* 51, 1479–1497.

Mace, Barbara J., 1991, Full insurance in the presence of aggregate uncertainty, *Journal of Political Economy* 99, 928–956.

Mankiw, N. Gregory, 1986, The equity premium and the concentration of aggregate shocks, *Journal of Financial Economics* 17, 211–219.

Mankiw, N. Gregory, and Matthew D. Shapiro, 1986, Risk and return: Consumption beta versus market beta, *Review of Economics and Statistics* 68, 452–459.

Mankiw, N. Gregory, and Stephen P. Zeldes, 1991, The consumption of stockholders and nonstockholders, *Journal of Financial Economics* 29, 97–112.

Marshall, Alfred, 1920, *Principles of Economics: An Introductory Volume* (Macmillan, London).

McGrattan, Ellen R., and Edward C. Prescott, 2001, Taxes, regulations, and asset prices, Working paper, Federal Reserve Bank of Minneapolis.

Mehra, Rajnish, 1998, On the volatility of stock prices: An exercise in quantitative theory, *International Journal of Systems Science* 29, 1203–1211.

Mehra, Rajnish, and Edward C. Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

Mehra, Rajnish, and Edward C. Prescott, 1988, The equity premium: A solution? *Journal of Monetary Economics* 22, 133–136.

Mehra, Rajnish, and Edward C. Prescott, 2002, The equity premium in retrospect, in George M. Constantinides, Milton Harris, and Rene Stulz, eds.: *Handbook of the Economics of Finance* (North-Holland Elsevier, Amsterdam), forthcoming.

Menzly, Lior, Tano Santos, and Pietro Veronesi, 2001, Habit formation and the cross section of stock returns, Working paper, University of Chicago.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Negishi, Takashi, 1960, Welfare economics and existence of an equilibrium for a competitive economy, *Metroeconomica* 12, 92–97.

Pollak, Robert A., 1970, Habit formation and dynamic demand functions, *Journal of Political Economy* 78, 745–763.

Poterba, James M., 2001, The rise of the "equity culture": U.S. stockownership patterns, 1989–1998, Working paper, MIT.

Rietz, Thomas A., 1988, The equity risk premium: A solution, *Journal of Monetary Economics* 22, 117–131.

Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics* 7, 407–425.

Ryder, Harl E., and Geoffrey M. Heal, 1973, Optimum growth with intertemporally dependent preferences, *Review of Economic Studies* 40, 1–43.

Shiller, Robert, 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457–498.

Shiller, Robert, 2000, *Irrational Exuberance* (Princeton University Press, Princeton, NJ).

Siegel, Jeremy J., 1998, *Stocks for the Long Run* (McGraw-Hill, New York).

Siegel, Jeremy J., 1999, The shrinking equity premium, *Journal of Portfolio Management* 26, 10–17.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, 2000, Consumption and risk sharing over the life cycle, Working paper, Carnegie Mellon University.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, 2001, Asset pricing with idiosyncratic risk and overlapping generations, Working paper, Carnegie Mellon University.

Sundaresan, Suresh M., 1989, Intertemporally dependent preferences and the volatility of consumption and wealth, *Review of Financial Studies* 2, 73–89.

Vissing-Jorgensen, Annette, 2002, Limited asset market participation and the elasticity of intertemporal substitution, *Journal of Political Economy* 110, forthcoming.

Vuolteenaho, Tuomo, 2000, Understanding the aggregate book-to-market ratio and its implications to current equity-premium expectations, Working paper, Harvard University.

Wachter, Jessica, 2001, Habit formation and returns on bonds and stocks, Working paper, New York University.

Weil, Philippe, 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401–421.

Welch, Ivo, 2001, The equity premium consensus forecast revisited, Working paper, Yale University.

FAJ
Financial Analysts Journal
Volume 66 • Number 1
©2010 CFA Institute

Economic Growth and Equity Investing

Bradford Cornell

The performance of equity investments is inextricably linked to economic growth. Nonetheless, few studies on investing have explicitly taken research on economic growth into account. This study bridges that gap by examining the implications for equity investing of both theoretical models and empirical results from growth theory. The study concludes that over the long run, investors should anticipate real returns on common stock to average no more than about 4 percent.

The performance of equity investments is inextricably linked to economic growth. Earnings, the source of value for equity investments, are themselves driven by economic activity. Unless corporate profits rise as a percentage of GDP, which cannot continue indefinitely, earnings growth is constrained by GDP growth. This dynamic means that the same factors that determine the rate of economic growth also place bounds on earnings growth and, thereby, the performance of equity investments. Despite these well-known facts, few studies on equity investing have explicitly taken the literature on economic growth into account. This observation is not meant to imply that research connecting economic growth with equity returns is sparse. Numerous contributions in that area include several provocative pieces by Arnott and Bernstein (2002), Arnott and Asness (2003), and Bernstein and Arnott (2003). Nonetheless, rarely has this research been expressly tied to the literature on the theory of economic growth. By bridging that gap, further insight can be gained into the relationship between economic growth and equity returns and forecasts regarding future returns can be placed on a more solid foundation.

Economic Growth: Theory and Data

The focus of economic growth theory is explaining expansion in the standard of living as measured by real per capita GDP. In the neoclassical model of economic growth, originally developed by Solow (1956), per capita GDP growth over the long run is entirely attributable to exogenous technological innovation.[1] This conclusion may surprise those not steeped in growth theory, given the intuitive thinking that output per capita can always be increased by simply adding more capital. Although adding capital does increase output per capita, it does so at a declining rate. Consequently, rational producers stop adding capital when the marginal product of capital drops to its marginal cost. When the economy reaches that point, it is said to be in a steady state. Once the economy reaches the steady state growth path, the ratio of capital to labor (C/L) remains constant and per capita GDP growth ceases unless the production function changes so as to increase the marginal product of capital.

The source of change in the production function is technological innovation. By increasing the marginal product of capital, technological progress breaks the deadlock imposed by diminishing returns and makes further growth in per capita output profitable. So long as the technological innovation continues, so too does the growth in per capita GDP.

This conclusion is not limited to such early models as Solow's, in which the rate of technological change is exogenous. Following Romer (1990), a variety of growth models have been developed in which the amount of investment in R&D—and thus the rate of technological progress—is endogenous. Even in these more sophisticated models, however, the declining marginal product of capital ensures that long-run per capita growth is bounded by the rate of technological progress. The word "bounded" is important because the ability of a society to exploit modern technology effectively is not a foregone conclusion. For example, from 1960 to 2005, all the countries of sub-Saharan Africa, with the exception of South Africa, experienced little or no growth. This failure of certain poor countries to grow is one of the fundamental mysteries of economics, but it is not a relevant consideration here.[2] Virtually the entire global stock

Bradford Cornell is professor of financial economics at California Institute of Technology, Pasadena.

 ©2010 CFA Institute

market capitalization is concentrated in a relatively few highly developed countries. For those countries, the impediments to effective adoption of technology have proved to be minor, at least to date.

Before turning to the data on economic growth, I need to address one remaining issue. The conclusion that growth is attributable exclusively to technological innovation is based on the assumption that the economy has reached the steady state. If the capital stock is below the steady state—and thus the marginal product of capital exceeds its marginal cost—room still exists for the deepening of capital. In that situation, a country's growth rate can exceed the steady state growth rate because it is spurred by capital deepening, as well as by technological innovation. As C/L rises toward its steady state value, the growth rate converges to the steady state level that is attributable to technological change.

The capital stock of a country may be below its steady state level for a variety of reasons. An obvious example is warfare. Another is the opening of a previously closed society. Whatever the reason, growth theory predicts that a country with a C/L below the steady state level will grow more rapidly during a period of capital deepening. Growth theorists refer to this "catch-up" as convergence.

Convergence is important to bear in mind when analyzing historical growth rates with the goal of forecasting future growth. If the historical sample includes growth rates of countries that are in the process of converging to a steady state, the historical growth rates will exceed the future rates that will apply once the steady state has been achieved.

Convergence also helps explain why long-run growth rates for a particular country are remarkably constant. To illustrate, **Figure 1** plots the log of real per capita GDP in the United States from 1802 through 2008. The long-run average growth rate of 1.8 percent is also shown. Over this period, even the largest downturns (associated with the U.S. Civil War and the Great Depression) appear only as temporary dips in a remarkably smooth progression. That smooth progression is attributable in part to the fact that accelerations in economic growth, associated with capital accumulation, followed the dips, which were tied to a drop in the capital stock below its steady state level.

With that background, **Table 1** presents Barro and Ursúa's (2008) update of Maddison's (2003) compilation of information on world economic growth from 1923 to 2006. The starting point in Table 1 is 1923, the first year for which Barro and Ursúa had data for all the countries in their sample. Extending the sample backward for those countries with longer time series available does not affect the essential nature of the findings. Table 1 also reports growth rates for a shorter sample period (beginning in 1960) to take into account the possibility of nonstationarity in the data.

Figure 1. Logarithm of Real per Capita GDP, 1802–2008

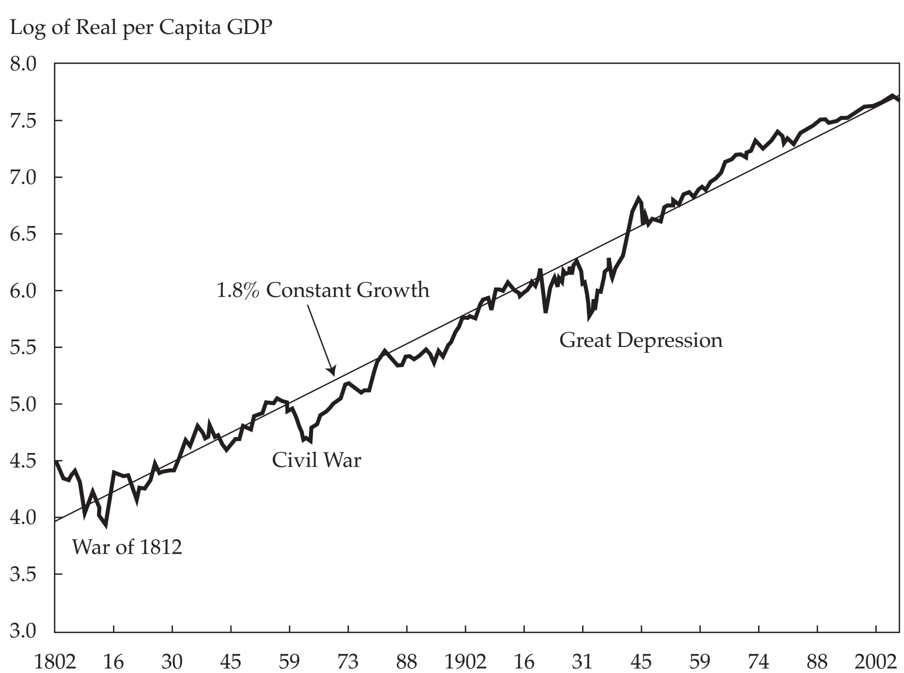


Table 1. Real Growth Rates in per Capita GDP, 1923–2006

Country	1923–2006	1960–2006
A. Mature Economies		
Australia	1.85%	2.16%
Austria	2.53	2.76
Belgium	2.11	2.62
Canada	2.22	2.27
Denmark	1.97	2.11
France	2.28	2.51
Germany	2.41	2.23
Italy	2.57	2.98
Japan	3.11	3.86
Netherlands	2.01	2.35
Spain	2.30	3.42
Sweden	2.50	2.25
Switzerland	1.63	1.51
United Kingdom	1.95	2.15
United States	1.42	1.14
Average	2.19%	2.42%
B. Developing and More Recently Developed Economies		
Argentina	1.10%	1.16%
Brazil	2.68	2.34
Chile	1.95	2.47
Colombia	2.18	2.24
Egypt	1.45	3.09
Finland	2.91	2.92
Greece	2.77	3.23
Iceland	3.24	2.87
India	1.74	2.88
Indonesia	1.81	3.08
S. Korea	3.55	5.72
Malaysia	1.91	2.14
Mexico	2.70	4.16
New Zealand	1.51	1.36
Norway	2.86	3.01
Peru	1.44	0.97
Philippines	1.32	1.46
Portugal	2.75	3.43
S. Africa	1.53	1.01
Singapore	3.33	5.72
Sri Lanka	1.93	3.06
Taiwan	3.78	6.24
Turkey	2.75	2.40
Uruguay	2.19	2.24
Venezuela	2.54	0.45
Average	2.32%	2.79%

Source: Barro and Ursúa (2008).

The results are reported in terms of compound growth rates. The following example illustrates why using compound growth rates is preferable to using averages of annual growth rates. Suppose that the ratio of corporate profits to GDP is stationary but not constant. In particular, assume (as the data will later show) that corporate profits are more variable than GDP. In that case, even though the compound growth rates of the two variables must converge in the long run, the arithmetic mean of annual growth rates for corporate profits will exceed that for GDP because of the variance effect.[3] The higher mean growth rate in earnings is illusory, however, because it fails to take into account the mean reversion in earnings growth that must occur for the ratio to be stationary.

The results reported in Table 1 are divided into two groups. The first group comprises mature economies that were already developed before World War II. These countries, which account for virtually the entire global stock market capitalization, are the focus of this study. The second group consists of economies that were developed more recently or are still considered developing. Results for the second group are presented for completeness and to provide perspective on the impact of convergence.

Consistent with the hypothesis that a common rate of technological advance is driving growth in all the developed countries, the results for the first group are remarkably homogeneous. Virtually all the growth rates for the full sample are close to the average of 2.19 percent. The exceptions are the United States, on the low end, and Japan, on the high end. The former's rate of 1.42 percent reflects the fact that the United States was the closest to steady state growth in 1923, after emerging from World War I relatively unscathed. The higher growth rate for Japan reflects convergence. At the start of the sample period, Japan was a relatively undeveloped country whose capital stock was below the steady state level. Convergence is also evident in the shorter sample period, beginning in 1960. The European countries and Japan, whose capital stocks were damaged in World War II, grew more rapidly than the United States, Switzerland, and Australia, all of which avoided war-related domestic destruction.

The results for the second group are more heterogeneous, reflecting the fact that growth in some countries (e.g., Peru and Venezuela) has stalled for reasons not fully understood whereas others (e.g., South Korea and Taiwan) have experienced rapid convergence. Despite the heterogeneity, however, the average growth rates of 2.32 percent for the sample period beginning in 1923 and 2.79 percent for the sample period beginning in 1960 are close to the averages for the first group of countries.

The averages reported in Table 1 are simple averages. If the growth rates for the first group of countries are weighted by market capitalization,

 ©2010 CFA Institute

the average falls to about 2 percent in both periods because of the predominant role of the United States. Giving the United States a higher weight is reasonable not only because of its large market capitalization but also because its economy is closest to steady state growth. Given the long period of time since World War II, to assume that all the countries in the first group will eventually converge to steady state growth is reasonable. Therefore, they are more likely to grow at rates comparable to the U.S. historical rate than at their own historical rates. This likelihood suggests that 2 percent real per capita growth, which exceeds the recent U.S. growth rate by 0.5 percent, is the most that investors can reasonably expect in the long run. Furthermore, although growth could be stalled by a catastrophe, such as another world war, the speed of technological innovation has proved almost impossible to accelerate meaningfully. In the remainder of this article, therefore, I will use 2 percent as the estimate of future per capita GDP growth. This number should be thought of as an achievable, but not necessarily expected, outcome.

In addition to the possibility of a catastrophe are two other reasons why 2 percent may prove to be an optimistic growth forecast. First, national income accounting does not deduct costs associated with pollution and environmental degradation in the calculation of GDP. Although these costs have been a tiny fraction of GDP in the past, concern that they are growing rapidly is widespread. If that concern is justified, properly accounting for these costs will reduce the future growth rate of per capita GDP. Second, whether the historical rate of technological innovation is sustainable is far from clear. Weil (2009, p. 260) noted that the rate of growth of real per capita GDP attributable to technological progress remained largely constant from 1950 to 2005, but over the same period, the number of researchers in the G–20 countries grew from 251,000 to 2.6 million. This finding suggests a declining marginal product of research as making and applying new discoveries become more difficult. If this trend continues, it could lead to falling rates of growth in per capita GDP.

Population Growth

Business opportunities depend on total economic activity, not per capita output. To see why, consider a hypothetical example of an economy for which technological innovation—and thus productivity growth—is zero but which is experiencing 5 percent population growth. Companies that provide goods and services in this economy will, on average, experience 5 percent growth in real revenues. Assuming that their margins remain constant, this rate translates into 5 percent growth in real earnings. Of course, in a dynamic economy, existing companies could lose business to start-ups, which could result in dilution for existing investors (which is a separate issue addressed later in the article). For companies in the aggregate, real earnings should be tied to real GDP, as data presented later in the article reveal to be the case.

Converting per capita growth to aggregate growth requires an estimate of population growth. Fortunately, population growth rates change even more slowly and are more predictable than growth rates of real per capita GDP.

Data on population growth for the sample countries are reported in **Table 2**. The first column presents historical growth rates from 2000 to 2007 taken from the U.S. Central Intelligence Agency's *2008 World Fact Book*. The second column presents United Nations (2007) forecasts of population growth rates from 2005 to 2010. That the two columns are very similar reflects the slowly changing nature of population growth.

The data in Table 2 are consistent with the widely documented fact that population growth is negatively correlated with per capita GDP.[4] The average population growth rate for the first group of countries is less than half that for the second group. Even for the second group, however, both the average historical growth rate and the average projected growth rate are less than 1 percent. Presumably, as per capita GDP continues to rise, these growth rates will continue to decline.

On the basis of the data presented in Table 2, population growth can be expected to add no more than 1 percent to the growth rate in per capita GDP. In fact, an assumption of a zero long-run future growth rate for the developed countries would not be unreasonable. Given real per capita growth of 2 percent, this assumption implies that investors cannot reasonably expect long-run future growth in real GDP to exceed 3 percent.

Earnings and GDP

The fundamental source of value for equity investors is earnings, not GDP. That long-run real GDP growth is reasonably bounded at 3 percent does not necessarily mean that the same is true of earnings, which depends on whether the ratio of earnings to GDP (E/GDP) is stationary. To test that hypothesis requires data on aggregate earnings.

Two primary measures of aggregate earnings are used in the United States. The first measure is derived from the national income and product accounts (NIPAs), produced by the U.S. Department

Table 2. Historical and Projected Population Growth Rates, 2000–2010

Country	Historical 2000–2007	Projected 2005–2010
A. Mature Economies		
Australia	1.22%	1.01%
Austria	0.06	0.36
Belgium	0.11	0.24
Canada	0.83	0.90
Denmark	0.30	0.90
France	0.57	0.49
Germany	–0.04	–0.07
Italy	0.00	0.13
Japan	–0.14	–0.02
Netherlands	0.44	0.21
Spain	0.10	0.77
Sweden	0.16	0.45
Switzerland	0.33	0.38
United Kingdom	0.28	0.42
United States	0.88	0.97
Average	0.34%	0.48%
B. Developing and More Recently Developed Economies		
Argentina	1.07%	1.00%
Brazil	1.23	1.26
Chile	0.91	1.00
Colombia	1.41	1.27
Egypt	1.68	1.76
Finland	0.11	0.29
Greece	0.15	0.21
Iceland	0.78	0.84
India	1.58	1.46
Indonesia	0.18	1.16
S. Korea	0.27	0.33
Malaysia	1.74	1.69
Mexico	1.14	1.12
New Zealand	0.97	0.90
Norway	0.35	0.62
Peru	1.26	1.15
Philippines	1.99	1.72
Portugal	0.31	0.37
S. Africa	0.83	0.55
Singapore	1.14	1.19
Sri Lanka	0.94	0.47
Taiwan	0.24	0.36
Turkey	1.01	1.26
Uruguay	0.49	0.29
Venezuela	1.50	1.67
Average	0.94%	0.96%

Sources: Central Intelligence Agency (2008) and the United Nations (2007).

of Commerce's Bureau of Economic Analysis. The NIPAs contain an estimate of aggregate corporate profits that is based on data collected from corporate income tax returns. The second measure of aggregate earnings is derived by Standard & Poor's from data collected from corporate financial reports. Because the two measures are not identical, distinguishing what is included in each measure before using the data is important.

The NIPA profit measure is designed to provide a time series of the income earned from the current production of all U.S. corporations. The sample is not limited to publicly traded companies. The tax rules on which the NIPAs are based are designed to expedite the timely and uniform completion of corporate tax returns. For that reason, all corporations use a highly uniform set of rules for tax accounting.

Because the NIPAs are designed to measure economic activity connected with current production, the NIPA definition of corporate profits includes only receipts arising from current production less associated expenses. The NIPA definition, therefore, excludes transactions that reflect the acquisition or sale of assets or liabilities. Dividend receipts from domestic corporations are excluded to avoid a double counting of profits. For the same reason, bad-debt expenses and capital losses are also excluded.

The Standard & Poor's estimate of aggregate earnings is derived from reported financial statements. Rather than being based on a unified set of tax rules, financial accounting is based on GAAP, which is designed to allow management to tailor financial statements so as to reveal information that is useful to a particular company. Furthermore, financial accounting provides for depreciation and amortization schedules that allow companies to attempt to match expenses with the associated stream of income.

The aggregate earnings data available from Standard & Poor's are for the companies in the S&P 500 Index. Each year's data consist of the aggregate GAAP after-tax earnings for the 500 companies in the S&P 500 for that year. Thus, the sample of companies in the aggregate is constantly changing as the index is updated. Because the S&P 500 earnings reflect a shifting sample of corporations, the series of reported earnings can be discontinuous over time. Fortunately, given the size of the index, these discontinuities are small and have little impact on estimated earnings growth.

The differences between financial and tax accounting create two dissimilarities between the measures of earnings for the same company.[5] First,

 ©2010 CFA Institute

intertemporal differences arise because of the timing of revenue, and expense recognition often differs between the two systems. The best example is depreciation because tax rules generally allow for more rapid depreciation than companies choose to report under GAAP. Second, permanent differences exist because the revenues and expenses recognized under the two systems are not the same. Although important in the short run, these differences tend to cancel out over long horizons, and thus, the long-run growth rates in the two measures are similar. For example, the average growth rate in NIPA real corporate profits from 1947 to 2008 was 3.23 percent, as compared with a growth rate of 3.17 percent in S&P 500 real aggregate earnings.

As an aid in examining the behavior of E/GDP, **Figure 2** plots after-tax corporate profits from the NIPAs as a fraction of GDP for 1947–2008. The figure reveals no overall trend. The fraction is approximately the same at the end as at the beginning, and thus, the growth rate of corporate profits is almost identical to that of GDP. The same is largely true of S&P 500 aggregate earnings as a fraction of GDP, which is plotted in **Figure 3** (normalized to start at 8.23 percent to facilitate comparison with Figure 2). The fraction for the S&P 500 earnings is smaller because the S&P 500 measure is less comprehensive than the NIPA measure. Unlike the NIPA data, the S&P 500 ratio exhibits a slight downward trend, reflecting the fact that as the economy has grown, the S&P 500 companies have become a progressively smaller fraction of total earnings. Therefore, the data are generally consistent with the hypothesis that over the long run, aggregate earnings are a stationary fraction of GDP. Certainly, no evidence exists of a persistent increase in the ratio, no matter which measure of earnings is chosen. This observation implies that the long-run growth rates of GDP place a limit on the long-run growth rates of earnings.

Although the data largely support the hypothesis that E/GDP is stationary, it is far from constant. Figure 2 shows that corporate profits vary between 3 percent and 11 percent of GDP. The variability of the ratio for S&P 500 earnings is even greater. This variability suggests that when earnings are low relative to GDP, they grow more quickly; the reverse is true when earnings are relatively high. This mean reversion in the growth rate of earnings maintains the stationarity of E/GDP.

Note that in an efficient market, the mean reversion in earnings growth would have no impact on stock returns because it would be impounded into current prices. Campbell and Shiller (1998), however, provided evidence that long-run average earnings are, in fact, predictive of future stock returns. Specifically, when the ratio of price to average earnings over the previous 10 years is high, future stock returns tend to be low; the reverse is true when the ratio is low. This finding suggests that the market does not fully account for the mean-reverting nature of long-run earnings growth.

Figure 2. Corporate Profits as a Percentage of GDP, 1947–2008

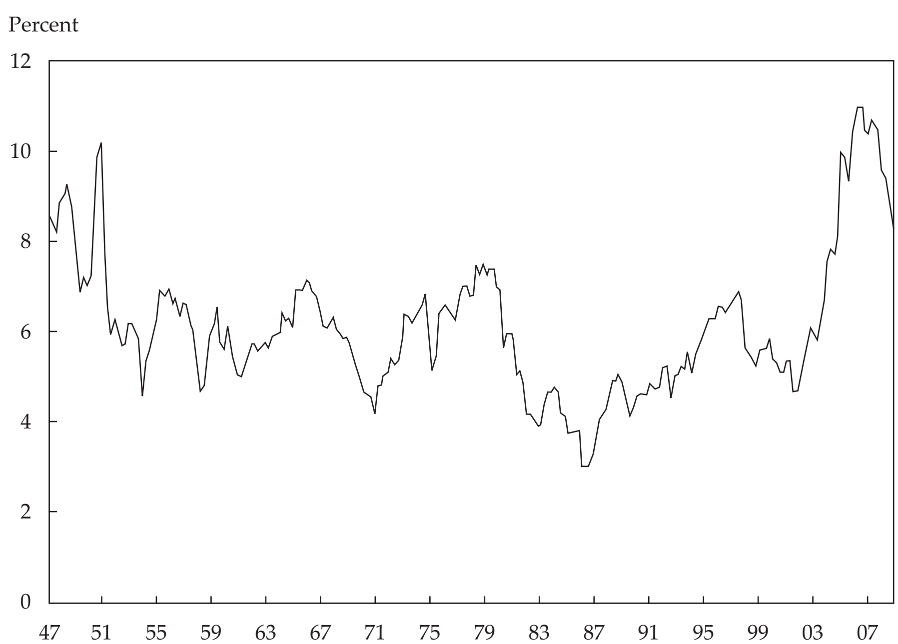
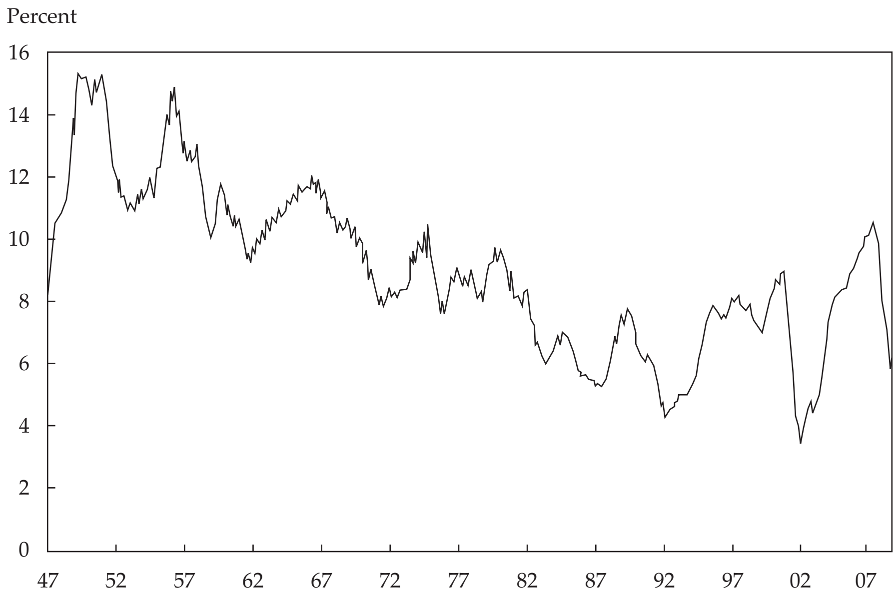


Figure 3. S&P 500 Earnings as a Percentage of GDP, 1947–2008



That the ratio of aggregate earnings to GDP is stationary implies that investors can expect aggregate real earnings growth to match, but not exceed, real GDP growth in the long run. Unfortunately, the same is not true of the earnings to which current investors have a claim. Two reasons explain this discrepancy. First, an investor's pro rata portion of a company's earnings will be affected by the company's share issuances and repurchases. If this dilution (or accretion) is ongoing, growth in aggregate earnings and earnings per share will diverge. Second and more important, current investors do not participate in the earnings of new businesses unless they dilute their current holdings to purchase shares in start-ups. Therefore, start-ups drive a wedge between the growth in aggregate earnings and the growth in the earnings to which current investors have a claim.

To illustrate the second effect, consider a simple example in which all companies in the economy are identical and earn $10 a share per period. Furthermore, assume that each company has a market value of $100 a share and has 1,000 shares outstanding. All earnings are paid out, so the values of the companies remain constant. Finally, assume that at the outset only two companies are in the economy, so aggregate earnings are $20,000. A current investor who holds 1 percent of each company has a pro rata share of aggregate earnings of $200. Now assume that the economy grows and a third company is started. As a result, aggregate earnings rise to $30,000, but the current investor does not participate in that growth and thus still holds 1 percent of the first two companies with rights to earnings of $200. If the current investor wanted to add the third company to the portfolio without investing new cash, the investor would have to dilute the portfolio's holdings in the first two companies. After the dilution, the investor would hold 0.67 percent of each of the three companies and would thus still have rights to earnings of $200. Therefore, the growth in earnings experienced by the current investor does not match the growth in aggregate earnings.

Bernstein and Arnott (2003) suggested an ingenious procedure for estimating the combined impact of both effects on the rate of growth of earnings to which current investors have a claim. They noted that total dilution on a marketwide basis can be measured by the ratio of the proportionate increase in market capitalization to the value-weighted proportionate increase in stock price. More precisely, net dilution for each period is given by the equation

$$\text{Net dilution} = \frac{1+c}{1+k} - 1, \tag{1}$$

where c is the percentage capitalization increase and k is the percentage increase in the value-weighted price index. Note that this dilution measure holds exactly only for the aggregate market portfolio. For narrower indices, the measure can be artificially affected if securities are added to or deleted from the index.

To account for the impact of dilution, the Bernstein–Arnott measure was estimated by using monthly data for the entire universe of CRSP stocks from 1926 to 2008. Using CRSP data for this purpose presents one problem. The CRSP universe was expanded twice during the sample period: in

 ©2010 CFA Institute

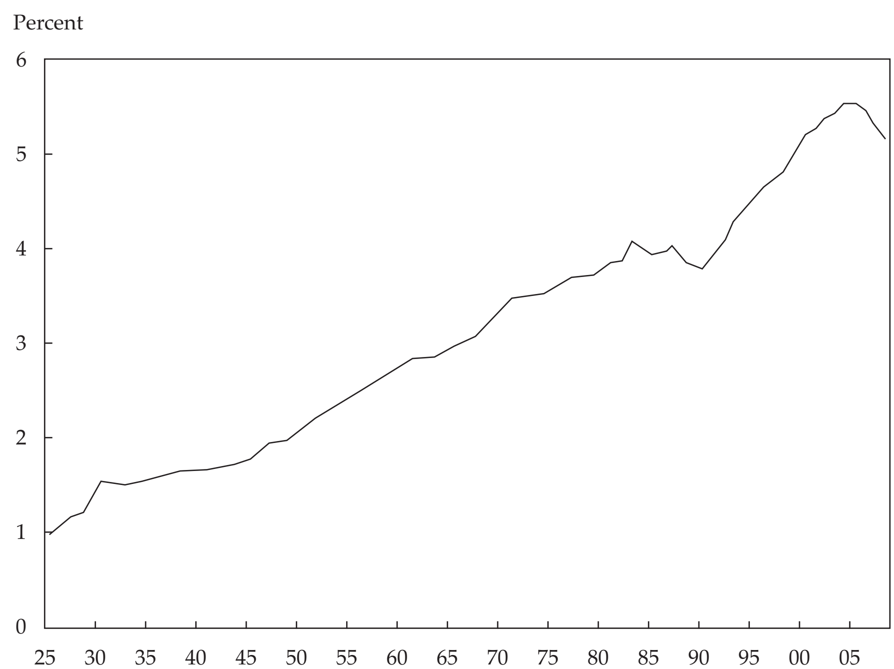
July 1962, when Amex stocks were added, and in July 1972, when NASDAQ stocks were added. Both these additions caused a significant increase in market capitalization unaccompanied by a corresponding increase in the value-weighted price. To eliminate the impact of these artificial discontinuities, I set the estimate of net dilution at zero for both July 1962 and July 1972.

Figure 4 plots the compounded estimate of net dilution from 1926 to 2008. It rises continuously except for downturns in the early 1990s and in 2006–2008. The average rate of dilution over the entire period is 2 percent. The primary source of dilution is the net creation of new shares as new companies capitalize their businesses with equity. The impact of start-ups is not surprising in light of the fact that more than half of U.S. economic growth comes from new enterprises, not from the growth of established businesses. Given the continuing importance of start-ups, the rate of dilution is highly unlikely to subside unless the rate of innovation slows. If the rate of innovation slows, however, GDP growth will also decline. Consequently, to conclude that the rate of growth of earnings, net of dilution, will remain largely constant is reasonable. Therefore, to estimate the growth rate of earnings to which current investors have a claim, approximately 2 percent must be deducted from the growth rate of aggregate earnings.

Putting the pieces together, we can see that growth theory predicts that current investors should count on long-run growth in real earnings of no more than 1 percent. This rate equals real growth of 3 percent in aggregate earnings, adjusted downward by 2 percent to account for dilution.

Arnott and Bernstein (2002) and Bernstein and Arnott (2003, p. 49) observed that "earnings and dividends grow at a pace very similar to that of per capita GDP." This observation correctly summarizes U.S. economic history, but it may not be true for other countries and it may not hold for the United States in the future. In terms of my analysis, the reason that earnings and dividends mirror per capita GDP is that population growth and dilution have both been about 2 percent between 1870 and 2008. Consequently, these two terms cancel each other out when we move from estimated growth in real per capita GDP to estimated growth in real earnings per share. But there is no theoretical reason why this cancellation should necessarily occur. For instance, population growth in Western Europe has fallen essentially to zero. If the United States were to follow suit but dilution were to continue at about 2 percent a year, growth in real earnings would be 2 percentage points *less* than growth in per capita GDP. In short, the Arnott–Bernstein observation is a shortcut that has historically held in the United States but is not a necessary condition. Therefore, a more complete analysis that takes into

Figure 4. The Impact of Dilution on Investor Earnings, 1926–2008



account both population growth and dilution is generally preferable. I do not present that analysis here because of limitations on dilution data for countries other than the United States.

Implications of Economic Growth Theory for Expected Stock Returns

The story thus far is that economic growth places a limit on the long-run growth of real earnings per share available to investors. On the basis of the data I have analyzed here, that limit is what many investors might consider a relatively anemic 1 percent. The next step is to explore the implications of that limitation for future returns on common stocks.

By definition, the rate of return on stock in period t is given by

$$R_t = \frac{D_t}{P_{t-1}} + GP_t, \tag{2}$$

where D_t is the dividend for year t, P_{t-1} is the price at the end of year $t - 1$, and $GP_t = (P_t - P_{t-1})/P_{t-1}$. Following Fama and French (2002), we can write Equation 2 in terms of long-run average values, denoted by A(), as

$$A(R_t) = A\left(\frac{D_t}{P_{t-1}}\right) + A(GP_t). \tag{3}$$

Equation 3 states that the long-run average return equals the average dividend yield plus the average capital gain.

Equation 3 holds *ex ante* as well as *ex post*. It implies that the long-run future average return equals the future average dividend yield plus the future average capital gain. Assuming that the earnings-to-price ratio is stationary, the long-run average earnings growth rate, $A(GE_t)$, can be substituted for the average capital gain rate, giving

$$A(R_t) = A\left(\frac{D_t}{P_{t-1}}\right) + A(GE_t). \tag{4}$$

My preceding analysis implies that $A(GE_t)$ in Equation 4 should be no more than about 1 percent in the future. In addition, as of December 2008, the current dividend yield was 3.1 percent and the previous 50-year average was 3.3 percent. Because the two are nearly equal, substituting either into Equation 4 as a proxy for the future average yield suggests that investors should not expect long-run real returns on common stocks to go much beyond 4 percent. Note that this calculation does not need to be adjusted for repurchases because the impact of repurchases is already accounted for in the dilution calculation. An adjustment is required only if future repurchases are expected to exceed their past average.

Equation 4 can also be used to approximate the equity risk premium. Because the real return on short-term government securities has averaged about 1 percent over the last 80 years, Equation 4 implies that the equity risk premium measured with respect to short-term government securities is approximately equal to the expected average dividend yield. Using either the current yield or the past average yield translates this number into a long-run average equity risk premium of just more than 3 percent. If the premium is measured with respect to longer-maturity government securities with greater expected real returns, the equity premium is commensurately less. This result is markedly less than the average historical risk premium measured over the 1926–2008 period that is commonly referenced. It is consistent, however, with a long-running body of empirical work that shows the *ex ante* risk premium to be significantly smaller than the historical average.[6]

Thus far, all the results have been stated in terms of compound growth rates. For many purposes, however, the object of interest is the annual expected return. For example, discounted-cash-flow valuations typically require annual estimates of the discount rate. To convert compound growth rates, which are geometric averages, into arithmetic averages requires taking the variance effect into account. This step can be well approximated by adding one-half of the annual variance of returns to the compound growth rate.

Because earnings are volatile, the variance effect adds about 1 percent to the compound growth rates. This result means that growth theory predicts that future annual real returns on common stocks should average no more than about 5 percent and that the annual equity risk premium for short-term government securities is about 4 percent.

Using annual data, we can tie the growth theory analysis to the long-run performance of company investments. If a company retains a fraction, b, of its earnings and invests those funds at a real rate of return, k, then basic finance theory teaches that the earnings per share will grow at the rate $(b)(k)$. Growth theory predicts that the annual long-run average growth in real earnings per share is about 2 percent, taking into account both dilution and the variance effect. From 1960 to 2008, companies in the S&P 500 retained, on average, 54 percent of their earnings. Solving for k, this retention ratio implies a real return on corporate investments of about 4 percent.

One possible adjustment might be made to the foregoing results. Recall that the dilution calculation was based on the assumption of a stable repurchase rate throughout the sample period. In fact,

 ©2010 CFA Institute

repurchases accelerated following the passage, in 1982, of U.S. SEC Rule 10b-18, which greatly reduced the legal risk associated with repurchases. More specifically, a pronounced trend toward repurchases as the preferred form of marginal payout to shareholders took place. Brav, Graham, Harvey, and Michaely (2005) reported that following the SEC ruling, managers began behaving as if a significant capital market penalty were associated with cutting dividends but not with reducing repurchases. Accordingly, dividends are set conservatively and repurchases are used to absorb variations in total payout. To the extent that this reliance on repurchases is expected to continue, the estimated 2 percent dilution effect might be too large and growth rates would have to be adjusted upward. Most of the 2 percent dilution, however, is associated not with the actions of existing companies but with start-ups that finance their businesses with new equity. Therefore, the adjustment in the overall rate of future dilution should not be large.

International Considerations

Thus far, I have limited my analysis to the United States. This restriction is an obvious shortcoming because most major corporations are becoming increasingly global. Although a detailed examination of international data is beyond the scope of this article, several general conclusions can be drawn. First, the data presented in Table 2 suggest that real per capita GDP growth rates for the other developed countries should be comparable to the U.S. growth rate in the future. Second, for the other developed countries, population growth rates are forecasted to be lower. As a result, the implied limitations on earnings growth remain largely unchanged and are perhaps even lower when other developed countries are included in the sample. Third, with respect to the developing countries—particularly India and China, which are the most important by virtue of their size—convergence predicts that they will experience higher growth rates in real per capita GDP than the United States. In addition, most developing countries are forecasted to have comparable or higher population growth rates than the United States. These forecasts suggest that companies doing business in the developing world will experience higher rates of earnings growth than they achieve in the developed world. Nonetheless, as those countries develop, both real GDP and population growth rates should decline. Furthermore, the fraction of total earnings attributable to business in the developing world is relatively small for most companies. Therefore, if a complete analysis were done on a global basis, the earnings bounds derived from U.S. data and the related predictions regarding stock returns would be unlikely to be markedly affected.

Conclusion

The long-run performance of equity investments is fundamentally linked to growth in earnings. Earnings growth, in turn, depends on growth in real GDP. This article demonstrates that both theoretical research and empirical research in development economics suggest relatively strict limits on future growth. In particular, real GDP growth in excess of 3 percent in the long run is highly unlikely in the developed world. In light of ongoing dilution in earnings per share, this finding implies that investors should anticipate real returns on U.S. common stocks to average no more than about 4–5 percent in real terms. Although more work needs to be done before equally definitive predictions can be made with respect to international equities, the basic outlook appears to be quite similar.

I thank Rob Arnott, Eugene Fama, Kenneth French, John Haut, John Hirshleifer, Jason Hsu, and Brian Palmer for helpful comments on earlier versions of this article. Data were graciously provided by Robert Barro and by Research Associates, LLC.

This article qualifies for 1 CE credit.

Notes

1. For details on the Solow model and more recent elaborations, see Barro and Sala-i-Martin (2004).
2. Hall and Jones (1999) described the problem in detail and offered an intriguing solution.
3. As a first-order approximation, the annual arithmetic mean equals the compound growth rate plus one-half the standard deviation of the annual growth rates.
4. See, for example, Weil (2009, ch. 4).
5. For further details on the relationship between reported earnings and NIPA profits, see Mead, Moulton, and Petrick (2004).
6. Contributions in this area include those of Rozeff (1984); Ross, Brown, and Goetzmann (1995); Claus and Thomas (2001); Fama and French (2002); and Cornell and Moroz (forthcoming).

References

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):70–87.

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal*, vol. 58, no. 2 (March/April):64–85.

Barro, Robert J., and Xavier Sala-i-Martin. 2004. *Economic Growth*. Cambridge, MA: MIT Press.

Barro, Robert J., and José F. Ursúa. 2008. "Macroeconomic Crises since 1870." NBER Working Paper 13940 (April).

Bernstein, William J., and Robert D. Arnott. 2003. "Earnings Growth: The Two Percent Dilution." *Financial Analysts Journal*, vol. 59, no. 5 (September/October):47–55.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely. 2005. "Payout Policy in the 21st Century." *Journal of Financial Economics*, vol. 77, no. 3 (September):483–527.

Campbell, John Y., and Robert J. Shiller. 1998. "Valuation Ratios and the Long-Run Stock Market Outlook." *Journal of Portfolio Management*, vol. 24, no. 2 (Winter):11–26.

Central Intelligence Agency. 2008. *The 2008 World Fact Book.* Washington, DC: Central Intelligence Agency.

Claus, James, and Jacob K. Thomas. 2001. "Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets." *Journal of Finance*, vol. 56, no. 5 (October):1629–1666.

Cornell, Bradford, and Max Moroz. Forthcoming. "The Equity Premium Revisited." *Journal of Portfolio Management*.

Fama, Eugene F., and Kenneth R. French. 2002. "The Equity Premium." *Journal of Finance*, vol. 57, no. 2 (April):637–659.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, vol. 114, no. 1 (February):83–116.

Maddison, Angus. 2003. *The World Economy: Historical Statistics*. Paris: OECD.

Mead, Charles Ian, Brent R. Moulton, and Kenneth Petrick. 2004. "NIPA Corporate Profits and Reported Earnings: A Comparison and Measurement Issues." Washington, DC: Bureau of Economic Analysis.

Romer, Paul M. 1990. "Endogenous Technical Change." *Journal of Political Economy*, vol. 98, no. S5 (October):S71–S102.

Ross, Stephen A., Stephen J. Brown, and William N. Goetzmann. 1995. "Survival." *Journal of Finance*, vol. 50, no. 3 (July):853–873.

Rozeff, Michael S. 1984. "Dividend Yields Are Equity Risk Premiums." *Journal of Portfolio Management*, vol. 11, no. 1 (Fall):68–75.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics*, vol. 70, no. 1 (February):65–94.

United Nations, Department of Economic and Social Affairs. 2007. "World Population Prospects: The 2006 Revision, Highlights." United Nations Working Paper No. ESA/P/WP.202 (www.un.org/esa/population/publications/wpp2006/WPP2006_Highlights_rev.pdf).

Weil, David N. 2009. *Economic Growth*. 2nd ed. Boston: Addison Wesley.

[ADVERTISEMENT]

©2010 CFA Institute

Historical Results II

Bradford Cornell

University of California
Los Angeles

The basic investment and constant-growth models, used with some justifiable simplifying assumptions about the U.S. market, indicate that the earnings growth rate cannot be greater than the GNP growth rate because of political forces and that the expected return, or cost of capital, in the long run should unconditionally be about 1.5 times the dividend-to-price ratio plus GNP growth. Adding reasonable assumptions about inflation produces a finding that equity risk premiums cannot be more than 3 percent (300 bps) because earnings growth is constrained by the real growth rate of the economy, which has been in the 1.5–3.0 percent range. In a consideration of today's market valuation, three reasons for the high market valuations seem possible: (1) stocks are simply seen as less risky, (2) valuation of equities is fundamentally determined by taxation, or (3) equity prices today are simply a mistake. A research question that remains and is of primary interest is the relationship between aggregate stock market earnings and GNP.

The very basic investment and constant-growth models from introductory finance courses can be used to interpret the long-run unconditional historical data on returns. So, let's begin with the basic model:

$$\frac{E_{t+1}}{E_t} = 1 + [(b)(ROE)],$$

where

E = earnings
b = the retention rate
ROE = return on equity

So that, with investment at time t denoted by I_t,

$$ROE = \frac{E_{t+1} - E_t}{I_t}$$

and

$$b = \frac{I_t}{E_t};$$

therefore, the growth rate of earnings is

$$(b)(ROE) = \frac{E_{t+1} - E_t}{E_t}.$$

This model implies that the growth rate in earnings is the retention rate times the return on equity, $(b)(ROE)$. In discussing the models, I would like to stress an important point: If you are interpreting the growth in earnings as being the retention rate times the return on equity, you have to be very careful when you are working with historical data. For example, does the retention rate apply only to dividends or to dividends and other payouts, such as share repurchases? The distinction is important because those proportions change in the more recent period. And if you make that distinction, you have to make a distinction between aggregate dividends and per share dividends because the per share numbers and the aggregate numbers will diverge. In working with the historical data, I have attempted to correct for that aspect.

Figure 1. S&P 500 Earnings and Dividends to GNP, 1950–July 2001

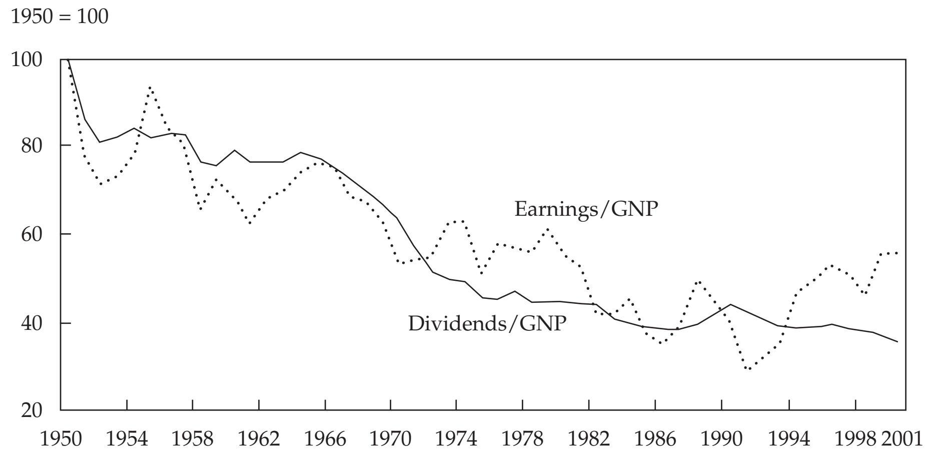


Table 1 gives the arithmetic average data for growth rates in GNP, earnings, and dividends for two periods: 1951–2000 and 1972–2000. (I used the 1972–2000 period because it mirrors the same period shown in Figure 1.) The earnings growth rates are so much more volatile than the dividend growth rates. And because of the volatility effect on arithmetic averages, GNP and earnings exhibit very similar growth rates from the early 1970s to the present. Dividends (and Table 1 shows the growth rate of actual dividends, not payouts) have grown much less than earnings for two reasons: First, dividends are less volatile, and second, dividend substitution is occurring. Corporations are not providing shareholders the same constant fraction of earnings (in the form of dividends) that they were in the past.

Despite the 1972–2000 data, it seems to me that earnings are not going to grow as fast as or faster than GNP in the future. This notion seems to be consistent with long-term historical data, and it fits my view of how politics works on the economy. If you accept that notion, it has immediate implications for the future.

Table 1. Historical Growth Rates of GNP, Earnings, and Dividends: Two Modern Periods

Period/Measure	GNP	Earnings	Dividends
1951–2000			
Mean	3.21%	2.85%	1.07%
Standard deviation	2.89	14.29	4.13
1972–2000			
Mean	2.62%	3.79%	0.96%
Standard deviation	2.94	15.72	3.58

Note: Growth rates for earnings and dividends are based on aggregate data.

First, under any reasonable underlying assumptions about inflation, equity risk premiums cannot be much more than 3 percent (300 bps) because the earnings growth rate is constrained unconditionally in the long run by the real growth rate of the economy, which has been in the range of 1.5–3.0 percent. Second, as **Table 2** shows, for an S&P level of about 1,000, you simply cannot have an equity risk premium any higher than 2 percent, 2.5 percent, or (at most) 3 percent.

Table 2. Value of the S&P 500 Index Given Various Real (Earnings or GNP) Growth Rates and Equity Risk Premiums

Real Growth Rate	Equity Risk Premium						
	2.0%	2.5%	3.0%	4.0%	5.0%	6.0%	7.0%
1.5%	845	724	634	507	423	362	317
2.0%	1,014	845	724	563	461	390	338
2.5%	1,268	1,014	845	634	507	423	362
3.0%	1,690	1,268	1,014	724	563	461	390

Assumptions: Inflation = 3 percent; long-term risk-free rate = 5.5 percent; payout = 1.5(S&P 500 dividend). The S&P 500 dividend used in the calculation was \$16.90, so $P = 1.5(\$16.90)/(k - g)$, where k = 5.5 percent (the risk-free rate minus 3 percent inflation plus the risk premium) and g = real growth rate.

What simplifying assumptions can be made to work with the unconditional data? I have made some relatively innocuous simplifying assumptions. First, that b should adjust until the cost of capital equals the ROE at the margin. To be very conservative, therefore, I will assume that the ROE equals the cost of capital, or expected returns, in the aggregate. The problem that arises is: What if the retention rate times the cost of capital (that is, the minimal expected return on equity), bk, is greater than GNP growth? The second assumption deals with this possibility: I assume bk cannot be greater than GNP growth because political forces will come into play that will limit the ROE if earnings start to rise as a fraction of GNP.

The relationship between aggregate earnings and GNP is one of the research questions that I have been unable to find interesting papers on—perhaps because I have not searched well enough—but I want to bring up the subject to this group. It seems to me that if aggregate earnings start to rise, and Robert Shiller mentioned several reasons why it can happen [see the "Current Estimates and Prospects for Change" session], then tax rates can change, antitrust regulation can change (one of Microsoft's problems probably was that it was making a great deal of money, which is an indication that some type of regulation may be necessary), labor regulation can change, and so forth. And these variables can change *ex post* as well as *ex ante*. So, once a company starts making superior returns using a particular technology, the government may step in *ex post* and limit those returns. The critical research question is how earnings relate to GNP.

The constant-growth model is

$$P = \frac{D}{k-g}$$

or

$$k = \frac{D}{p} + g,$$

where

P = price
D = dividends
k = cost of capital
g = growth rate

What I am going to do is just an approximation because I am going to work with aggregate, not per share, data. I am going to assume that total payouts are 1.5 times dividends.[1] Payouts will probably be lower in the future, but if I work with aggregate payouts, then g should be the growth rate in aggregate potential payouts, which I will characterize as earnings.

One of the implications of the simplifying assumptions I have made, and it relates to the data that Jeremy Siegel just produced ["Historical Results I"], is that the expected returns on stocks should be equal to the earnings-to-price ratio. (In the more complicated equations, you have situations in which the ROE is not exactly equal to expected returns, but for my long-run data, the simplifying assumption that earnings yield equals the expected ROE is fine.) So, with these assumptions,

$$\begin{aligned} P &= \frac{D}{k-g} \\ &= \frac{D}{k-bk} \\ &= 1-(b)\left(\frac{E}{1-b}\right)(k) \\ &= \frac{E}{k} \end{aligned}$$

or

$$k = \frac{E}{P}.$$

A further implication is that if g is constrained to be close to the growth of GNP, then it is reasonable to substitute GNP growth for g in the constant-growth model. The implication of this conclusion is that the expected return, or cost of capital, in the long run should unconditionally be about 1.5 times the dividend-to-price ratio plus GNP growth:

$$k = 1.5\frac{D}{P} + \text{GNP growth}.$$

With this background, we can now look at some of the data.

Earnings and GNP

Figure 1 allows a comparison of dividends/GNP and (after-tax) earnings/GNP for 1950 through July 2001.[2] The data begin in 1950 because Fama believed that the data before then were unreliable. Figure 1 shows that, historically, earnings have declined as a fraction of GNP in this period. My assumption that earnings keep up with GNP works from about 1970 on, but I am looking at the picture in Figure 1 in order to make that conclusion. The ratio of earnings to GNP depends on a lot of things: the productivity of labor, capital, the labor-to-capital ratio, taxes, and (as I said earlier) a host of political forces. Figure 1 shows that earnings have, at best, kept up with GNP.

[1] This choice is based on recent findings by Jagannathan, Stephens, and Weisbach (2000) that we are seeing significant payouts today.

[2] These data were provided by Eugene Fama, who attributed them to Robert Shiller.

Valuation

Why is the market so high? As an aside, and this concern is not directed toward our topic today of the equity risk premium, but I think it is an interesting question: Why is the market where it is today relative to where it was on September 10 or September 9 or just before the events of September 11, 2001? The market then and now is at about the same level. Almost every economist and analyst has said that the September 11 attacks accelerated a recession, that they changed perceptions of risk, and so forth. It is curious to me that such a situation does not seem to be reflected in market prices.

But in general, why is the market so high? I believe three possible explanations exist. One idea, and I consider it a "rational" theory, is that stocks are simply seen as less risky than in the past. I do not know whether the behavioral theories are rational or not, in the sense that prices are high because of behavioral phenomena that are real and are going to persist. If so, then those phenomena—as identified by Jeremy Siegel and Richard Thaler [see the "Theoretical Foundations" session]—are also rational. In that case, the market is not "too high"; it is not, in a sense, a mistake. It is simply reflecting characteristics of human beings that are not fully explained by economic theories.

Another rational explanation has been given less attention but is the subject of a recent paper by McGrattan and Prescott (2001). It is that the valuation of equities is fundamentally determined by taxation. McGrattan and Prescott argue that the move toward holding equities in nontaxable accounts has led to a drop in the relative tax rate on dividends. Therefore, stock prices should rise relative to the valuation of the underlying capital and expected returns should fall. This effect is a rational tax effect.

Both this theory and the theory that stocks are now seen as less risky say that the market is high because it should be high and that, looking ahead, equities are going to have low expected returns, or low risk premiums—about 2 percent—but that investors have nothing to worry about.

The final explanation, which I attribute to John Campbell and Robert Shiller, focuses on the view that equity prices today are simply a mistake. (I suppose mistakes are a behavioral phenomenon, but presumably, they are not as persistent as an underlying psychological condition.) Now, when people realize they have made a mistake, they attempt to correct the behavior. And those corrections imply a period of *negative* returns from the U.S. equity market before the risk premium can return to a more normal level.

Closing

To close, I want to repeat that, to me, the fundamental historical piece of data that needs more explanation is the relationship between the aggregate behavior of earnings and GNP—what it has been in the past and what it can reasonably be going forward. This relationship is interesting, and I look forward to hearing what all of you have to say about it. In my view, it is the key to unlocking the mystery of the equity risk premium's behavior.

FAJ
Financial Analysts Journal
Volume 66 • Number 1
©2010 CFA Institute

Economic Growth and Equity Investing

Bradford Cornell

The performance of equity investments is inextricably linked to economic growth. Nonetheless, few studies on investing have explicitly taken research on economic growth into account. This study bridges that gap by examining the implications for equity investing of both theoretical models and empirical results from growth theory. The study concludes that over the long run, investors should anticipate real returns on common stock to average no more than about 4 percent.

The performance of equity investments is inextricably linked to economic growth. Earnings, the source of value for equity investments, are themselves driven by economic activity. Unless corporate profits rise as a percentage of GDP, which cannot continue indefinitely, earnings growth is constrained by GDP growth. This dynamic means that the same factors that determine the rate of economic growth also place bounds on earnings growth and, thereby, the performance of equity investments. Despite these well-known facts, few studies on equity investing have explicitly taken the literature on economic growth into account. This observation is not meant to imply that research connecting economic growth with equity returns is sparse. Numerous contributions in that area include several provocative pieces by Arnott and Bernstein (2002), Arnott and Asness (2003), and Bernstein and Arnott (2003). Nonetheless, rarely has this research been expressly tied to the literature on the theory of economic growth. By bridging that gap, further insight can be gained into the relationship between economic growth and equity returns and forecasts regarding future returns can be placed on a more solid foundation.

Economic Growth: Theory and Data

The focus of economic growth theory is explaining expansion in the standard of living as measured by real per capita GDP. In the neoclassical model of economic growth, originally developed by Solow (1956), per capita GDP growth over the long run is entirely attributable to exogenous technological innovation.[1] This conclusion may surprise those not steeped in growth theory, given the intuitive thinking that output per capita can always be increased by simply adding more capital. Although adding capital does increase output per capita, it does so at a declining rate. Consequently, rational producers stop adding capital when the marginal product of capital drops to its marginal cost. When the economy reaches that point, it is said to be in a steady state. Once the economy reaches the steady state growth path, the ratio of capital to labor (C/L) remains constant and per capita GDP growth ceases unless the production function changes so as to increase the marginal product of capital.

The source of change in the production function is technological innovation. By increasing the marginal product of capital, technological progress breaks the deadlock imposed by diminishing returns and makes further growth in per capita output profitable. So long as the technological innovation continues, so too does the growth in per capita GDP.

This conclusion is not limited to such early models as Solow's, in which the rate of technological change is exogenous. Following Romer (1990), a variety of growth models have been developed in which the amount of investment in R&D—and thus the rate of technological progress—is endogenous. Even in these more sophisticated models, however, the declining marginal product of capital ensures that long-run per capita growth is bounded by the rate of technological progress. The word "bounded" is important because the ability of a society to exploit modern technology effectively is not a foregone conclusion. For example, from 1960 to 2005, all the countries of sub-Saharan Africa, with the exception of South Africa, experienced little or no growth. This failure of certain poor countries to grow is one of the fundamental mysteries of economics, but it is not a relevant consideration here.[2] Virtually the entire global stock

Bradford Cornell is professor of financial economics at California Institute of Technology, Pasadena.

 ©2010 CFA Institute

market capitalization is concentrated in a relatively few highly developed countries. For those countries, the impediments to effective adoption of technology have proved to be minor, at least to date.

Before turning to the data on economic growth, I need to address one remaining issue. The conclusion that growth is attributable exclusively to technological innovation is based on the assumption that the economy has reached the steady state. If the capital stock is below the steady state—and thus the marginal product of capital exceeds its marginal cost—room still exists for the deepening of capital. In that situation, a country's growth rate can exceed the steady state growth rate because it is spurred by capital deepening, as well as by technological innovation. As C/L rises toward its steady state value, the growth rate converges to the steady state level that is attributable to technological change.

The capital stock of a country may be below its steady state level for a variety of reasons. An obvious example is warfare. Another is the opening of a previously closed society. Whatever the reason, growth theory predicts that a country with a C/L below the steady state level will grow more rapidly during a period of capital deepening. Growth theorists refer to this "catch-up" as convergence.

Convergence is important to bear in mind when analyzing historical growth rates with the goal of forecasting future growth. If the historical sample includes growth rates of countries that are in the process of converging to a steady state, the historical growth rates will exceed the future rates that will apply once the steady state has been achieved.

Convergence also helps explain why long-run growth rates for a particular country are remarkably constant. To illustrate, **Figure 1** plots the log of real per capita GDP in the United States from 1802 through 2008. The long-run average growth rate of 1.8 percent is also shown. Over this period, even the largest downturns (associated with the U.S. Civil War and the Great Depression) appear only as temporary dips in a remarkably smooth progression. That smooth progression is attributable in part to the fact that accelerations in economic growth, associated with capital accumulation, followed the dips, which were tied to a drop in the capital stock below its steady state level.

With that background, **Table 1** presents Barro and Ursúa's (2008) update of Maddison's (2003) compilation of information on world economic growth from 1923 to 2006. The starting point in Table 1 is 1923, the first year for which Barro and Ursúa had data for all the countries in their sample. Extending the sample backward for those countries with longer time series available does not affect the essential nature of the findings. Table 1 also reports growth rates for a shorter sample period (beginning in 1960) to take into account the possibility of nonstationarity in the data.

Figure 1. Logarithm of Real per Capita GDP, 1802–2008

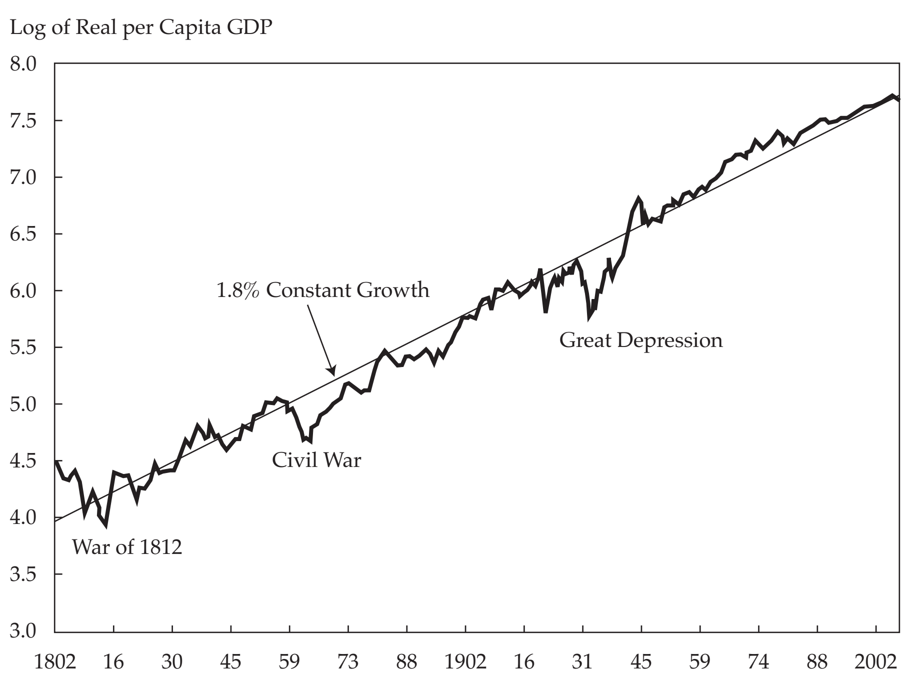


Table 1. Real Growth Rates in per Capita GDP, 1923–2006

Country	1923–2006	1960–2006
A. Mature Economies		
Australia	1.85%	2.16%
Austria	2.53	2.76
Belgium	2.11	2.62
Canada	2.22	2.27
Denmark	1.97	2.11
France	2.28	2.51
Germany	2.41	2.23
Italy	2.57	2.98
Japan	3.11	3.86
Netherlands	2.01	2.35
Spain	2.30	3.42
Sweden	2.50	2.25
Switzerland	1.63	1.51
United Kingdom	1.95	2.15
United States	1.42	1.14
Average	2.19%	2.42%
B. Developing and More Recently Developed Economies		
Argentina	1.10%	1.16%
Brazil	2.68	2.34
Chile	1.95	2.47
Colombia	2.18	2.24
Egypt	1.45	3.09
Finland	2.91	2.92
Greece	2.77	3.23
Iceland	3.24	2.87
India	1.74	2.88
Indonesia	1.81	3.08
S. Korea	3.55	5.72
Malaysia	1.91	2.14
Mexico	2.70	4.16
New Zealand	1.51	1.36
Norway	2.86	3.01
Peru	1.44	0.97
Philippines	1.32	1.46
Portugal	2.75	3.43
S. Africa	1.53	1.01
Singapore	3.33	5.72
Sri Lanka	1.93	3.06
Taiwan	3.78	6.24
Turkey	2.75	2.40
Uruguay	2.19	2.24
Venezuela	2.54	0.45
Average	2.32%	2.79%

Source: Barro and Ursúa (2008).

The results are reported in terms of compound growth rates. The following example illustrates why using compound growth rates is preferable to using averages of annual growth rates. Suppose that the ratio of corporate profits to GDP is stationary but not constant. In particular, assume (as the data will later show) that corporate profits are more variable than GDP. In that case, even though the compound growth rates of the two variables must converge in the long run, the arithmetic mean of annual growth rates for corporate profits will exceed that for GDP because of the variance effect.[3] The higher mean growth rate in earnings is illusory, however, because it fails to take into account the mean reversion in earnings growth that must occur for the ratio to be stationary.

The results reported in Table 1 are divided into two groups. The first group comprises mature economies that were already developed before World War II. These countries, which account for virtually the entire global stock market capitalization, are the focus of this study. The second group consists of economies that were developed more recently or are still considered developing. Results for the second group are presented for completeness and to provide perspective on the impact of convergence.

Consistent with the hypothesis that a common rate of technological advance is driving growth in all the developed countries, the results for the first group are remarkably homogeneous. Virtually all the growth rates for the full sample are close to the average of 2.19 percent. The exceptions are the United States, on the low end, and Japan, on the high end. The former's rate of 1.42 percent reflects the fact that the United States was the closest to steady state growth in 1923, after emerging from World War I relatively unscathed. The higher growth rate for Japan reflects convergence. At the start of the sample period, Japan was a relatively undeveloped country whose capital stock was below the steady state level. Convergence is also evident in the shorter sample period, beginning in 1960. The European countries and Japan, whose capital stocks were damaged in World War II, grew more rapidly than the United States, Switzerland, and Australia, all of which avoided war-related domestic destruction.

The results for the second group are more heterogeneous, reflecting the fact that growth in some countries (e.g., Peru and Venezuela) has stalled for reasons not fully understood whereas others (e.g., South Korea and Taiwan) have experienced rapid convergence. Despite the heterogeneity, however, the average growth rates of 2.32 percent for the sample period beginning in 1923 and 2.79 percent for the sample period beginning in 1960 are close to the averages for the first group of countries.

The averages reported in Table 1 are simple averages. If the growth rates for the first group of countries are weighted by market capitalization,

 ©2010 CFA Institute

the average falls to about 2 percent in both periods because of the predominant role of the United States. Giving the United States a higher weight is reasonable not only because of its large market capitalization but also because its economy is closest to steady state growth. Given the long period of time since World War II, to assume that all the countries in the first group will eventually converge to steady state growth is reasonable. Therefore, they are more likely to grow at rates comparable to the U.S. historical rate than at their own historical rates. This likelihood suggests that 2 percent real per capita growth, which exceeds the recent U.S. growth rate by 0.5 percent, is the most that investors can reasonably expect in the long run. Furthermore, although growth could be stalled by a catastrophe, such as another world war, the speed of technological innovation has proved almost impossible to accelerate meaningfully. In the remainder of this article, therefore, I will use 2 percent as the estimate of future per capita GDP growth. This number should be thought of as an achievable, but not necessarily expected, outcome.

In addition to the possibility of a catastrophe are two other reasons why 2 percent may prove to be an optimistic growth forecast. First, national income accounting does not deduct costs associated with pollution and environmental degradation in the calculation of GDP. Although these costs have been a tiny fraction of GDP in the past, concern that they are growing rapidly is widespread. If that concern is justified, properly accounting for these costs will reduce the future growth rate of per capita GDP. Second, whether the historical rate of technological innovation is sustainable is far from clear. Weil (2009, p. 260) noted that the rate of growth of real per capita GDP attributable to technological progress remained largely constant from 1950 to 2005, but over the same period, the number of researchers in the G–20 countries grew from 251,000 to 2.6 million. This finding suggests a declining marginal product of research as making and applying new discoveries become more difficult. If this trend continues, it could lead to falling rates of growth in per capita GDP.

Population Growth

Business opportunities depend on total economic activity, not per capita output. To see why, consider a hypothetical example of an economy for which technological innovation—and thus productivity growth—is zero but which is experiencing 5 percent population growth. Companies that provide goods and services in this economy will, on average, experience 5 percent growth in real revenues. Assuming that their margins remain constant, this rate translates into 5 percent growth in real earnings. Of course, in a dynamic economy, existing companies could lose business to start-ups, which could result in dilution for existing investors (which is a separate issue addressed later in the article). For companies in the aggregate, real earnings should be tied to real GDP, as data presented later in the article reveal to be the case.

Converting per capita growth to aggregate growth requires an estimate of population growth. Fortunately, population growth rates change even more slowly and are more predictable than growth rates of real per capita GDP.

Data on population growth for the sample countries are reported in **Table 2**. The first column presents historical growth rates from 2000 to 2007 taken from the U.S. Central Intelligence Agency's *2008 World Fact Book*. The second column presents United Nations (2007) forecasts of population growth rates from 2005 to 2010. That the two columns are very similar reflects the slowly changing nature of population growth.

The data in Table 2 are consistent with the widely documented fact that population growth is negatively correlated with per capita GDP.[4] The average population growth rate for the first group of countries is less than half that for the second group. Even for the second group, however, both the average historical growth rate and the average projected growth rate are less than 1 percent. Presumably, as per capita GDP continues to rise, these growth rates will continue to decline.

On the basis of the data presented in Table 2, population growth can be expected to add no more than 1 percent to the growth rate in per capita GDP. In fact, an assumption of a zero long-run future growth rate for the developed countries would not be unreasonable. Given real per capita growth of 2 percent, this assumption implies that investors cannot reasonably expect long-run future growth in real GDP to exceed 3 percent.

Earnings and GDP

The fundamental source of value for equity investors is earnings, not GDP. That long-run real GDP growth is reasonably bounded at 3 percent does not necessarily mean that the same is true of earnings, which depends on whether the ratio of earnings to GDP (E/GDP) is stationary. To test that hypothesis requires data on aggregate earnings.

Two primary measures of aggregate earnings are used in the United States. The first measure is derived from the national income and product accounts (NIPAs), produced by the U.S. Department

Table 2. Historical and Projected Population Growth Rates, 2000–2010

Country	Historical 2000–2007	Projected 2005–2010
A. Mature Economies		
Australia	1.22%	1.01%
Austria	0.06	0.36
Belgium	0.11	0.24
Canada	0.83	0.90
Denmark	0.30	0.90
France	0.57	0.49
Germany	–0.04	–0.07
Italy	0.00	0.13
Japan	–0.14	–0.02
Netherlands	0.44	0.21
Spain	0.10	0.77
Sweden	0.16	0.45
Switzerland	0.33	0.38
United Kingdom	0.28	0.42
United States	0.88	0.97
Average	0.34%	0.48%
B. Developing and More Recently Developed Economies		
Argentina	1.07%	1.00%
Brazil	1.23	1.26
Chile	0.91	1.00
Colombia	1.41	1.27
Egypt	1.68	1.76
Finland	0.11	0.29
Greece	0.15	0.21
Iceland	0.78	0.84
India	1.58	1.46
Indonesia	0.18	1.16
S. Korea	0.27	0.33
Malaysia	1.74	1.69
Mexico	1.14	1.12
New Zealand	0.97	0.90
Norway	0.35	0.62
Peru	1.26	1.15
Philippines	1.99	1.72
Portugal	0.31	0.37
S. Africa	0.83	0.55
Singapore	1.14	1.19
Sri Lanka	0.94	0.47
Taiwan	0.24	0.36
Turkey	1.01	1.26
Uruguay	0.49	0.29
Venezuela	1.50	1.67
Average	0.94%	0.96%

Sources: Central Intelligence Agency (2008) and the United Nations (2007).

of Commerce's Bureau of Economic Analysis. The NIPAs contain an estimate of aggregate corporate profits that is based on data collected from corporate income tax returns. The second measure of aggregate earnings is derived by Standard & Poor's from data collected from corporate financial reports. Because the two measures are not identical, distinguishing what is included in each measure before using the data is important.

The NIPA profit measure is designed to provide a time series of the income earned from the current production of all U.S. corporations. The sample is not limited to publicly traded companies. The tax rules on which the NIPAs are based are designed to expedite the timely and uniform completion of corporate tax returns. For that reason, all corporations use a highly uniform set of rules for tax accounting.

Because the NIPAs are designed to measure economic activity connected with current production, the NIPA definition of corporate profits includes only receipts arising from current production less associated expenses. The NIPA definition, therefore, excludes transactions that reflect the acquisition or sale of assets or liabilities. Dividend receipts from domestic corporations are excluded to avoid a double counting of profits. For the same reason, bad-debt expenses and capital losses are also excluded.

The Standard & Poor's estimate of aggregate earnings is derived from reported financial statements. Rather than being based on a unified set of tax rules, financial accounting is based on GAAP, which is designed to allow management to tailor financial statements so as to reveal information that is useful to a particular company. Furthermore, financial accounting provides for depreciation and amortization schedules that allow companies to attempt to match expenses with the associated stream of income.

The aggregate earnings data available from Standard & Poor's are for the companies in the S&P 500 Index. Each year's data consist of the aggregate GAAP after-tax earnings for the 500 companies in the S&P 500 for that year. Thus, the sample of companies in the aggregate is constantly changing as the index is updated. Because the S&P 500 earnings reflect a shifting sample of corporations, the series of reported earnings can be discontinuous over time. Fortunately, given the size of the index, these discontinuities are small and have little impact on estimated earnings growth.

The differences between financial and tax accounting create two dissimilarities between the measures of earnings for the same company.[5] First,

 ©2010 CFA Institute

intertemporal differences arise because of the timing of revenue, and expense recognition often differs between the two systems. The best example is depreciation because tax rules generally allow for more rapid depreciation than companies choose to report under GAAP. Second, permanent differences exist because the revenues and expenses recognized under the two systems are not the same. Although important in the short run, these differences tend to cancel out over long horizons, and thus, the long-run growth rates in the two measures are similar. For example, the average growth rate in NIPA real corporate profits from 1947 to 2008 was 3.23 percent, as compared with a growth rate of 3.17 percent in S&P 500 real aggregate earnings.

As an aid in examining the behavior of E/GDP, **Figure 2** plots after-tax corporate profits from the NIPAs as a fraction of GDP for 1947–2008. The figure reveals no overall trend. The fraction is approximately the same at the end as at the beginning, and thus, the growth rate of corporate profits is almost identical to that of GDP. The same is largely true of S&P 500 aggregate earnings as a fraction of GDP, which is plotted in **Figure 3** (normalized to start at 8.23 percent to facilitate comparison with Figure 2). The fraction for the S&P 500 earnings is smaller because the S&P 500 measure is less comprehensive than the NIPA measure. Unlike the NIPA data, the S&P 500 ratio exhibits a slight downward trend, reflecting the fact that as the economy has grown, the S&P 500 companies have become a progressively smaller fraction of total earnings. Therefore, the data are generally consistent with the hypothesis that over the long run, aggregate earnings are a stationary fraction of GDP. Certainly, no evidence exists of a persistent increase in the ratio, no matter which measure of earnings is chosen. This observation implies that the long-run growth rates of GDP place a limit on the long-run growth rates of earnings.

Although the data largely support the hypothesis that E/GDP is stationary, it is far from constant. Figure 2 shows that corporate profits vary between 3 percent and 11 percent of GDP. The variability of the ratio for S&P 500 earnings is even greater. This variability suggests that when earnings are low relative to GDP, they grow more quickly; the reverse is true when earnings are relatively high. This mean reversion in the growth rate of earnings maintains the stationarity of E/GDP.

Note that in an efficient market, the mean reversion in earnings growth would have no impact on stock returns because it would be impounded into current prices. Campbell and Shiller (1998), however, provided evidence that long-run average earnings are, in fact, predictive of future stock returns. Specifically, when the ratio of price to average earnings over the previous 10 years is high, future stock returns tend to be low; the reverse is true when the ratio is low. This finding suggests that the market does not fully account for the mean-reverting nature of long-run earnings growth.

Figure 2. Corporate Profits as a Percentage of GDP, 1947–2008

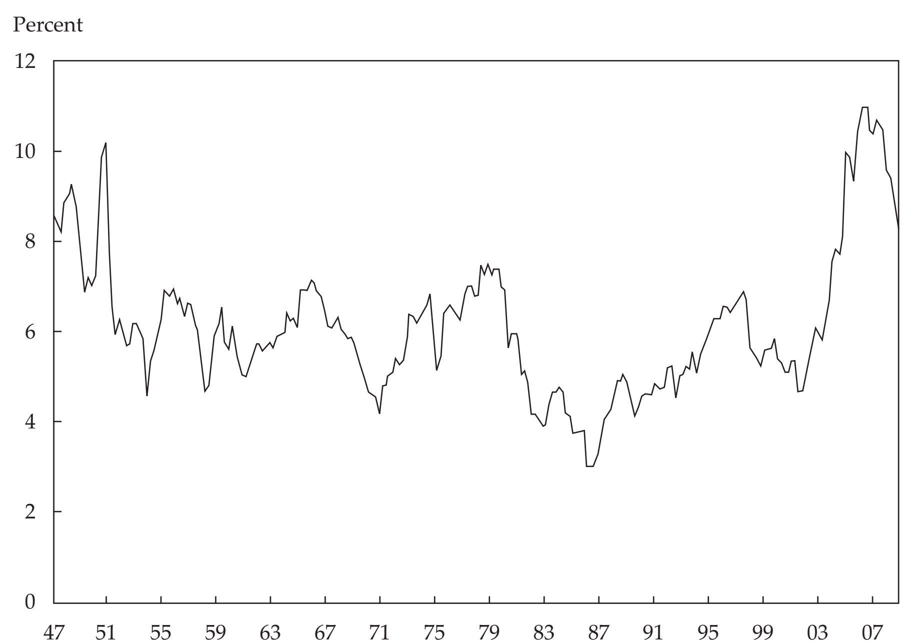
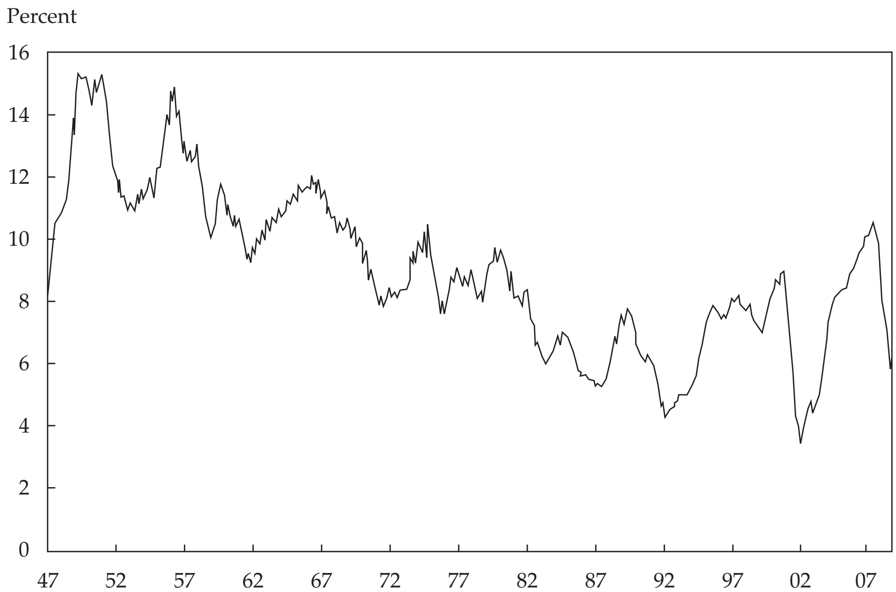


Figure 3. S&P 500 Earnings as a Percentage of GDP, 1947–2008



That the ratio of aggregate earnings to GDP is stationary implies that investors can expect aggregate real earnings growth to match, but not exceed, real GDP growth in the long run. Unfortunately, the same is not true of the earnings to which current investors have a claim. Two reasons explain this discrepancy. First, an investor's pro rata portion of a company's earnings will be affected by the company's share issuances and repurchases. If this dilution (or accretion) is ongoing, growth in aggregate earnings and earnings per share will diverge. Second and more important, current investors do not participate in the earnings of new businesses unless they dilute their current holdings to purchase shares in start-ups. Therefore, start-ups drive a wedge between the growth in aggregate earnings and the growth in the earnings to which current investors have a claim.

To illustrate the second effect, consider a simple example in which all companies in the economy are identical and earn \$10 a share per period. Furthermore, assume that each company has a market value of \$100 a share and has 1,000 shares outstanding. All earnings are paid out, so the values of the companies remain constant. Finally, assume that at the outset only two companies are in the economy, so aggregate earnings are \$20,000. A current investor who holds 1 percent of each company has a pro rata share of aggregate earnings of \$200. Now assume that the economy grows and a third company is started. As a result, aggregate earnings rise to \$30,000, but the current investor does not participate in that growth and thus still holds 1 percent of the first two companies with rights to earnings of \$200. If the current investor wanted to add the third company to the portfolio without investing new cash, the investor would have to dilute the portfolio's holdings in the first two companies. After the dilution, the investor would hold 0.67 percent of each of the three companies and would thus still have rights to earnings of \$200. Therefore, the growth in earnings experienced by the current investor does not match the growth in aggregate earnings.

Bernstein and Arnott (2003) suggested an ingenious procedure for estimating the combined impact of both effects on the rate of growth of earnings to which current investors have a claim. They noted that total dilution on a marketwide basis can be measured by the ratio of the proportionate increase in market capitalization to the value-weighted proportionate increase in stock price. More precisely, net dilution for each period is given by the equation

$$\text{Net dilution} = \frac{1+c}{1+k} - 1, \tag{1}$$

where c is the percentage capitalization increase and k is the percentage increase in the value-weighted price index. Note that this dilution measure holds exactly only for the aggregate market portfolio. For narrower indices, the measure can be artificially affected if securities are added to or deleted from the index.

To account for the impact of dilution, the Bernstein–Arnott measure was estimated by using monthly data for the entire universe of CRSP stocks from 1926 to 2008. Using CRSP data for this purpose presents one problem. The CRSP universe was expanded twice during the sample period: in

 ©2010 CFA Institute

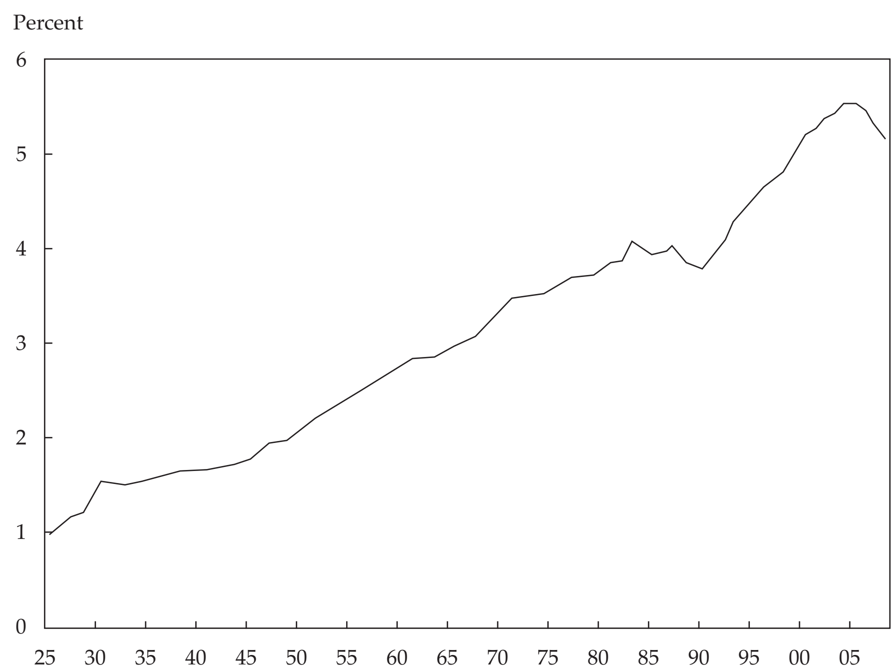
July 1962, when Amex stocks were added, and in July 1972, when NASDAQ stocks were added. Both these additions caused a significant increase in market capitalization unaccompanied by a corresponding increase in the value-weighted price. To eliminate the impact of these artificial discontinuities, I set the estimate of net dilution at zero for both July 1962 and July 1972.

Figure 4 plots the compounded estimate of net dilution from 1926 to 2008. It rises continuously except for downturns in the early 1990s and in 2006–2008. The average rate of dilution over the entire period is 2 percent. The primary source of dilution is the net creation of new shares as new companies capitalize their businesses with equity. The impact of start-ups is not surprising in light of the fact that more than half of U.S. economic growth comes from new enterprises, not from the growth of established businesses. Given the continuing importance of start-ups, the rate of dilution is highly unlikely to subside unless the rate of innovation slows. If the rate of innovation slows, however, GDP growth will also decline. Consequently, to conclude that the rate of growth of earnings, net of dilution, will remain largely constant is reasonable. Therefore, to estimate the growth rate of earnings to which current investors have a claim, approximately 2 percent must be deducted from the growth rate of aggregate earnings.

Putting the pieces together, we can see that growth theory predicts that current investors should count on long-run growth in real earnings of no more than 1 percent. This rate equals real growth of 3 percent in aggregate earnings, adjusted downward by 2 percent to account for dilution.

Arnott and Bernstein (2002) and Bernstein and Arnott (2003, p. 49) observed that "earnings and dividends grow at a pace very similar to that of per capita GDP." This observation correctly summarizes U.S. economic history, but it may not be true for other countries and it may not hold for the United States in the future. In terms of my analysis, the reason that earnings and dividends mirror per capita GDP is that population growth and dilution have both been about 2 percent between 1870 and 2008. Consequently, these two terms cancel each other out when we move from estimated growth in real per capita GDP to estimated growth in real earnings per share. But there is no theoretical reason why this cancellation should necessarily occur. For instance, population growth in Western Europe has fallen essentially to zero. If the United States were to follow suit but dilution were to continue at about 2 percent a year, growth in real earnings would be 2 percentage points *less* than growth in per capita GDP. In short, the Arnott–Bernstein observation is a shortcut that has historically held in the United States but is not a necessary condition. Therefore, a more complete analysis that takes into

Figure 4. The Impact of Dilution on Investor Earnings, 1926–2008



account both population growth and dilution is generally preferable. I do not present that analysis here because of limitations on dilution data for countries other than the United States.

Implications of Economic Growth Theory for Expected Stock Returns

The story thus far is that economic growth places a limit on the long-run growth of real earnings per share available to investors. On the basis of the data I have analyzed here, that limit is what many investors might consider a relatively anemic 1 percent. The next step is to explore the implications of that limitation for future returns on common stocks.

By definition, the rate of return on stock in period t is given by

$$R_t = \frac{D_t}{P_{t-1}} + GP_t, \tag{2}$$

where D_t is the dividend for year t, P_{t-1} is the price at the end of year $t - 1$, and $GP_t = (P_t - P_{t-1})/P_{t-1}$. Following Fama and French (2002), we can write Equation 2 in terms of long-run average values, denoted by A(), as

$$A(R_t) = A\left(\frac{D_t}{P_{t-1}}\right) + A(GP_t). \tag{3}$$

Equation 3 states that the long-run average return equals the average dividend yield plus the average capital gain.

Equation 3 holds *ex ante* as well as *ex post*. It implies that the long-run future average return equals the future average dividend yield plus the future average capital gain. Assuming that the earnings-to-price ratio is stationary, the long-run average earnings growth rate, $A(GE_t)$, can be substituted for the average capital gain rate, giving

$$A(R_t) = A\left(\frac{D_t}{P_{t-1}}\right) + A(GE_t). \tag{4}$$

My preceding analysis implies that $A(GE_t)$ in Equation 4 should be no more than about 1 percent in the future. In addition, as of December 2008, the current dividend yield was 3.1 percent and the previous 50-year average was 3.3 percent. Because the two are nearly equal, substituting either into Equation 4 as a proxy for the future average yield suggests that investors should not expect long-run real returns on common stocks to go much beyond 4 percent. Note that this calculation does not need to be adjusted for repurchases because the impact of repurchases is already accounted for in the dilution calculation. An adjustment is required only if future repurchases are expected to exceed their past average.

Equation 4 can also be used to approximate the equity risk premium. Because the real return on short-term government securities has averaged about 1 percent over the last 80 years, Equation 4 implies that the equity risk premium measured with respect to short-term government securities is approximately equal to the expected average dividend yield. Using either the current yield or the past average yield translates this number into a long-run average equity risk premium of just more than 3 percent. If the premium is measured with respect to longer-maturity government securities with greater expected real returns, the equity premium is commensurately less. This result is markedly less than the average historical risk premium measured over the 1926–2008 period that is commonly referenced. It is consistent, however, with a long-running body of empirical work that shows the *ex ante* risk premium to be significantly smaller than the historical average.[6]

Thus far, all the results have been stated in terms of compound growth rates. For many purposes, however, the object of interest is the annual expected return. For example, discounted-cash-flow valuations typically require annual estimates of the discount rate. To convert compound growth rates, which are geometric averages, into arithmetic averages requires taking the variance effect into account. This step can be well approximated by adding one-half of the annual variance of returns to the compound growth rate.

Because earnings are volatile, the variance effect adds about 1 percent to the compound growth rates. This result means that growth theory predicts that future annual real returns on common stocks should average no more than about 5 percent and that the annual equity risk premium for short-term government securities is about 4 percent.

Using annual data, we can tie the growth theory analysis to the long-run performance of company investments. If a company retains a fraction, b, of its earnings and invests those funds at a real rate of return, k, then basic finance theory teaches that the earnings per share will grow at the rate $(b)(k)$. Growth theory predicts that the annual long-run average growth in real earnings per share is about 2 percent, taking into account both dilution and the variance effect. From 1960 to 2008, companies in the S&P 500 retained, on average, 54 percent of their earnings. Solving for k, this retention ratio implies a real return on corporate investments of about 4 percent.

One possible adjustment might be made to the foregoing results. Recall that the dilution calculation was based on the assumption of a stable repurchase rate throughout the sample period. In fact,

 ©2010 CFA Institute

repurchases accelerated following the passage, in 1982, of U.S. SEC Rule 10b-18, which greatly reduced the legal risk associated with repurchases. More specifically, a pronounced trend toward repurchases as the preferred form of marginal payout to shareholders took place. Brav, Graham, Harvey, and Michaely (2005) reported that following the SEC ruling, managers began behaving as if a significant capital market penalty were associated with cutting dividends but not with reducing repurchases. Accordingly, dividends are set conservatively and repurchases are used to absorb variations in total payout. To the extent that this reliance on repurchases is expected to continue, the estimated 2 percent dilution effect might be too large and growth rates would have to be adjusted upward. Most of the 2 percent dilution, however, is associated not with the actions of existing companies but with start-ups that finance their businesses with new equity. Therefore, the adjustment in the overall rate of future dilution should not be large.

International Considerations

Thus far, I have limited my analysis to the United States. This restriction is an obvious shortcoming because most major corporations are becoming increasingly global. Although a detailed examination of international data is beyond the scope of this article, several general conclusions can be drawn. First, the data presented in Table 2 suggest that real per capita GDP growth rates for the other developed countries should be comparable to the U.S. growth rate in the future. Second, for the other developed countries, population growth rates are forecasted to be lower. As a result, the implied limitations on earnings growth remain largely unchanged and are perhaps even lower when other developed countries are included in the sample. Third, with respect to the developing countries—particularly India and China, which are the most important by virtue of their size—convergence predicts that they will experience higher growth rates in real per capita GDP than the United States. In addition, most developing countries are forecasted to have comparable or higher population growth rates than the United States. These forecasts suggest that companies doing business in the developing world will experience higher rates of earnings growth than they achieve in the developed world. Nonetheless, as those countries develop, both real GDP and population growth rates should decline. Furthermore, the fraction of total earnings attributable to business in the developing world is relatively small for most companies. Therefore, if a complete analysis were done on a global basis, the earnings bounds derived from U.S. data and the related predictions regarding stock returns would be unlikely to be markedly affected.

Conclusion

The long-run performance of equity investments is fundamentally linked to growth in earnings. Earnings growth, in turn, depends on growth in real GDP. This article demonstrates that both theoretical research and empirical research in development economics suggest relatively strict limits on future growth. In particular, real GDP growth in excess of 3 percent in the long run is highly unlikely in the developed world. In light of ongoing dilution in earnings per share, this finding implies that investors should anticipate real returns on U.S. common stocks to average no more than about 4–5 percent in real terms. Although more work needs to be done before equally definitive predictions can be made with respect to international equities, the basic outlook appears to be quite similar.

I thank Rob Arnott, Eugene Fama, Kenneth French, John Haut, John Hirshleifer, Jason Hsu, and Brian Palmer for helpful comments on earlier versions of this article. Data were graciously provided by Robert Barro and by Research Associates, LLC.

This article qualifies for 1 CE credit.

Notes

1. For details on the Solow model and more recent elaborations, see Barro and Sala-i-Martin (2004).
2. Hall and Jones (1999) described the problem in detail and offered an intriguing solution.
3. As a first-order approximation, the annual arithmetic mean equals the compound growth rate plus one-half the standard deviation of the annual growth rates.
4. See, for example, Weil (2009, ch. 4).
5. For further details on the relationship between reported earnings and NIPA profits, see Mead, Moulton, and Petrick (2004).
6. Contributions in this area include those of Rozeff (1984); Ross, Brown, and Goetzmann (1995); Claus and Thomas (2001); Fama and French (2002); and Cornell and Moroz (forthcoming).

References

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):70–87.

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal*, vol. 58, no. 2 (March/April):64–85.

Barro, Robert J., and Xavier Sala-i-Martin. 2004. *Economic Growth*. Cambridge, MA: MIT Press.

Barro, Robert J., and José F. Ursúa. 2008. "Macroeconomic Crises since 1870." NBER Working Paper 13940 (April).

Bernstein, William J., and Robert D. Arnott. 2003. "Earnings Growth: The Two Percent Dilution." *Financial Analysts Journal*, vol. 59, no. 5 (September/October):47–55.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely. 2005. "Payout Policy in the 21st Century." *Journal of Financial Economics*, vol. 77, no. 3 (September):483–527.

Campbell, John Y., and Robert J. Shiller. 1998. "Valuation Ratios and the Long-Run Stock Market Outlook." *Journal of Portfolio Management*, vol. 24, no. 2 (Winter):11–26.

Central Intelligence Agency. 2008. *The 2008 World Fact Book.* Washington, DC: Central Intelligence Agency.

Claus, James, and Jacob K. Thomas. 2001. "Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets." *Journal of Finance*, vol. 56, no. 5 (October):1629–1666.

Cornell, Bradford, and Max Moroz. Forthcoming. "The Equity Premium Revisited." *Journal of Portfolio Management*.

Fama, Eugene F., and Kenneth R. French. 2002. "The Equity Premium." *Journal of Finance*, vol. 57, no. 2 (April):637–659.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, vol. 114, no. 1 (February):83–116.

Maddison, Angus. 2003. *The World Economy: Historical Statistics*. Paris: OECD.

Mead, Charles Ian, Brent R. Moulton, and Kenneth Petrick. 2004. "NIPA Corporate Profits and Reported Earnings: A Comparison and Measurement Issues." Washington, DC: Bureau of Economic Analysis.

Romer, Paul M. 1990. "Endogenous Technical Change." *Journal of Political Economy*, vol. 98, no. S5 (October):S71–S102.

Ross, Stephen A., Stephen J. Brown, and William N. Goetzmann. 1995. "Survival." *Journal of Finance*, vol. 50, no. 3 (July):853–873.

Rozeff, Michael S. 1984. "Dividend Yields Are Equity Risk Premiums." *Journal of Portfolio Management*, vol. 11, no. 1 (Fall):68–75.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics*, vol. 70, no. 1 (February):65–94.

United Nations, Department of Economic and Social Affairs. 2007. "World Population Prospects: The 2006 Revision, Highlights." United Nations Working Paper No. ESA/P/WP.202 (www.un.org/esa/population/publications/wpp2006/WPP2006_Highlights_rev.pdf).

Weil, David N. 2009. *Economic Growth*. 2nd ed. Boston: Addison Wesley.

[ADVERTISEMENT]

 ©2010 CFA Institute

Copyright of Financial Analysts Journal is the property of CFA Institute and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Chapter 2

Evaluating the Historical Record

Primitive peoples, with no knowledge of modern science, express confidence in the proposition that the sun will rise tomorrow. The reason is that the historical record is unambiguous on this point. Ask whether it will rain tomorrow, though, and doubt arises. Because of random variation in weather, the historical record is a good deal more ambiguous. Rain today does not necessarily mean rain tomorrow.

With respect to the equity premium, the confidence that can be placed in the assumption that the future will be like the past depends on two related characteristics of the historical data: how accurately the historical premium can be measured and the extent to which the measured premium depends on the choice of the sample period. Before those questions can be addressed, however, there is the issue of how the average returns that go into the premium should be computed in the first place.

Computing the Average Premium: Arithmetic versus Geometric

The historical equity risk premium equals the difference between the average return on equities and the average return on treasury

securities calculated over a specified time period. It can be seen in Table 1.2, for instance, that over the full sample period between 1926 and 1997, the average return on stocks was 13.0% and the average return on treasury bills was 3.8%, so the equity risk premium over bills was 9.2%. Those are arithmetic averages. They are computed in the standard way: Add up all the annual returns and divide by the numbers of years (in this case, 72).

Although it is familiar, the arithmetic average has a peculiar property. As an illustration, suppose that an investor earns returns of 10%, 20%, –25%, and 15% in 4 consecutive years. The arithmetic average of the four returns is 5%. Now consider an investor who starts with \$100. If he or she earns 10%, 20%, –25%, and 15% in each of 4 years, his or her ending wealth will be \$113.85. However, if that investor earns 5% per year for 4 years, he or she will end up with \$121.55. This is a general problem. Investors who earn the arithmetic average of a series of returns wind up with more money than investors who earn the series of returns that are being averaged.

The geometric average solves this problem. By definition, the geometric average is the constant return an investor must earn every year to arrive at the same final value that would be produced by a series of variable returns. The geometric average is calculated using the formula

$$\text{Geometric Average} = (\text{Final Value}/\text{Initial Value})^{1/n} - 1$$

where n is the number of periods in the average. When the formula is applied to the preceding example, the results are as follows:

$$\text{Geometric Average} = (113.85/100)^{1/4} - 1 = 3.29\%$$

An investor who earns 3.29% for 4 years will end up with \$113.85.

There are four properties of arithmetic and geometric averages that are worth noting:

- The geometric average is always less than or equal to the arithmetic average. For instance, in Table 1.2 the arithmetic average stock return is 13.0%, but the geometric average is only 11.0%. (The geometric averages are reported at the bottom of the path of wealth columns in Table 1.2.)
- The more variable the series of returns, the greater the difference between the arithmetic and geometric average. For example, the returns for common stock are highly variable. As a result, the arithmetic average exceeds the geometric average by 200 basis points. For treasury bonds, whose returns are less variable, the difference between the two averages is only 40 basis points.
- For a given sample period, the geometric average is independent of the length of the observation interval.[1] The arithmetic average, however, tends to rise as the observation interval is shortened. For instance, the arithmetic average of monthly returns for the S&P 500 (calculated on an annualized basis by compounding the monthly arithmetic average) over the period between 1926 and 1997 is 13.1%, compared with the 13.0% average of annual returns.
- The difference between the geometric averages for two series does not equal the geometric average of the difference. Consider, for instance, stock returns and inflation. Table 1.2 reveals that the geometric average stock return is 11.0% and the average inflation rate is 3.1%, for a difference of 7.9%. However, Table 1.3 shows that the geometric average real return on common stock was 7.7%. This discrepancy does not arise for arithmetic averages, where the mean difference always equals the difference of the means.

With respect to the equity risk premium, the manner in which the average is calculated makes a significant difference. When compared with treasury bills over the full 1926-to-1997 period, the arithmetic average risk premium is 9.2%, whereas the geometric average premium is only 7.2%. Which average is the more appropriate choice? That depends on the question being asked. Assuming that the returns being averaged are largely independent and that the future is like the past, the best estimate of expected returns over a given future holding period is the arithmetic average of past returns over the same holding period. For instance, if the goal is to estimate future stock-market returns on a year-by-year basis, the appropriate average is the annual arithmetic risk premium. On the other hand, if the goal is to estimate what the average equity risk premium will be over the next 50 years, the geometric average is a better choice. Because the ultimate goal in this book is to arrive at reasonable forward-looking estimates of the equity risk premium, both arithmetic and geometric averages are employed where they are useful.

It is worth reiterating that projection of any past average is based on the implicit assumption that the future will be like the past. If the assumption is not reasonable, both the arithmetic and geometric averages will tend to be misleading.

How Accurately Can the Historical Risk Premium Be Measured?

The accuracy with which the historical risk premium can be measured depends on the variability of the observations from which the average is calculated. In an assessment of the impact of that variability, the best place to start is with an expanded version of Table 1.2 that includes monthly returns for the four asset classes over the period between 1926 and 1997. Given this expanded data set, one way to assess the variability of the ex-post risk premium, defined as the difference between the observed returns for stocks and the related treasury securities, is to plot one histogram for stocks versus bonds and another for stocks versus bills. Each bar on the histogram represents the fraction of the 864 monthly

[1] This follows immediately from the fact that the geometric average depends only on the initial and final values of the investment.

WILEY FRONTIERS IN FINANCE

THE EQUITY RISK PREMIUM

The Long-Run Future of the Stock Market

BRADFORD CORNELL

Chapter 6

The Equity Risk Premium and the Long-Run Outlook for Common Stocks

So that there is no suspense, here is the bottom line: The future will not be as bright as the past. The data of Ibbotson Associates showed that over the period from 1926 to 1997, the average equity risk premium was 7.4% over treasury bonds and 9.2% over treasury bills. Investors cannot reasonably expect equities to produce such large premiums going forward. Instead, premiums are much more likely to be on the order of 300 to 400 basis points lower. Reasonable forward-looking ranges for the future equity risk premium in the long run are 3.5% to 5.5% over treasury bonds and 5.0% to 7.0% over treasury bills.

This relatively pessimistic conclusion is based on two considerations. The first is an overall assessment of the empirical data and theoretical arguments presented in Chapters 1 through 4. The second is the analysis of the level of stock prices presented in Chapter 5. Although forecasting future stock returns, even over the long run, is hazardous at best, when all the evidence is taken into account, the conclusion that the future will be less rosy than the past has strong support.

Investment Banking Relationships and Analyst Affiliation Bias: The Impact of Global Settlement on Sanctioned and Non-Sanctioned Banks

Shane A. Corwin*
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556
scorwin@nd.edu

Stephannie Larocque
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556
larocque.1@nd.edu

Mike Stegemoller
Hankamer School of Business
Baylor University
Waco, TX 76798
michael_stegemoller@baylor.edu

January 2015

Abstract

We examine the impact of the Global Analyst Research Settlement on analyst affiliation bias in stock recommendations. Using a comprehensive measure of investment bank-firm relationships, including equity and debt underwriting and M&A advising, we find that affiliation bias is substantially reduced, but not eliminated, for analysts employed by banks named in the settlement. In contrast, we find strong evidence of analyst affiliation bias both before and after the Global Settlement for analysts at non-sanctioned banks. The results hold after controlling for shifts in the recommendation schemes used by investment banks and are robust to alternative empirical specifications.

JEL classification: G10, G24, G34, L14

Keywords: Analysts, Recommendations, Investment Banking, Investment Banking Relationships

* We thank Robert Battalio, Larry Brown, Gus De Franco, Marcus Kirk, Tim Loughran, and Hai Lu for helpful comments. Steven Carroll, Brian Ford, and Travis Johnson provided excellent research assistance. Any remaining errors are the responsibility of the authors.

Electronic copy available at: http://ssrn.com/abstract=2548771

1. INTRODUCTION

Conflicts of interest within investments banks and other financial institutions have been the subject of numerous academic studies (see Mehran and Stulz 2007 for a discussion). One particular conflict that has received significant attention from both regulators and academics is analyst affiliation bias. Specifically, prior research provides strong evidence that analysts are overly optimistic when their employers have equity underwriting relationships with the covered firms. Early in the 2000s, several attempts to reduce conflicts of interest were implemented in the securities industry, culminating in the 2003 Global Analyst Research Settlement (Global Settlement). In particular, a major purpose of the Global Settlement reached between the SEC, NYSE, NASD, New York Attorney General, and North American Securities Administrators Association and 12 of the largest investment banks was to reduce the conflicts of interest between the investment banking and research departments within the major banks.[1] Subsequent research suggests that investment banks changed their behavior following the Global Settlement[2], but provides little evidence on affiliation bias for analysts employed by sanctioned and non-sanctioned banks nor on relationships beyond the well-studied equity underwriting relationship. In this study, we use a broad measure of investment banking relationships, including equity and debt underwriting and mergers and acquisitions (M&A) advising, to examine analyst affiliation bias for a large sample of sanctioned and non-sanctioned investment banks (IBs) in the periods before and after the Global Settlement and contemporaneous regulatory changes.

Sell-side financial analysts provide buy/sell recommendations and earnings forecasts for a set of covered firms. In general, analysts are compensated and earn a reputation based on the quality of the information they provide. Despite these incentives to produce accurate information, however, analysts can also face pressure to issue optimistic or biased coverage. In particular, the financial services firms that employ analysts also compete for lucrative underwriting and M&A advisory mandates and may seek to

[1] The original settlement in April 2003 named ten investment banks, including Bear Stearns, CSFB, Goldman Sachs, Lehman Brothers, J.P. Morgan, Merrill Lynch, Morgan Stanley, Citigroup (Salomon Smith Barney), UBS Warburg, and U.S. Bancorp Piper Jaffray. Similar settlements with Deutsche Bank and Thomas Weisel were added later. We refer to these banks (including other name variations of the same banks) as "sanctioned" banks.
[2] See, for example, Kadan, Madureira, Wang, and Zach (2009).

Electronic copy available at: http://ssrn.com/abstract=2548771

use biased coverage as one means of winning potential clients. As a result, analysts face a conflict between their role in providing quality information to financial markets (and the associated reputational concerns) and the motivations of their employers to win future investment banking business.

Following prior research, we define an affiliated analyst as one whose employer also has an investment banking relationship with the covered firm. Existing research suggests that affiliated analysts tend to produce optimistic (i.e., upward biased) recommendations and earnings forecasts relative to unaffiliated analysts (see, for example, Dugar and Nathan 1995, Lin and McNichols 1998). This research focuses primarily on affiliation through equity underwriting relationships, with a particular emphasis on affiliation at the time of an equity issue.[3] However, equity underwriting is only one of many services that investment banks provide to firms. In the fourth quarter of 2013, for example, equity underwriting accounted for only 36% of total investment banking revenues at Goldman Sachs, compared to 34% for financial advising and 30% for debt underwriting. This suggests that investment banking relationships may have an impact beyond that evidenced through equity underwriting.[4]

To better understand the impact of investment banking relationships on analyst behavior, we examine the individual equity, debt, and M&A components of the relationship, as well as the overall investment banking relationship. We expect the results to be strongest for the overall relationship for two reasons. First, since equity, debt, and M&A transactions are discrete observations of the firm-bank relationship, viewing all of these transactions together allows us to observe the relationship at more points in time, better capturing the ongoing nature of the relationship. Second, we expect investment banking relationships that span multiple functional areas to put more pressure on analysts than narrow relationships.

To analyze affiliation bias, we study recommendations on a large sample of U.S. non-financial firms between 1998 and 2009 by analysts whose employers are either sanctioned investment banks or top

[3] One exception is Ljungqvist, Marston, Starks, Wei, and Yan (2007) who control for both equity and debt underwriting affiliations. This study is discussed in more detail below.

[4] The importance of firm-wide relationships may also change over time. For example, Corwin and Stegemoller (2014) find that the tendency of firms to use the same investment bank in multiple functional areas (i.e., equity underwriting, debt underwriting, or M&A advising) has increased significantly over time.

non-sanctioned banks. Our main variable of interest is the analyst's relative recommendation, defined as the difference between the analyst recommendation (with strong buy=5 and strong sell=1) and the median recommendation across all analysts covering the stock. Following Ljungqvist et al. (2007), we construct this variable at the end of each quarter, using the most recent recommendation by each analyst during the preceding twelve months. In our main tests, we regress this variable on proxies for investment banking relationships and a set of control variables shown in prior literature to have an association with analyst recommendations. Our primary relationship variable is an indicator variable equal to one if, during the prior three years, the firm hired the investment bank as a lead or co-manager on an equity or debt deal or as an advisor on an M&A transaction. However, we also provide tests using a continuous measure of relationships, defined as the proportion of a firm's total transaction value during a three-year window for which the investment bank acted as a lead manager, co-manager, or advisor. We define these relationship variables separately for equity, debt, and M&A transactions, as well as for the combined set of transactions across all types.

Consistent with prior research, we find strong evidence of analyst affiliation bias prior to the Global Settlement in 2003. For banks named in the Global Settlement (sanctioned banks), this bias is evident for all individual transaction types and for the overall relationship measure. For non-sanctioned banks in the period prior to the Global Settlement, we find mixed evidence of an affiliation bias based on individual transaction type relationship measures, but strong evidence of an affiliation bias based on the overall relationship measure. This evidence is consistent with our prediction that the overall measure better captures the ongoing nature of the investment banking relationship. The more striking results appear during the period following the Global Settlement. During this period, there remains evidence of an affiliation bias for sanctioned banks, but the bias is substantially reduced from the pre-Global Settlement effect. In contrast, non-sanctioned banks continue to exhibit strong analyst affiliation bias even after the Global Settlement. This bias is evident across all types of transactions and for the overall relationship measure. These results suggest that while the Global Settlement was successful at reducing analyst affiliation bias for the banks named in the settlement, conflicts of interest persist, especially for non-sanctioned investment banks.

Our results are robust to several alternative specifications and robustness checks. While our main results are based on relationship indicator variables, we find similar results based on continuous measures of relationships. The results are also robust to alternative fixed effects specifications, including firm, analyst, and investment bank fixed effects. Most importantly, our results are not driven by the shift of many investment banks from a five-tier to a three-tier recommendation scheme following the Global Settlement (Kadan et al. 2009). We find similar results when we repeat our analysis on a relative recommendation variable based on a three-tier recommendation scheme.

As an alternative specification, we use logistic regressions to examine the impact of investment banking relationships on the likelihood of issuing a buy or strong buy and the likelihood of issuing a sell or strong sell. Consistent with the relative recommendation results, this analysis suggests that prior to the Global Settlement, analysts at both sanctioned and non-sanctioned banks were significantly more likely to issue a buy or strong buy recommendation and significantly less likely to issue a sell or strong sell recommendation when affiliated with the firm through an investment banking relationship. After the Global Settlement, the bias for sanctioned banks is reduced, but remains significant. For non-sanctioned banks, the bias is significant both before and after the Global Settlement. For both groups of banks, the logit results suggest that a significant affiliation bias remains following the Global Settlement, with the effect being substantially larger for non-sanctioned banks.

As a final test, we examine whether incorporating lending data has an impact on the measurement of analyst affiliation bias. We find only weak evidence that lending relationships have an incremental effect on the measurement of analyst affiliation bias. Thus, affiliation bias appears to be best captured through the equity, debt, and M&A relationships. We assert that an overall measure, incorporating equity underwriting, debt underwriting, and M&A advising, is better able to capture investment banking relationships and their effects than measures based on any one type of transaction.

In summary, our findings suggest that conflicts of interest within investment banks have not been completely eliminated by the Global Settlement and contemporary regulatory changes. Our results suggest that the Global Settlement reduced, but did not eliminate, analyst affiliation bias in recommendations from banks named in the Global Settlement. Further, for large banks not named in the Global Settlement, we find strong evidence of a continued affiliation bias in the post-Global Settlement period. This suggests that our findings are driven by the punitive and bank-specific requirements imposed by the Global Settlement, rather than the broader regulatory changes that accompanied the settlement.

The remainder of the paper is organized as follows. Section 2 summarizes the literature related to analyst affiliation bias, provides background information on the Global Settlement, and describes our main hypothesis. In Section 3, we describe our data and sample construction. Section 4 presents our main results related to analyst affiliation bias and Section 5 examines the incremental impact of lending relationships. Section 6 concludes.

2. BACKGROUND AND HYPOTHESIS DEVELOPMENT

2.1. Analyst Affiliation Bias

Sell-side financial analysts have been widely studied as proxies for the market's expectations. At the same time, however, analysts' recommendations, target prices, and forecasts have been shown to be optimistic (Beneish 1991; Bradshaw 2004; La Porta 1996). In particular, prior research provides strong evidence of a link between analyst optimism (or bias) and investment banking relationships between covered firms and the banks that employ analysts. Dugar and Nathan (1995) find that recommendations and earnings forecasts are more optimistic for analysts who also have an investment banking relationship with the covered firm than for non-affiliated analysts and Lin and McNichols (1998) show that analysts employed by lead and co-managing underwriters issue growth forecasts and recommendations on the issuing firms that are significantly more favorable than those made by unaffiliated analysts. Further, Dechow, Hutton, and Sloan (2000) provide evidence that analysts employed by lead managers of equity offerings make more optimistic long-term growth forecasts around equity offerings and O'Brien, McNichols, and Lin (2005) conclude that investment banking relationships increase analysts' reluctance to reveal negative news.

Prior studies also point to factors that appear to mitigate analyst affiliation bias. Cowen, Groysberg, and Healy (2006) find that the bias is lower for bulge bracket investment banks than for lower-tier banks, suggesting that the reputational concerns of bulge bracket banks outweigh the benefits of biased analyst coverage. Ljungqvist et al. (2007) argue that, because analysts rely on institutional investors for trading commissions and ratings, they will be less likely to produce biased coverage on affiliated stocks that are also highly visible to institutional investors. Their results confirm that relative recommendations are negatively related to the presence of institutional investors.

Other research examines the impact of analyst bias on investors and the post-recommendation performance of covered firms. De Franco, Lu, and Vasvari (2007) examine the investor consequences of analysts' misleading behavior in the period prior to the Global Settlement. Using a sample of 50 firm-events identified in the Global Settlement in which analysts' private beliefs differed from their public disclosures, they provide evidence that these events are associated with selling by sophisticated investors and a wealth transfer from individuals to institutions. Michaely and Womack (1999) report that in the month following the post-IPO quiet period, affiliated analysts issue more buy recommendations for the IPO firm than do unaffiliated analysts, and the IPOs recommended by affiliated analysts substantially under-perform IPOs recommended by unaffiliated analysts. Similarly, Barber, Lehavy, and Trueman (2007) find that the "buy" and "strong buy" ratings of IB-employed analysts tend to underperform those of other analysts.

Research also examines whether analyst coverage affects the investment bank's ability to win future business from the covered firm. Bradshaw, Richardson, and Sloan (2006) surmise that all analysts bias their recommendations and forecasts in an attempt to win underwriting business. Ljungqvist, Marston, and Wilhelm (2006) find little evidence that optimistic analyst coverage affects an investment bank's likelihood of winning future lead underwriting mandates. However, Ljungqvist, Marston, and

Wilhelm (2009) show that optimistic analyst coverage does increase the likelihood of winning future co-managing appointments, which in turns leads to an increased likelihood of future lead mandates.

Existing research focuses primarily on affiliation through equity underwriting relationships. However, some recent research extends the analysis of affiliation bias to other areas. Ljungqvist et al. (2007) examine both equity and debt underwriting relationships and find that affiliation bias is stronger with respect to equity relationships. Kolasinski and Kothari (2006) investigate affiliation bias in analyst recommendations issued around M&A deals. They find that analysts affiliated with acquirer advisors upgrade acquirer stocks around M&A deals and target-affiliated analysts issue optimistic coverage on acquirers after exchange ratios (for all-stock deals) have been set.

2.2. The Global Settlement

During 2000, the securities industry attempted to reduce investment banking conflicts of interest, with the Securities Industries Association endorsing best practices around research and investment banking and the Association for Investment Management and Research (since renamed CFA Institute) releasing a white paper titled “Preserving the Integrity of Research.” In 2002, the Sarbanes-Oxley Act (SOX) amended the Securities and Exchange Act of 1934 with the creation of Section 15D, which required the NYSE and the NASD to adopt rules designed to address research analysts’ conflicts of interest. To comply with SOX, in 2002 the NYSE amended its Rule 351 (Reporting Requirement) and Rule 472 (Communication with the Public), while the NASD released Rule 2711 (Research Analysts and Research Report).[5] These rules were approved by the SEC in May 2002.

In 2001, following allegations of research tainted by investment banking conflicts of interest, the

[5] NYSE Rule 472 (Communication with the Public) requires that research reports be approved by a supervisory analyst, that research analysts not be subject to the supervision of any member of the investment banking department, that research analysts not purchase issuer securities prior to an IPO, that an IB not distribute research regarding an issuer 40 calendar days following an IPO offering in which the IB acted as a manager or co-manager, that an IB not issue a favorable research report in return for business, that analysts not receive compensation for investment banking business, and that the above be disclosed in the analyst’s research reports. NYSE Rule 351(f) requires an annual letter of attestation by the investment bank that it is in compliance with Rule 472. Similarly, NASD Rule 2711 (Research Analysts and Research Report) restricts relationships between investment banking and research departments and restricts the review of research reports by the subject company. It also prohibits analyst compensation based upon investment banking services, prohibits the promise of favorable research, imposes a 40 (10) day quiet period for research following an IPO (SEO), restricts personal trading by analysts in their covered stocks, and requires additional disclosures in research reports as well as additional supervisory procedures at the investment bank.

New York Attorney General began investigating Merrill Lynch and, subsequently, several other large investment banks. This investigation culminated in April 2003 with the Global Analyst Research Settlement reached by the SEC, NYSE, NASD, New York Attorney General, and North American Securities Administrators Association with ten of the largest investment banks – Bear Stearns, CSFB, Goldman, Lehman, J.P. Morgan, Merrill Lynch, Morgan Stanley, Citigroup (Salomon Smith Barney), UBS Warburg, and U.S. Bancorp Piper Jaffray (with Deutsche Bank and Thomas Weisel added later).[6] The Global Settlement required the payment of $875 million in penalties and disgorgement, $432.5 million to fund independent research, and $80 million to fund investor education. In addition, the settlement made numerous structural reforms including the physical separation of investment banking and research departments, the inability to compensate research analysts based upon investment banking revenues, and the prohibition of research analysts taking part in investment banking pitches and roadshows.

Subsequent research suggests that these regulatory changes affected the behavior of analysts within investment banks. Kadan, Madureira, Wang, and Zach (2009) find that the overall informativeness of recommendations (measured using absolute price reactions) declined following the Global Settlement. They also document that sanctioned banks shifted their stock recommendations from a 5-tier scale to a 3-tier scale. Barniv, Hope, Myring, and Thomas (2009) and Chen and Chen (2009) both document that the mapping between analysts' forecasts and target prices improved following the regulatory changes of the early 2000s. Clarke, Khorana, Patel, and Rau (2011) investigate market reactions to independent, affiliated, and unaffiliated analysts before and after the Global Settlement. They find that affiliated (independent) analysts issued fewer (more) strong buys following the settlement, with recommendation upgrades by affiliated analysts being more informative in the post-period. Moreover, Guan, Lu, and Wong (2012) find that forecasts by research firms are more optimistic than those of brokerage firms, syndicate firms, and investment banks following the regulatory changes in the early 2000s, but that forecast

[6] See http://www.sec.gov/news/press/2003-54.htm for the April 2003 press release and http://www.sec.gov/news/press/2003-144.htm for the SEC's October 2003 approval of Global Settlement.

accuracy and recommendation profitability for research firms are not significantly different from those of investment banks after the reforms.

Despite these behavior changes, there is some evidence that the Global Settlement may not have eliminated analyst affiliation bias. Using data from 1994 through 2008, Malmendier and Shanthikumar (2014) distinguish between strategic and non-strategic distortions in analyst behavior. Consistent with their expectations for strategic behavior, they find that affiliated analysts tend to issue more positive recommendations, but similar or more negative forecasts, than unaffiliated analysts.[7] In a recent survey of sell-side analysts, Brown, Call, Clement, and Sharp (2014) report that analysts view the generation of investment banking business as an important driver of their compensation and feel pressure from their research management to issue optimistic forecasts and/or recommendations. Recent actions by FINRA against Citigroup and Goldman Sachs also provide evidence of analyst involvement in IPO road shows and of analysts tipping selected clients, even after the Global Settlement.

2.3. Hypothesis

We contribute to the literature on analyst affiliation bias by examining the differential impact of the Global Settlement and contemporaneous regulatory changes on affiliation bias for sanctioned and non-sanctioned banks. We also provide a detailed analysis of the link between affiliation bias and the equity, debt, and M&A components of investment banking relationships. Our primary hypothesis is that analyst affiliation bias was eliminated following the Global Settlement. However, by separating sanctioned and non-sanctioned banks, we are able to examine two variations of this hypothesis. If the Global Settlement and concurrent regulatory changes imposed on the industry eliminated the conflicts of interest within investment banks that lead to analyst affiliation bias, we expect the bias to be eliminated for both sanctioned and non-sanctioned banks. However, if the principal effects of the Global Settlement

[7] Although not the main subject of our analysis, we also examined the relation between investment banking relationships and the bias and accuracy of analyst earnings forecasts. We define bias and accuracy by comparing each analyst's most recent forecast to actual earnings, where bias and accuracy are scaled by the standard deviation of forecasts across all analysts following the stock and normalized by subtracting the consensus (median) level of bias/accuracy. We find some evidence of optimistic forecasts by GS banks in the period prior to Global Settlement, but little evidence of a link between investment banking relationships and forecasts for GS banks in the post period or for non-sanctioned banks in either the pre or post period. We find little evidence of a consistent relation between analyst affiliation and forecast accuracy for either class of banks.

result from the punitive aspects or bank-specific requirements of the settlement, we expect affiliation bias to be eliminated only for sanctioned banks. We test these alternative versions of the hypothesis below.

3. Data and Sample Characteristics

To construct our sample, we use two main data sources. First, we use SDC to identify all equity, debt, and M&A activity by a large sample of U.S. firms, allowing us to measure the relationships between firms and their investment banks. Second, we use I/B/E/S data to identify the stock recommendations of sell-side analysts and the brokerage firms for which the analysts work. Together, these two datasets allow us to provide a detailed examination of the link between analyst recommendations and investment banking relationships both before and after the Global Settlement.

3.1. Sample Firms and Investment Banking Activity

We begin with the sample of all U.S. firms with listed common stock (CRSP share codes 10 or 11) between 1996 and 2009. After eliminating financials, utilities, and government agencies, the resulting sample includes 8,322 unique firms. For these firms, we then use the Securities Data Company (SDC) database to collect information on all public and private issues of equity and debt by the firm and any M&A transactions in which the firm is either the acquirer or the target. Firms are identified based on PERMCO in the CRSP data and based on CIDGEN in the SDC data. Firms are matched between the two databases using Cusip and, where possible, Ticker. To provide meaningful analysis of investment banking relationships, we exclude transactions for which either the transaction value or the identity of the underwriter/advisor is missing.

To identify affiliation through investment banking relationships, we focus on the most important investment banks in the sample. To identify these banks, we begin with the full sample of banks identified as lead or co-managing underwriters in the equity and debt samples or as advisors in the M&A sample.[8] We then compute market share ranks on an annual basis for each transaction type (equity, debt, and M&A). Finally, we compute each bank's average market share rank in each transaction type category

[8] Investment bank names are cleaned to eliminate multiple variations of the same investment bank name and to adjust for mergers and acquisitions among investment banks.

across all years during which the bank appears in the sample and limit our analysis to those investment banks with an average market share rank of 25 or higher in at least one transaction type category. In cases where one of the top 25 banks reflects the merger of two or more predecessor banks, all predecessor banks are also included. As shown in Table A2 in the Appendix, the resulting sample includes 57 different investment bank names during the sample period, with 48 active at the beginning of the sample period and 28 active at the end of the sample period.[9]

3.2. Analyst Recommendations

To test analyst affiliation bias, we focus on analyst stock recommendations, one of the analysts' primary and most visible outputs. We collect recommendations data, including the identity of the broker employing the analyst, from I/B/E/S. We then link the recommendations to the sample of CRSP firms using CUSIPs and hand-match the broker names in I/B/E/S to the sample investment banks using the I/B/E/S broker translation file.

Following Ljungqvist et al. (2007) we examine recommendations at a quarterly frequency. For each calendar quarter end and each firm in our sample, we select the most recent recommendation issued during the preceding 12 months by each analyst covering the stock. We code recommendations as 1 (strong sell) through 5 (strong buy). We then define each analyst's relative recommendation, *RelRec*, by subtracting the consensus (i.e., median) recommendation across all analysts covering the firm in the same one-year window.[10] Finally, we limit our sample to stocks covered by at least one analyst employed by a sample investment bank. The resulting sample includes 216,242 quarterly observations, involving 4,628 analysts and 5,111 sample stocks.

3.3 Variable Construction and Sample Characteristics

Our main empirical tests examine the relation between the relative recommendations of analysts

[9] For clarity following large investment bank mergers, we assign a new name to the combined bank. For example, we refer to the combination of Citibank and Salomon Smith Barney as Citigroup Salomon Smith Barney and the combination of UBS Warburg and Paine Webber as UBS Paine Webber. The 28 ultimate banks considered here compares to 16 studied in Ljungqvist et al. (2006) and Ljungqvist et al. (2007). Lehman and Merrill Lynch are eliminated from the sample because their recommendations are excluded from the I/B/E/S database for all or part of our sample period.

[10] In order to compute relative recommendations, our sample is restricted to firms that are followed by two or more analysts. As discussed in Section 3 below, we also provide robustness tests based on a redefined three-point recommendation scale. Our main conclusions are robust to this alternative specification.

(*RelRec*) and investment banking relationships between the analyst's firm and the covered stock, after controlling for firm, analyst, and investment bank characteristics that have been shown to affect recommendations. Our empirical model closely follows that in Ljungqvist et al. (2007), with several important differences. First, we examine investment banking relationships across a wider set of transaction types, including equity, debt, and M&A transactions. Second, we define relationships both within specific functional areas and across all functional areas. Finally, we examine affiliation bias both before and after the Global Settlement, allowing for differences between investment banks named in the Global Settlement and other banks. Table A1 in the Appendix contains all variable definitions.

Summary statistics for our sample of quarterly observations are provided in Panel A of Table 1. Consistent with previous research, we find that analysts primarily issue "buy" or "strong buy" recommendations, giving a mean (median) analyst recommendation across our sample of 3.6 (4.0). As noted earlier, our main variable of interest is the relative recommendation of the analyst (*RelRec*), defined as the difference between the analyst's recommendation and the consensus (i.e., median) recommendation across all analysts following the stock. *RelRec* has a range from -4 to +3, with a mean (median) of 0.0025 (0.0000) across our sample observations.

To proxy for investment banking relationships, we examine each firm's equity, debt, and M&A transactions during the 36 months preceding each quarter end. We then define relationship dummy variables (*IBRel*) for each investment bank-firm pair that equal one if the investment bank acted as lead or co-managing underwriter on an equity or debt issue, or as an advisor on an M&A transaction. While the majority of our tests are based on these relationship dummy variables, we also analyze continuous relationship variables based on the proportion of each firm's equity, debt, and M&A transaction value for which the bank acted as lead or co-managing underwriter, or advisor.

We define relationship measures both by transaction type (equity, debt, or M&A) and across all combined transactions (overall relationship).[11] We expect affiliation bias to be better captured by overall relationships than by type-specific relationships for two reasons. First, equity, debt, and M&A transactions are discrete measures of what is likely an ongoing relationship. Thus, the use of multiple transaction types will better capture the ongoing nature of any underlying relationship. Second, if there is any pressure placed on the analyst to produce optimistic coverage, then this pressure will only be magnified when the investment banking relationship spans multiple functional areas.

To illustrate the potential benefits of the overall relationship measure, Figure 1 plots the time series of relationships between Convergys Corp. and Citi-Salomon-Smith, based on 36-month windows. Convergys used this bank as a lead equity underwriter on their August 1998 IPO, as a lead debt underwriter in September 2000 and December 2004, and as an M&A advisor in April 2001. When we incorporate all three transaction types, we are able to capture the ongoing nature of the relationship between Convergys and Citi-Salomon-Smith over the entire period from 1998 through 2007. However, when we define relationships based on any individual transaction type (equity, debt, or M&A) the relationship measure is spotty and only covers sub-periods from August 1998 through December 2007.

Summary statistics for our type-specific and overall relationship measures are provided in the second section of Table 1. Across all quarterly observations, the mean transaction type-specific relationship ranges from 2.43% for M&A transactions to 3.24% for equity transactions. Incorporating all transaction types, the mean overall relationship is 5.90%. In untabulated results, we find that the proportion of quarterly observations with no relationship equals 87.2% for the overall relationship measure, compared to 93.5% for equity, 93.6% for debt, and 96.3% for M&A. This provides one indication that the overall relationship measure may better identify ongoing relationships in cases where type-specific relationship measures do not.

Our remaining control variables are motivated by prior literature and closely follow the specification in Ljungqvist et al. (2007). To control for investment bank characteristics, we define two

[11] For the overall relationship variable, we measure at each quarter end date the proportion of a firm's combined equity, debt, and M&A transaction value during the preceding 36 months for which each investment bank acted as lead underwriter, co-managing underwriter, or adviser, and an indicator variable for whether this value is greater than zero.

continuous variables and a set of indicator variables. We define investment bank size (*IB_Size*), as the number of analysts employed by the investment bank during quarter *t*, based on I/B/E/S recommendations.[12] Investment bank market share, *IB_MktShare*, is the proportion of total deal value across all firms during the previous 12 months for which the investment bank acted as a lead or co-managing underwriter or M&A advisor. Like the relationship measures, *IB_MktShare* is defined by transaction type (equity, debt, or M&A) and across all combined transactions (overall). As shown in Table 1, the mean (median) number of analysts employed by an investment bank is 89 (85) and investment bank market shares average 4.55%, 4.77%, and 4.38% for equity, debt, and M&A, respectively. We also define two indicator variables, *IB_GS* and *IB_NonGS*, to distinguish between those investment banks sanctioned in the Global Settlement (including subsequent name variations of the same banks) and other non-sanctioned banks, respectively. Based on this categorization, 57% of our quarterly observations are from sanctioned banks and 43% from non-sanctioned banks. Appendix Table A2 lists the sample investment banks in each category.

We define six analyst-level characteristics. Four of these variables are defined directly from the I/B/E/S recommendations data. *Seniority* is the number of years since the analyst first appeared in I/B/E/S and *Seasoning* is the number of years since the analyst initiated coverage on the particular stock. *NFollow* is the number of firms followed by the analyst during the quarter and *JobMove* is an indicator variable that equals 1 if the analyst changed employers during the quarter. Following Hong and Kubik (2003) and Ljungqvist et al. (2007), we define relative forecast accuracy (*RelAccuracy*) based on the analyst's average earnings forecast accuracy across all followed stocks.[13] Finally, *AllStar* is an indicator variable that equals 1 if the analyst is a ranked as an All-Star by *Institutional Investor* magazine during year *t-1*,

[12] Ljungqvist et al. measure investment bank size as the number of registered representatives employed by the IB.
[13] For each analyst following each firm, we first estimate the absolute value of the difference between the analyst's most recent forecast of fiscal-year earnings and actual earnings, scaled by prior year price. We then rescale such that the most accurate analyst following the firm scores 1 and the least accurate analyst scores 0. Finally, each analyst's relative forecast accuracy is defined as their mean score across all stocks followed over years *t*-2 through *t*. See Appendix Table A1 for a more complete description.

and 0 otherwise. For the mean (median) observation in our sample the analyst has seniority of 5.4 (4.9) years, seasoning of 2.3 (1.4) years, and follows 11 (10) stocks. The mean and median values of relative accuracy are 41.23% and 40.96%, respectively. Finally, 18.9% of the recommendation observations in our sample are issued by All-Star analysts and 3.2% by analysts that changed employers during the quarter.

Our last set of control variables is related to firm characteristics. *ANF* is the number of analysts issuing recommendations for the firm during the previous 12 months, based on I/B/E/S recommendations. *MV* is the firm's market value of equity at the end of the prior calendar year, as defined by CRSP. *InstHoldings* is the percentage of shares held by institutional investors at the end of the quarter, based on Thomson Reuters' 13F filings. Lastly, *Proceeds* is the total value of transaction by the firm during the previous 36 months, defined for each transaction type (equity, debt, or M&A) and across all combined transactions (overall). Across all observations in our sample, mean (median) values are 11 (1) for analyst following, $9.6 ($1.9) billion for market capitalization, and 62% (70%) for institutional holdings. Three-year proceeds average $77 million, $428 million, and $1,055 million for equity, debt, and M&A, respectively. Across quarterly observations with non-zero proceeds, these averages increase to $300 million, $1,145 million, and $2,981 million.

Panel B of Table 1 provides mean values of all variables for the subsamples of observations involving sanctioned and non-sanctioned banks. As expected, sanctioned banks tend to be larger and have higher market shares than non-sanctioned banks. For example, the mean values of *IB_Size* (i.e., number of analysts) and equity market share are 116.2 and 7.2% for sanctioned banks, compared to 52.1 and 1.01% for non-sanctioned banks. Other categories of market share and measures of investment banking relationships provide similar results. Analyst and firm characteristics also differ significantly between the two groups of banks, though the differences are smaller economically than the differences in bank size and market share. Analysts employed by sanctioned banks are more likely to be ranked as All Stars, have higher seniority and seasoning, and follow more stocks than analysts employed by non-sanctioned banks.

In addition, analysts employed by sanctioned banks tend to follow larger stocks, with higher institutional ownership and more equity, debt, and M&A activity. While forecast bias and accuracy are similar across the two groups of analysts, recommendations and relative recommendations tend to be higher for analysts at non-sanctioned banks, on average. As a result, we control for differences between sanctioned and non-sanctioned banks in our analysis to follow. Despite the observed differences described above, non-sanctioned banks and the firms that hire them are involved in a significant fraction of equity, debt, and M&A activity over our sample period and account for a large fraction (43%) of the quarterly analyst observations in our data.

To highlight the relation between investment banking relationships and analyst recommendations, Figure 2 plots the frequency of various recommendations for sanctioned and non-sanctioned banks across the entire sample of quarterly observations. Frequencies are further categorized by whether or not the analyst was affiliated with the covered firm, where affiliation is defined based on the overall investment banking relationship over the previous 36 months. Results for the period prior to the Global Settlement are provided in Panel A and results for the period following Global Settlement are provided in Panel B.

The plots on the left show frequencies based on a 5-tier recommendation scale. From these graphs, it is clear that Sell and Strong Sell recommendations are rare in the period before the Global Settlement. While negative recommendations are more common in the post period, they remain relatively rare. Most importantly, the graph shows that affiliated analysts are more likely to issue Strong Buy recommendations and less likely to issue Hold or Sell recommendations than unaffiliated analysts. Although the bias is reduced in the period after the Global Settlement, it does not appear to be eliminated for either sanctioned or non-sanctioned banks, and remains particularly strong for non-sanctioned banks.

Kadan et al. (2009) note that, following the Global Settlement, many large investment banks shifted from 5-tier to 3-tier recommendation schemes. This shift is also evident in our data. For example, from 1998-2001, Deutsche Bank's investment recommendations included the five categories: Strong Buy, Buy, Hold, Underperform, and Sell. In contrast, from 2004-2009, Deutsche Alex Brown's investment recommendations included the three categories: Buy, Hold, and Sell. To ensure that our results are robust to this shift in recommendation schemes, we reassign all recommendations to a 3-tier scale. Frequencies based on this redefined scale are shown on the right side of Figure 2. The results from this redefined scale are consistent with those from the 5-tier scale, with affiliated analysts being less likely to issue Sell or Hold recommendations and more likely to issue Buy recommendations.

The results in Figure 2 suggest that analyst affiliation bias persists following the Global Settlement. However, these frequencies do not control for other factors that may affect analyst recommendations. In the next section, we therefore analyze analyst recommendations in a multivariate framework.

4. Results

In this section, we describe our main results related to analyst affiliation bias. Using the quarterly data described above, we estimate variations of the following general model specification:

$$\begin{aligned} RelRec_{ijkt} = \alpha + \beta_1 \times IB\_GS + \beta_2 \times IB\_NonGS + \beta_3 \times IBRel_{jkt} \times IB\_GS + \beta_4 \times IBRel_{jkt} \times IB\_NonGS \\ + \sum_{i=1}^{I} \delta_i \times AnalystChar_i + \sum_{j=1}^{J} \gamma_j \times IBChar_j + \sum_{k=1}^{K} \lambda_k \times StockChar_k + \varepsilon_{ijkt}, \end{aligned} \quad (1)$$

where $IBRel_{jkt}$ indicates an investment banking relationship between investment bank j and firm k during the 36 months ending in quarter t, and the remaining variables represent controls for analyst, investment bank, and stock characteristics. Our main tests are based on a comparison of the relationship interaction terms involving *IB_GS* and *IB_NonGS*, which are dummy variables that distinguish between investment banks that were and were not sanctioned in the Global Settlement, respectively. To examine the impact of the Global Settlement on analyst affiliation bias, we provide two sets of analysis. In the full period analysis, we interact the relationship variables with a dummy variable equal to one for all quarters after the Global Settlement and zero otherwise. We also provide separate analyses for the sub-periods 1998-2001 and 2003-2009. Following Kadan et al. (2009), we define the implementation date for the Global Settlement as September 2002, but because the investigations related to investment banking conflicts of interest were ongoing during 2002, we exclude 2002 from the sub-period analysis. Our general specifications also include year and firm fixed effects.

4.1 Relative Recommendations and Investment Banking Relationships

The full period regression results are presented in Table 2. *P*-values based on robust standard errors clustered by firm are reported below the coefficients. Examining the coefficients on the control variables, we see that relative recommendations are lower for large investment banks and for analysts that cover a large number of stocks, and higher for more experienced analysts and for stocks followed by a large number of analysts. Investment bank market share is positively related to relative recommendations for equity, M&A, and overall relationships, but negatively related for debt relationships. The coefficient signs for investment bank market share, for analyst All-Star ranking, seasoning, and number of firms followed, and for the firm's analyst following are generally consistent with results reported in Ljungqvist et al. (2007), but the negative coefficient on investment bank size differs from their results.[14] Consistent with expectations, the coefficient on the post-Global Settlement dummy variable indicates that relative recommendations dropped in the post period. As in Table 1, there is also evidence that non-sanctioned banks tend to have higher recommendations than sanctioned banks, especially in the post-Global Settlement period.

Turning to the results for investment banking relationships, we find strong evidence that both sanctioned and non-sanctioned banks exhibited significant affiliation bias in the pre-Global Settlement period. This result holds for each type-specific relationship (equity, debt, and M&A), as well as for the overall relationship. However, the post-GS interaction terms point to significant differences between sanctioned and non-sanctioned banks in the period following the Global Settlement. For sanctioned banks, the interaction terms suggest that analyst affiliation bias is significantly reduced in the post-Global Settlement period. In particular, the combined post-Global Settlement effects listed at the bottom of the table show that analyst affiliation bias is insignificant in the post period for equity relationships, and marginally significant for debt and M&A relationships. The results for overall relationships point to

[14] In our analysis of the sub-period from 1998-2001 (Table 3 Panel A), we obtain a positive and significant coefficient on investment bank size, consistent with Ljungqvist et al.'s (2007) results for the 1994-2000 sample period.

statistically significant affiliation bias for sanctioned banks in the period after the Global Settlement, but the magnitude of the effect is substantially reduced from the pre period. Based on the coefficients on the overall relationship variable (0.160) and the post-GS interaction term (-0.129), affiliation bias is reduced by approximately 81% in the post Global Settlement period for sanctioned banks.

The results for non-sanctioned banks provide a sharp contrast. For these investment banks, analyst affiliation bias is not reduced significantly in the period following the Global Settlement. The results provide strong evidence of a continued analyst affiliation bias in the period following the Global Settlement for non-sanctioned banks, regardless of whether relationships are measured based on equity, debt, or M&A transactions, or across all combined transactions. Based on the coefficients on the overall relationship variable (0.171) and the post-GS interaction term (-0.010), affiliation bias is reduced by only 5.9% in the post Global Settlement period for non-sanctioned banks and this reduction is statistically insignificant.

To better understand the effects of analyst affiliation bias in the periods before and after the Global Settlement, we estimate models using two sub-periods: 1998-2001 and 2003-2009. The results are presented in Panels A and B of Table 3, respectively. As in Table 2, the results for the first sub-period point to significant analyst affiliation bias for both sanctioned and non-sanctioned banks. For sanctioned banks, the coefficient on *IBRel* is positive and significant for all type-specific and overall relationships. For non-sanctioned banks, the coefficient is positive and insignificant for equity and debt relationships, positive and marginally significant for M&A, and significantly positive for the overall relationship measure. . Equality of coefficients between sanctioned and non-sanctioned banks cannot be rejected for any of the relationships measures in the pre-settlement sub-period.

The results for the second sub-period (Panel B) confirm the findings from Table 2. For sanctioned banks, the coefficient on *IBRel* is positive but insignificant for equity relationships, positive and marginally significant for debt and M&A, and significantly positive for overall relationships. However, as in Table 2, the impact of investment banking relationships on relative recommendations is substantially reduced for sanctioned banks in the post-Global Settlement period. For non-sanctioned banks, significant analyst affiliation bias remains in the post-Global Settlement period, regardless of the relationship measure used. Indeed, the coefficients uniformly increase in the second sub-period for non-sanctioned banks. Equality of coefficients between sanctioned and non-sanctioned banks is rejected in the second sub-period for equity (*p*-value=0.002), M&A (0.014), and overall relationships (0.000), but is not rejected for debt relationships (0.145).

The results from Tables 2 and 3 suggest that overall investment banking relationships better capture analyst affiliation bias than relationship measures based solely on equity, debt, or M&A transactions. As noted earlier, this may reflect that relationships spanning multiple functional areas put more pressure on analysts to produce optimistic recommendations or it may be the result of the overall measure better capturing the continuous nature of the underlying investment banking relationship. In unreported results, we examine whether any of the type-specific relationship measures have incremental explanatory power when included in the regression with the overall measure. In each case, the effects of type-specific relationships are subsumed by the overall relationship measure. Given these results, we focus on overall investment banking relationships throughout the rest of the paper.

The specifications described in Tables 2 and 3 follow prior literature by including firm fixed effects. To examine the robustness of the results to this choice and to the specification of the relationship measure, Table 4 reports results from alternative specifications incorporating analyst and investment bank fixed effects using both the indicator and continuous relationship measures. Results for the sub-periods before and after the Global Settlement are provided in Panels A and B, respectively. The first column in each panel of Table 4 repeats the overall relationship specification from Table 3. Comparing this specification to those based on alternative fixed effects and continuous relationship measures shows that the main results are robust to these alternative specifications. For both continuous and discrete measures of investment banking relationships, the results point to significant analyst affiliation bias in the first sub-period, regardless of specification. In the second sub-period, the results become somewhat weaker after incorporating investment bank fixed effects, but remain significant, especially for non-sanctioned banks. Interestingly, results for sanctioned banks are statistically significant based on relationship dummy variables, but insignificant based on continuous relationship measures.

In unreported results, we estimated two other robustness checks. First, we re-estimated the basic model for the subsets of sanctioned and non-sanctioned banks. Second, we re-estimated the model for the subset of firms covered by at least one affiliated and one non-affiliated analyst. In all cases, the findings are consistent with the overall results reported above.

Taken together, the results in Tables 2 through 4 provide strong evidence of analyst affiliation bias in the period following the Global Settlement for at least some investment banks. While this bias is substantially reduced in the post-Global Settlement period for investment banks named in the settlement, it remains significant when measured based on overall investment banking relationships. The coefficients from Table 2 suggest an 81% reduction in the magnitude of the bias for sanctioned banks when measured with the overall relationship. For the banks not named in the Global Settlement, analyst affiliation bias remains large and significant even after the Global Settlement. These results suggest that the reduction in affiliation bias is driven by the punitive and bank-specific requirements of the Global Settlement, rather than the broader regulatory changes that accompanied the settlement.

4.2. Relative Recommendations based on a 3-Tier System

Kadan et al. (2009) point out that, following the Global Settlement, many brokerages shifted from 5-tier to 3-tier recommendation scales, with all ten of the original Global Settlement banks adopting 3-tier scales in 2002 or soon thereafter. If only sanctioned banks shifted to this new recommendation scale or if the shift differs by bank type, it is possible that our measure of relative recommendations is inflated for non-sanctioned banks relative to sanctioned banks. To ensure that our results are not driven by this shift in recommendation scales, we re-estimate our main regressions after redefining all recommendations based on a 3-tier scale. Specifically, we redefine I/B/E/S recommendations such that a 3 represents a Strong Buy or Buy and a 1 represents a Sell or Strong Sell, and recalculate relative recommendations accordingly.

Table 5 reports regression results based on this redefined relative recommendation variable, with results for the sub-periods before and after the Global Settlement reported in Panels A and B, respectively. For completeness, we provide results based on transaction type relationships (equity, debt, and M&A), as well as overall relationships. For both sub-periods, the results are generally consistent with the main results presented in Tables 2 and 3. In the first sub-period, there is evidence of analyst affiliation bias for sanctioned banks based on all relationship measures. For non-sanctioned banks, there is evidence of analyst affiliation bias based on M&A and overall relationships, but insignificant results based on equity and debt relationships.

In the second sub-period, the impact of analyst affiliation is reduced for sanctioned banks, though it remains statistically significant for all relationship measures. For non-sanctioned banks, we again find strong evidence of analyst affiliation bias in the post-settlement period based on both transaction type and overall relationship measures. Thus, our results are not driven by the shift of some investment banks from a 5-tier to a 3-tier recommendation scale.

4.3. Logit Models for Buy/Sell Recommendations

As an alternative test, we follow Kadan et al. (2009) in estimating logit models for the likelihood of buy/strong buy recommendations and the likelihood of sell/strong sell recommendations, where we focus on affiliation effects and differences between sanctioned and non-sanctioned banks. The models follow the specification described in equation (1). However, we define two alternative dependent variables. The first is an indicator variable equal to one if the analyst issues a buy or strong buy recommendation and zero otherwise. The second is an indicator variable equal to one if the analyst issues a sell or strong sell recommendation and zero otherwise. The logit framework has two advantages over the regression specifications presented earlier. First, like the analysis in Table 5, the dependent variables in the logit models are defined based on a 3-tier recommendation scale and are therefore robust to a shift in recommendation scales by some investment banks. Second, the dependent variables in the logit model are defined directly from I/B/E/S recommendations and are therefore unaffected by the definition of "consensus" ranking used in the construction of *RelRec*.

Table 6 presents the results from the logit models for both the full period and the pre/post Global Settlement sub-periods. Again, the findings point to significant analyst affiliation bias. In the models for buy/strong buy recommendations, the results suggest that both sanctioned and non-sanctioned banks are significantly more likely to issue buy or strong buy recommendations when affiliated with the covered firm through an investment banking relationship. For sanctioned banks, this effect is strongest during the first sub-period, but remains statistically significant even after the Global Settlement. For non-sanctioned banks, affiliation bias is statistically significant and similar in magnitude both before and after the Global Settlement.

The logit results for sell/strong sell recommendations point to symmetric effects in terms of pessimistic recommendations, although the results appear to be driven primarily by the period after the Global Settlement. Specifically, during the post-Global Settlement period, both sanctioned and non-sanctioned banks are less likely to issue sell or strong sell recommendations when affiliated with the firm through an investment banking relationship.

The results from the logit models are largely consistent with those based on relative recommendations and suggest that analysts tend to issue more optimistic (or less pessimistic) recommendations on firms with which their employer has an investment banking relationship.

5. The Impact of Lending Activity on Analyst Affiliation Bias

The passage of the Gramm-Leach-Bliley Act in 1999 led to a substantial increase in the role of commercial banks in investment banking and more direct ties between lending and underwriting relationships. For example, Ljungqvist et al. (2006), Drucker and Puri (2005), Yasuda (2005), and Bharath, Dahiya, Saunders, and Srinivasan (2007) find that lending relationships increase the likelihood of a bank being awarded future debt and equity underwriting business, and Corwin and Stegemoller (2014) identify important links between lending and the cross-functional nature of investment banking relationships. In this section, we examine whether lending relationships have any incremental impact on analyst affiliation bias, after controlling for investment banking relationships based on equity, debt, and M&A transactions.[15]

To examine lending relationships, we use Dealscan data to collect the sample of syndicated loans involving our sample firms. We match CRSP firms to Dealscan data using the link table provided by Michael Roberts and Wharton Research Data Services (see Chava and Roberts (2008)). For each loan, we identify the loan amount and all lenders identified as having lead arranger credit. Notably, the Dealscan data include both loans and revolving credit line agreements. We believe credit lines are an important part of a lending relationship, regardless of whether or not the loan is drawn down. However, the fact that these loans may not be drawn down suggests that the total loan values in Dealscan will not be comparable to the transaction values in the equity, debt, and M&A datasets.

To integrate the lending and investment banking datasets, we hand match lender names to our sample of large investment banks. Following the construction of the investment banking variables, we calculate investment bank market share, firm loan proceeds, and firm-lender relationships at the end of each quarter. For each investment bank in our sample, we calculate lending market share based on all loans over the prior twelve months. For each firm in our sample, we calculate lending proceeds as the sum of all loans received over the preceding 36 months. Finally, for each firm-investment bank pair, we calculate the lending relationship as the proportion of the firm's total loan value over the preceding 36 months for which the investment bank was assigned lead arranger credit and we calculate a revised "overall" relationship measure combining lending with equity, debt, and M&A transaction values.

Summary statistics for the lending variables are provided in Panel A of Table 7. Across all quarterly observations in our sample, the lending relationship has a mean value of 2.82% and the overall relationship incorporating lending has a mean value of 5.84%. Investment bank market share has a mean (median) value of 4.56% (0.74%) based on lending alone and 4.58% (2.05%) based on the combined values of lending, equity, debt, and M&A transactions. The average value of three-year lending proceeds

[15] Although they do not analyze recommendations, Chen and Martin (2011) examine the relation between earnings forecast accuracy and lending relationships. They find that forecast accuracy improves after a firm borrows from an affiliated bank, suggesting that lending provides affiliated analysts with an informational advantage over other analysts.

for the firms in our sample is \$964.1 million across all observations and \$1,818.3 million across observations with positive lending proceeds.

Table 7 describes coefficients from regressions of relative recommendations on the set of control variables and investment banking relationship variables, after incorporating lending, with results for the pre and post-Global Settlement sub-periods in Panels B and C, respectively. To conserve space, coefficients on control variables are not included. The table provides results from four different specifications. The first specification includes only lending relationship indicators. This specification suggests that lending relationships have a positive impact on analyst affiliation bias in the 1998-2001 sub-period, but an insignificant effect after 2002. In the second specification, we include the lending relationship indicator in addition to the overall relationship indicator based on equity, debt, and M&A transactions. This regression suggests that lending may have some incremental impact on affiliation bias beyond that captured by the investment banking relationship, but the impact is again strongest during the first sub-period.

In the third specification, we again include the overall relationship indicator based on combined equity, debt, and M&A transactions, but we add an interaction with the lending relationship indicator. The results from this specification suggest that the affiliation bias associated with investment banking relationships is magnified in cases where there is also a lending relationship, especially during the first sub-period. Finally, in the fourth specification, we provide results based on the redefined overall relationship indicator that incorporates equity, debt, M&A, and lending transactions. This combined measure produces results that are similar to those from the overall relationship measure without lending, with affiliation bias being significant for non-sanctioned banks in both sub-periods and strongest for sanctioned banks in the first sub-period.

The results in Table 7 provide weak evidence that lending leads to incremental affiliation bias effects beyond those captured by investment banking relationships, at least during the first sub-period. However, unlike the main results based on equity, debt, and M&A relationships, the findings in Table 7 are sensitive to the inclusion of alternative fixed effects. In untabulated results, we find that when either analyst or investment bank fixed effects are included in these models, the incremental effects of lending become insignificant. Thus, there is limited evidence of any incremental impact of lending relationships on analyst affiliation bias in the period after the Global Settlement.

6. Conclusion

Previous research provides strong evidence of conflicts of interest between investment banking and research departments within large investment banks. In particular, research shows that analysts tend to issue optimistic recommendations on firms with which their employer has an equity underwriting relationship. One of the major purposes of the 2003 Global Analyst Research Settlement reached between the SEC, NYSE, NASD, New York Attorney General, and North American Securities Administrators Association and 12 of the largest investment banks was to reduce these conflicts of interest. In this study, we use a comprehensive measure of relationships between investment banks and firms to examine the impact of the Global Settlement on analyst affiliation bias.

Our data include all equity, debt, and M&A transactions by U.S. firms, allowing us to analyze a more comprehensive measure of investment banking relationships than has been studied in prior literature. In general, we find evidence of analyst affiliation bias for each individual type of investment banking relationship. However, our results suggest that an overall measure spanning all functional areas does a better job of capturing investment banking relationships and the related affiliation bias.

To better understand the impact of the Global Settlement and contemporaneous regulatory changes on analyst behavior, we separate analysts employed by investment banks named in the Global Settlement (sanctioned banks) and other top investment banks (non-sanctioned banks). Consistent with prior research, our results provide strong evidence of analyst affiliation bias for both groups of banks in the period prior to the Global Settlement. Following the Global Settlements, affiliation bias is substantially reduced, but not eliminated, for those banks named in the Global Settlement. In contrast, we find strong evidence of analyst affiliation bias for non-sanctioned banks even after the Global Settlement.

These findings suggest that the Global Settlement and related regulatory changes were only partially successful in mitigating conflicts of interest between investment banking and analyst research. In particular, the impact appears limited to the subset of sanctioned banks, suggesting that the decline in analyst affiliation bias is driven by the punitive aspects or bank-specific requirements of the Global Settlement more than the broader regulatory changes imposed on the industry.

References

Barber, B., R. Lehavy, and B. Trueman. 2007. Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics* 85: 490–517.

Barniv, R., O.-K. Hope, M. Myring, and W. Thomas. 2009. Do analysts practice what they preach and should investors listen? Effects of recent regulations. *The Accounting Review* 84 (4): 1015-1039.

Beneish, M. 1991. Stock prices and the dissemination of analysts' recommendations. *Journal of Business* 64 (3): 393-416.

Bharath, S., S. Dahiya, A. Saunders, and A. Srinivasan, 2007, So what do I get? The bank's view of lending relationships, *Journal of Financial Economics* 85, 368-419.

Bradshaw, M. 2004. How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review* 79 (1): 25-50.

Bradshaw, M., S. Richardson, and R. Sloan. 2006. The relation between corporate financing activities, analysts' forecasts and stock returns. *Journal of Accounting and Economics* 42: 53-85.

Brown, L., A. Call, M. Clement, and N. Sharp. 2014. Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research*, forthcoming.

Chava, S., and M. Roberts, 2008, How does financing impact investment? The role of debt covenants, *Journal of Finance* 63, 2085-2121.

Chen, T., and X. Martin. 2011. Do bank-affiliated analysts benefit from lending relationship? *Journal of Accounting Research* 49 (3): 633-675.

Chen, C.-Y., and P. Chen. 2009. NASD Rule 2711 and changes in analysts' independence in making stock recommendations. *The Accounting Review* 84 (4): 1041-1071.

Clarke, J., A. Khorana, A. Patel, and P. Rau. 2011. Independents' day? Analyst behavior surrounding the Global Settlement. *Annals of Finance* 7: 529-547.

Corwin, S. A., and M. Stegemoller. 2014, The changing nature of investment banking relationships. Working paper, University of Notre Dame and Baylor University.

Cowen, A., B. Groysberg, and P. Healy. 2006. Which types of analyst firms are more optimistic? *Journal of Accounting and Economics* 41: 119-146.

Dechow, P., A. Hutton, and R. Sloan. 2000. The relation between analysts' forecast of long-term earnings growth and stock price performance following equity offerings. *Contemporary Accounting Research* 17 (1): 1-32.

De Franco, G., H. Lu, and F. Vasvari. 2007. Wealth transfer effects of analysts' misleading behavior. *Journal of Accounting Research* 45 (1): 71-110.

Drucker, S., and M. Puri, 2005, On the benefits of concurrent lending and underwriting, *Journal of Finance* 60, 2763-2799.

Dugar, A., and S. Nathan. 1995. The effect of investment banking relationships on financial analysts' earnings forecasts and investment recommendations. *Contemporary Accounting Research* 12 (1): 131-160.

Francis, J. and D. Philbrick. 1993. Analysts' decisions as products of a multi-task environment. *Journal of Accounting Research* 31: 216-230.

Guan, Y., H. Lu, and F. Wong. 2012. Conflict-of-interest reforms and investment bank analysts' research biases. *Journal of Accounting, Auditing, and Finance* 27 (4): 443-470.

Hong, H., and J. Kubik. 2003. Analyzing the analysts: career concerns and biased earnings forecasts. *Journal of Finance* 58: 313–351.

Irvine, P. 1996. Do analysts generate trade for their firms? Evidence from the Toronto stock exchange. *Journal of Accounting and Economics* 30: 209-226.

Kadan, O., L. Madureira, R. Wang, and T. Zach. 2009. Conflicts of interest and stock recommendations – the effect of the Global Settlement and related regulations. *Review of Financial Studies* 22 (10): 4189-4217.

Kadan, O., L. Madureira, R. Wang, and T. Zach. 2012. Analysts' industry expertise. *Journal of Accounting and Economics* 54: 95-120.

Kolasinski, A., and S.P. Kothari. 2008. Investment banking and analyst objectivity: Evidence from analysts affiliated with mergers and acquisitions advisors. *Journal of Financial and Quantitative Analysis* 43 (4): 817-842.

La Porta, R. 1996. Expectations and the cross-sections of returns. *The Journal of Finance* 51 (5): 1715-1742.

Lin, H., and M. McNichols. 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. Journal of Accounting and Economics 25 (1): 101-127.

Ljungqvist, A., F. Marston, and W. J. Wilhelm, Jr., 2006, Competing for securities underwriting mandates: Banking relationships and analyst recommendations, *Journal of Finance* 61, 301-340.

Ljungqvist, A., F. Marston, L. Starks, K. Wei, and H. Yan. 2007. Conflicts of interest in sell-side research and the moderating role of institutional investors. *Journal of Financial Economics* 85: 420-456.

Ljungqvist, A., F. Marston, and W. Wilhelm. 2009. Scaling the hierarchy: How and why investment banks compete for syndicate co-management appointments. *The Review of Financial Studies* 22 (10): 3977-4007.

McNichols, M., and P. O'Brien. 1997. Self-selection and analyst coverage. *Journal of Accounting Research* 35 (Supplement): 167-199.

Mehran, H., and R. Stulz. 2007. The economics of conflicts of interest in financial institutions. *Journal of Financial Economics* 85: 267-296.

Michaely, R. and K. Womack. 1999. Conflict of interest and the credibility of underwriter analyst recommendations. *The Review of Financial Studies* 12 (4): 653-686.

O'Brien, P., McNichols, M., Lin, H., 2005. Analyst impartiality and investment banking relationships. *Journal of Accounting Research* 43, 623–650.

Yasuda, A., 2005, Do bank relationships affect the firm's underwriter choice in the corporate bond underwriting market? *Journal of Finance* 60, 1259-1292.

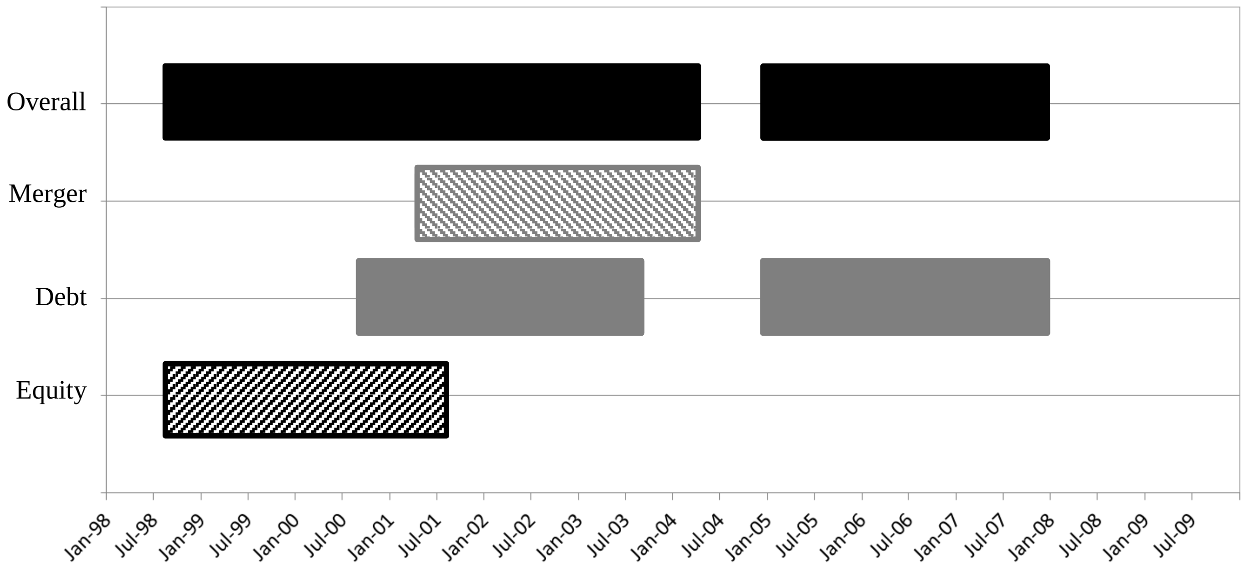
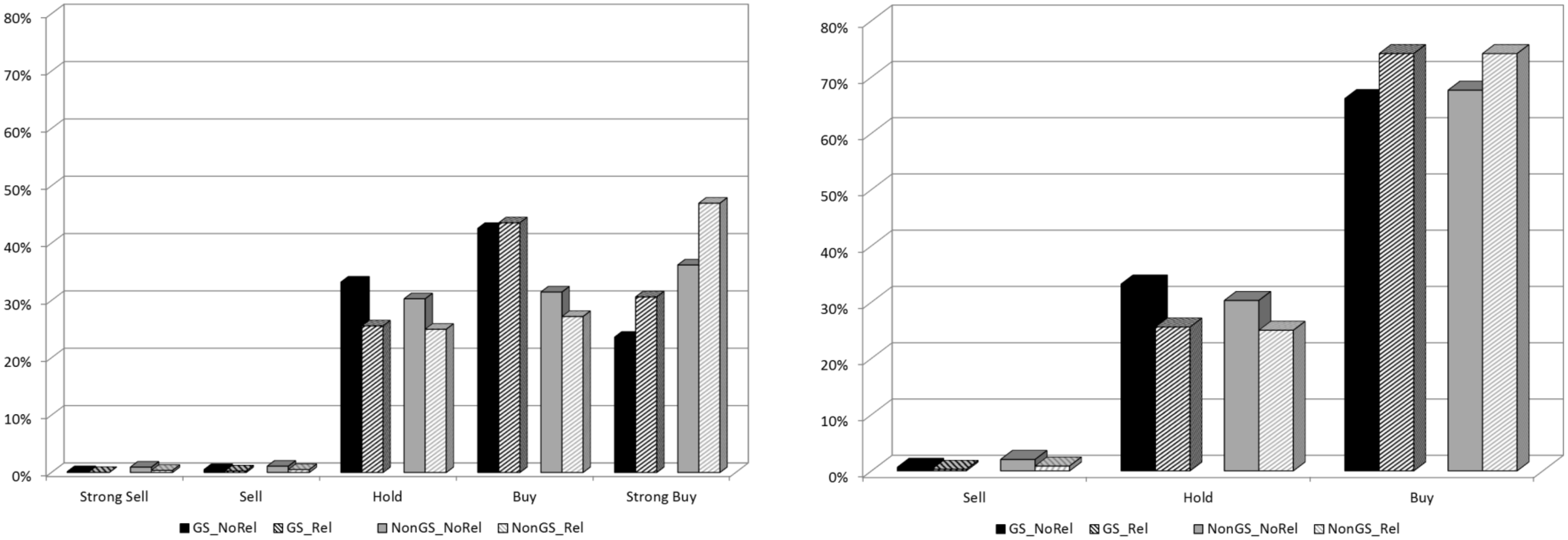


Figure 1 – Relationship Illustration for Convergys Corp and Citi Salomon Smith

This figure provides an illustration of our measures of investment banking relationships. We define a firm-bank pair as having a relationship if at any point during the preceding 36 months, the firm had an equity, debt, or M&A transaction for which the investment bank served as a lead or co-managing underwriter or M&A advisor. Equity, debt, and M&A relationships are defined based only on transactions within each category. The overall relationship is defined based on transactions across all three categories.

Panel A: Pre-Global Settlement



Panel B: Post-Global Settlement

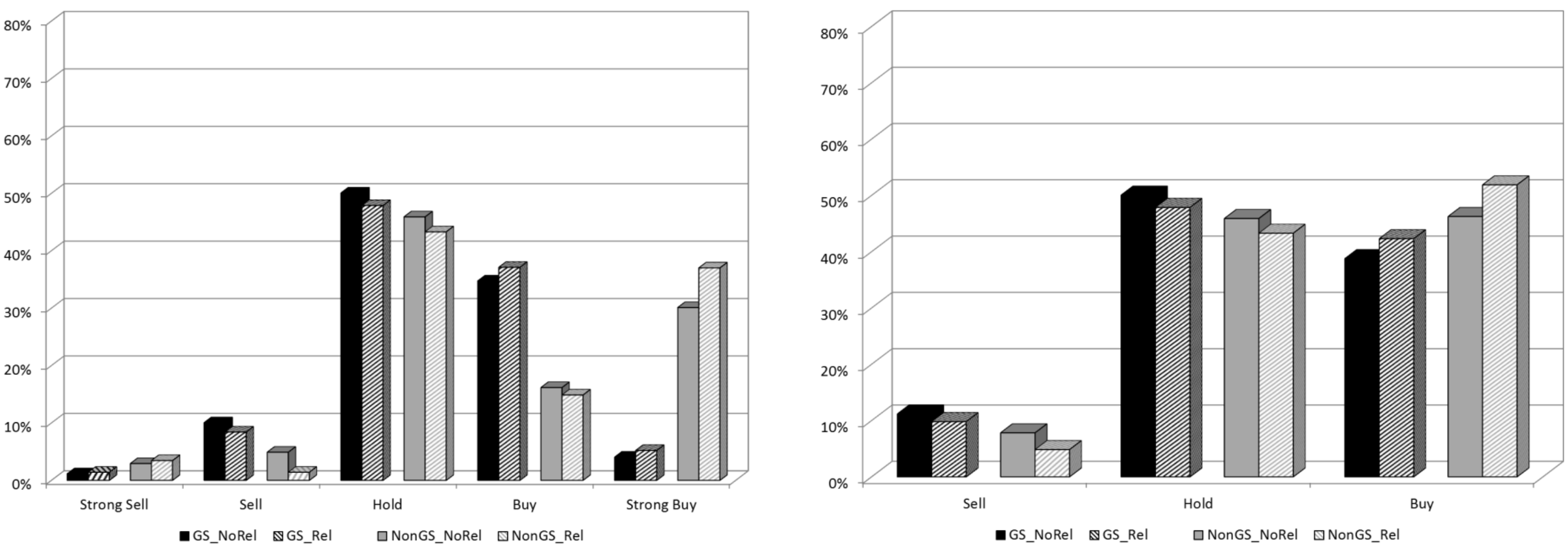


Figure 2 – Recommendation Frequency Before and After Global Settlement

The figure plots recommendation frequencies for our sample of quarterly data, where frequencies are classified on both a five-tier and a three-tier scale. Analysts are classified as being affiliated with either a Global Settlement bank or a non-Global Settlement bank and firm-analyst observations are separated into those that are associated with an investment bank relationship and those that are not, based on the overall investment banking relationship.

Table 1 – Summary Statistics

This table provides descriptive statistics for the variables used in this study. Variable definitions are contained in Appendix Table A1. Panel A provides summary statistics for the full sample, including 216,242 quarterly observations. The non-zero proceeds variables are based on 55,221 observations for equity, 80,823 observations for debt, 76,491 observations for M&A, and 140,997 observations for all combined transactions (overall). Panel B provides mean values for the subsamples of observations related to sanctioned and Non-sanctioned bank analysts. The *p*-value in the last column of Panel B is from a test of difference in means across sanctioned and Non-sanctioned banks based on analysis of variance.

Panel A: Full Sample Summary Statistics

	Mean	Median	Min	Max	Std. Dev.
Recommendation and Forecast Measures:					
Analyst Recommendation	3.61	4.00	1.00	5.00	0.91
Relative Recommendation	0.0025	0.00	-4.00	3.00	0.80
Adjusted Forecast Bias	-0.0351	0.00	-9.24	5.57	0.96
Adjusted Forecast Accuracy	0.0437	0.00	-9.11	5.34	0.87
IB Relationship Measures:					
IBRel_Equity (%)	3.24	0.00	0.00	1.00	15.51
IBRel_Debt (%)	2.72	0.00	0.00	1.00	13.03
IBRel_Merger (%)	2.43	0.00	0.00	1.00	14.14
IBRel_Overall (%)	5.90	0.00	0.00	1.00	19.49
IB Characteristics:					
IB_Size	88.74	85.00	1.00	250.00	49.65
IB_MktShare_Equity (%)	4.55	2.81	0.00	22.11	4.84
IB_MktShare_Debt (%)	4.77	2.13	0.00	21.64	5.63
IB_MktShare_Merger (%)	4.38	1.70	0.00	34.13	5.67
IB_MktShare_Overall (%)	4.47	2.18	0.00	23.06	5.17
Analyst Characteristics:					
RelAccuracy (%)	41.23	40.96	0.00	100.00	10.33
AllStar	0.19	0.00	0.00	1.00	0.39
Seniority	5.43	4.92	0.00	16.18	3.47
Seasoning	2.33	1.39	0.00	16.18	2.46
NFollow	10.96	10.00	1.00	103.00	7.22
JobMove	0.03	0.00	0.00	1.00	0.18
Firm/Stock Characteristics:					
ANF	10.02	9.00	2.00	51.00	6.18
InstHoldings (%)	62.10	69.81	0.00	100.00	29.44
MV	9,592.51	1,886.44	0.76	602,432.92	28,686.62
Proceeds_Equity	76.61	0.00	0.00	12,189.10	312.10
Proceeds_Debt	427.87	0.00	0.00	34,879.74	1,335.85
Proceeds_Merger	1,054.52	0.00	0.00	153,653.35	5,672.22
Proceeds_Overall	1,575.53	152.30	0.00	178,009.68	6,477.18
Proceeds_Equity^{+}	300.01	139.20	0.70	12,189.10	560.73
Proceeds_Debt^{+}	1,144.78	491.25	3.00	34,879.74	1,988.39
Proceeds_Merger^{+}	2,981.15	591.59	0.95	153,653.35	9,231.15
Proceeds_Overall^{+}	2,416.34	498.18	0.70	178,009.68	7,893.76

Table 1 – continued

Panel B: Sanctioned vs. Non-Sanctioned Banks			
	Sanctioned Banks	Non-Sanctioned Banks	*p*-value for difference
N	123,708	92,534	-
Recommendation and Forecast Measures:			
Analyst Recommendation	3.48	3.78	0.000
Relative Recommendation	-0.0777	0.1098	0.000
Adjusted Forecast Bias	-0.0395	-0.0293	0.013
Adjusted Forecast Accuracy	0.0442	0.0430	0.739
IB Relationship Measures:			
IBRel_Equity (%)	4.42	1.67	0.000
IBRel_Debt (%)	4.46	0.81	0.000
IBRel_Merger (%)	3.45	1.07	0.000
IBRel_Overall (%)	8.32	2.67	0.000
IB Characteristics:			
IB_Size	116.15	52.09	0.000
IB_MktShare_Equity (%)	7.20	1.01	0.000
IB_MktShare_Debt (%)	7.35	1.31	0.000
IB_MktShare_Merger (%)	7.20	0.60	0.000
IB_MktShare_Overall (%)	7.24	0.78	0.000
Analyst Characteristics:			
RelAccuracy (%)	41.05	41.47	0.000
AllStar	0.28	0.06	0.000
Seniority	5.48	5.37	0.000
Seasoning	2.46	2.16	0.000
NFollow	11.49	10.25	0.000
JobMove	0.03	0.04	0.000
Firm/Stock Characteristics:			
ANF	10.12	9.88	0.000
InstHoldings (%)	63.18	60.66	0.000
MV	10,253.75	8,708.50	0.000
Proceeds_Equity	81.28	70.37	0.000
Proceeds_Debt	479.30	359.12	0.000
Proceeds_Merger	1,131.00	952.27	0.000
Proceeds_Overall	1,708.67	1,397.54	0.000
Proceeds_Equity^{+}	343.35	251.06	0.000
Proceeds_Debt^{+}	1,195.89	1,063.66	0.000
Proceeds_Merger^{+}	3,102.64	2,806.65	0.000
Proceeds_Overall^{+}	2,593.51	2,173.63	0.000

Table 2 – Full Period Regressions for Relative Recommendations

This table provides the results from estimating regressions of relative recommendations on investment bank relationship measures, investment bank characteristics, analyst characteristics, and stock characteristics for the full sample period 1998 to 2009. Columns 1 through 3 respectively use equity, debt, and M&A investment banking relationship measures while column 4 uses an overall relationship measure. *p*-values based on robust standard errors are presented in parentheses below the coefficients, where standard errors are clustered by firm. Each model contains year and firm fixed effects. *GS* and *NonGS* refer to sanctioned and non-sanctioned banks, respectively. Variable definitions are contained in Appendix Table A1.

	Equity Relationship	Debt Relationship	M&A Relationship	Overall Relationship
Intercept	0.168	0.263	0.162	0.169
	(.001)	(.000)	(.002)	(.001)
Post	-0.134	-0.139	-0.143	-0.122
	(.000)	(.000)	(.000)	(.000)
IB Relationship Measures:				
IBRel_GS	0.122	0.129	0.108	0.160
	(.000)	(.000)	(.000)	(.000)
IBRel_GS*Post	-0.121	-0.102	-0.068	-0.129
	(.000)	(.000)	(.024)	(.000)
IBRel_NonGS	0.171	0.162	0.172	0.171
	(.000)	(.004)	(.001)	(.000)
IBRel_NonGS*Post	-0.030	-0.055	-0.023	-0.010
	(.590)	(.390)	(.748)	(.789)
IB Characteristics:				
Ln(IB_Size)	-0.044	-0.084	-0.042	-0.048
	(.000)	(.000)	(.000)	(.000)
IB_MktShare	-0.573	0.735	-0.650	-0.548
	(.000)	(.000)	(.000)	(.000)
IB_NonGS	0.019	0.064	0.011	0.028
	(.071)	(.000)	(.296)	(.009)
IB_NonGS*Post	0.200	0.198	0.205	0.187
	(.000)	(.000)	(.000)	(.000)
Analyst Characteristics:				
RelAccuracy	-0.010	-0.004	-0.008	-0.008
	(.707)	(.878)	(.760)	(.778)
AllStar	-0.013	-0.034	-0.013	-0.018
	(.153)	(.000)	(.156)	(.038)
Ln(Seniority)	0.023	0.023	0.023	0.023
	(.000)	(.000)	(.000)	(.000)
Ln(Seasoning)	0.010	0.013	0.010	0.010
	(.084)	(.033)	(.101)	(.088)
Ln(NFollow)	-0.045	-0.037	-0.043	-0.043
	(.000)	(.000)	(.000)	(.000)
JobMove	-0.006	-0.004	-0.007	-0.004
	(.565)	(.698)	(.499)	(.717)
Stock Characteristics:				
Ln(ANF)	0.048	0.046	0.047	0.048
	(.000)	(.000)	(.000)	(.000)
Ln(MV)	0.005	0.005	0.006	0.005
	(.325)	(.297)	(.267)	(.329)
Ln(Proceeds)	-0.001	0.000	-0.001	0.000
	(.670)	(.905)	(.505)	(.783)
InstHoldings	-0.165	-0.201	-0.196	-0.157
	(.467)	(.375)	(.386)	(.489)

Table 2 - continued

Combined Post Effects:				
GS Banks	0.001	0.028	0.041	0.031
	(.951)	(.087)	(.038)	(.009)
Non-GS Banks	0.142	0.107	0.150	0.161
	(.000)	(.019)	(.001)	(.000)
Adjusted R^2	.051	.052	.051	.052
N	216,242	216,242	216,242	216,242

Table 3 – Sub-period Regressions for Relative Recommendations

This table provides the results from estimating regressions of relative recommendations on investment bank relationship measures, investment bank characteristics, analyst characteristics, and stock characteristics. Results for the sub-periods before (1998-2001) and after (2003-2009) Global Settlement period are provided in Panels A and B, respectively. Columns 1 through 3 respectively use equity, debt, and M&A investment banking relationship measures while column 4 uses an overall relationship measure. *p*-values based on robust standard errors are presented in parentheses below the coefficients, where standard errors are clustered by firm. Each model contains year and firm fixed effects. *GS* and *NonGS* refer to sanctioned and non-sanctioned banks, respectively. Variable definitions are contained in Appendix Table A1.

	Equity Relationship	Debt Relationship	M&A Relationship	Overall Relationship
		Panel A: 1998 – 2001		
Intercept	-0.272	-0.214	-0.265	-0.237
	(.003)	(.022)	(.004)	(.011)
IB Relationship Measures:				
IBRel_GS	0.072	0.121	0.063	0.119
	(.005)	(.000)	(.022)	(.000)
IBRel_NonGS	0.050	0.097	0.136	0.106
	(.294)	(.122)	(.029)	(.003)
IB Characteristics:				
Ln(IB_Size)	0.065	0.031	0.058	0.052
	(.000)	(.002)	(.000)	(.000)
IB_MktShare	-0.223	1.126	0.236	0.259
	(.043)	(.000)	(.032)	(.027)
IB_NonGS	0.104	0.156	0.120	0.129
	(.000)	(.000)	(.000)	(.000)
Analyst Characteristics:				
RelAccuracy	0.049	0.062	0.052	0.053
	(.284)	(.178)	(.260)	(.253)
AllStar	-0.013	-0.053	-0.027	-0.036
	(.363)	(.000)	(.054)	(.011)
Ln(Seniority)	-0.007	-0.006	-0.008	-0.008
	(.554)	(.607)	(.539)	(.501)
Ln(Seasoning)	0.054	0.051	0.053	0.052
	(.000)	(.000)	(.000)	(.000)
Ln(NFollow)	-0.049	-0.037	-0.045	-0.043
	(.000)	(.000)	(.000)	(.000)
JobMove	-0.039	-0.040	-0.037	-0.033
	(.008)	(.007)	(.012)	(.023)
Stock Characteristics:				
Ln(ANF)	0.036	0.035	0.036	0.038
	(.008)	(.010)	(.009)	(.006)
Ln(MV)	-0.004	-0.004	-0.004	-0.005
	(.664)	(.648)	(.670)	(.631)
Ln(Proceeds)	0.000	-0.003	-0.001	-0.005
	(.989)	(.405)	(.593)	(.171)
InstHoldings	-0.845	-0.855	-0.852	-0.838
	(.024)	(.022)	(.022)	(.025)
Adjusted R^2	.047	.052	.047	.049
N	59,703	59,703	59,703	59,703
PERMCO clusters	3,367	3,367	3,367	3,367
GS – NonGS = 0	.694	.709	.275	.743

Table 3 – continued

	Equity Relationship	Debt Relationship	M&A Relationship	Overall Relationship
	Panel B: 2003 – 2009			
Intercept	0.307	0.408	0.302	0.298
	(.000)	(.000)	(.000)	(.000)
IB Relationship Measures:				
IBRel_GS	0.010	0.037	0.045	0.042
	(.612)	(.025)	(.032)	(.001)
IBRel_NonGS	0.161	0.107	0.176	0.179
	(.000)	(.020)	(.000)	(.000)
IB Characteristics:				
Ln(IB_Size)	-0.076	-0.131	-0.080	-0.080
	(.000)	(.000)	(.000)	(.000)
IB_MktShare	-1.124	0.648	-1.023	-1.000
	(.000)	(.000)	(.000)	(.000)
IB_NonGS	0.170	0.230	0.171	0.173
	(.000)	(.000)	(.000)	(.000)
Analyst Characteristics:				
RelAccuracy	-0.044	-0.042	-0.037	-0.037
	(.233)	(.249)	(.312)	(.308)
AllStar	-0.007	-0.024	-0.009	-0.012
	(.583)	(.039)	(.444)	(.331)
Ln(Seniority)	0.028	0.027	0.026	0.027
	(.000)	(.001)	(.001)	(.001)
Ln(Seasoning)	-0.005	-0.001	-0.006	-0.006
	(.480)	(.940)	(.449)	(.456)
Ln(NFollow)	-0.036	-0.032	-0.031	-0.033
	(.000)	(.000)	(.000)	(.000)
JobMove	0.022	0.022	0.018	0.020
	(.124)	(.127)	(.208)	(.165)
Stock Characteristics:				
Ln(ANF)	0.033	0.031	0.031	0.033
	(.001)	(.002)	(.002)	(.001)
Ln(MV)	-0.004	-0.003	-0.002	-0.003
	(.639)	(.678)	(.769)	(.720)
Ln(Proceeds)	-0.001	-0.001	-0.001	-0.001
	(.793)	(.726)	(.513)	(.598)
InstHoldings	-0.003	-0.011	-0.014	0.009
	(.992)	(.975)	(.967)	(.980)
Adjusted R^2	.068	.067	.069	.068
N	136,193	136,193	136,193	136,193
PERMCO clusters	3,473	3,473	3,473	3,473
GS – NonGS = 0	.002	.145	.014	.000

Table 4 – Alternative Models for Relative Recommendations

This table provides results from regressions of relative recommendations on overall investment bank relationship measures, investment bank characteristics, analyst characteristics, and stock characteristics. Results for the sub-periods before (1998-2001) and after (2003-2009) Global Settlement period are provided in Panels A and B, respectively. Columns 1 through 3 use an indicator variable for the overall investment banking relationship while columns 4 through 6 use a continuous variable for the overall relationship measure. Columns 1 and 4 include firm fixed effects, columns 2 and 5 use analyst fixed effects, and columns 3 and 6 use investment bank fixed effects. All models contain year fixed effects. *p*-values based on robust standard errors are presented in parentheses below the coefficients, where standard errors are clustered by firm. Variable definitions are contained in Appendix Table A1.

	Overall Relationship Dummy			Overall Relationship Continuous		
			Panel A: 1998 – 2001			
Intercept	-0.237	-0.098	-0.684	-0.245	-0.099	-0.691
	(.011)	(.355)	(.000)	(.008)	(.347)	(.000)
IB Relationship Measures:						
IBRel_GS	0.119	0.098	0.104	-	-	-
	(.000)	(.000)	(.000)			
IBRel_NonGS	0.106	0.072	0.070	-	-	-
	(.003)	(.009)	(.011)			
IBRelC_GS				0.098	0.098	0.102
	-	-	-	(.000)	(.000)	(.000)
IBRelC_NonGS				0.118	0.085	0.090
	-	-	-	(.014)	(.019)	(.011)
IB Characteristics:						
Ln(IB_Size)	0.052	0.002	0.135	0.052	0.002	0.135
	(.000)	(.922)	(.000)	(.000)	(.938)	(.000)
IB_MktShare	0.259	0.517	0.281	0.356	0.562	0.341
	(.027)	(.003)	(.141)	(.002)	(.001)	(.073)
IB_NonGS	0.129	0.028		0.127	0.027	
	(.000)	(.249)	-	(.000)	(.270)	-
Analyst Characteristics:						
RelAccuracy	0.053	0.121	0.123	0.054	0.120	0.123
	(.253)	(.066)	(.001)	(.246)	(.068)	(.001)
AllStar	-0.036	0.003	-0.013	-0.034	0.003	-0.012
	(.011)	(.887)	(.272)	(.016)	(.900)	(.334)
Ln(Seniority)	-0.008	-0.031	-0.006	-0.008	-0.030	-0.006
	(.501)	(.317)	(.524)	(.524)	(.328)	(.546)
Ln(Seasoning)	0.052	0.030	0.042	0.052	0.030	0.042
	(.000)	(.001)	(.000)	(.000)	(.001)	(.000)
Ln(NFollow)	-0.043	-0.041	-0.018	-0.043	-0.041	-0.018
	(.000)	(.000)	(.014)	(.000)	(.000)	(.012)
JobMove	-0.033	-0.029	-0.032	-0.035	-0.030	-0.033
	(.023)	(.038)	(.020)	(.017)	(.035)	(.016)
Stock Characteristics:						
Ln(ANF)	0.038	0.048	0.044	0.037	0.047	0.043
	(.006)	(.000)	(.000)	(.007)	(.000)	(.000)
Ln(MV)	-0.005	0.011	0.004	-0.004	0.011	0.005
	(.631)	(.001)	(.125)	(.654)	(.001)	(.101)
Ln(Proceeds)	-0.005	-0.001	-0.002	-0.003	0.000	-0.001
	(.171)	(.625)	(.190)	(.305)	(.870)	(.563)
InstHoldings	-0.838	-0.711	-0.738	-0.846	-0.715	-0.746
	(.025)	(.003)	(.001)	(.023)	(.003)	(.001)
Fixed Effects	Firm	Analyst	IB	Firm	Analyst	IB
Adjusted R^2	.049	.122	.052	.047	.122	.051
N	59,703	59,703	59,703	59,703	59,703	59,703

Table 4 – continued

	Overall Relationship Dummy			Overall Relationship Continuous		
	Panel B: 2003 – 2009					
Intercept	0.298	-0.278	0.157	0.284	-0.280	0.155
	(.000)	(.008)	(.002)	(.000)	(.008)	(.002)
IB Relationship Measures:						
IBRel_GS	0.042	0.039	0.020	-	-	-
	(.001)	(.001)	(.090)			
IBRel_NonGS	0.179	0.097	0.066	-	-	-
	(.000)	(.000)	(.014)			
IBRelC_GS				-0.003	0.029	-0.003
	-	-	-	(.884)	(.143)	(.895)
IBRelC_NonGS				0.260	0.117	0.084
	-	-	-	(.000)	(.005)	(.042)
IB Characteristics:						
Ln(IB_Size)	-0.080	-0.078	-0.103	-0.078	-0.077	-0.102
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
IB_MktShare	-1.000	-0.427	-0.745	-0.939	-0.387	-0.728
	(.000)	(.021)	(.000)	(.000)	(.038)	(.000)
IB_NonGS	0.173	0.162		0.175	0.165	
	(.000)	(.000)	-	(.000)	(.000)	-
Analyst Characteristics:						
RelAccuracy	-0.037	0.046	0.007	-0.040	0.046	0.006
	(.308)	(.385)	(.837)	(.274)	(.386)	(.856)
AllStar	-0.012	-0.012	-0.004	-0.009	-0.011	-0.003
	(.331)	(.452)	(.723)	(.447)	(.479)	(.779)
Ln(Seniority)	0.027	0.060	0.009	0.027	0.061	0.009
	(.001)	(.006)	(.198)	(.001)	(.006)	(.183)
Ln(Seasoning)	-0.006	0.002	0.006	-0.006	0.001	0.005
	(.456)	(.794)	(.404)	(.448)	(.836)	(.431)
Ln(NFollow)	-0.033	-0.012	-0.018	-0.034	-0.012	-0.018
	(.000)	(.113)	(.001)	(.000)	(.108)	(.001)
JobMove	0.020	0.013	0.028	0.020	0.013	0.028
	(.165)	(.356)	(.041)	(.166)	(.366)	(.042)
Stock Characteristics:						
Ln(ANF)	0.033	0.053	0.035	0.033	0.053	0.035
	(.001)	(.000)	(.000)	(.001)	(.000)	(.000)
Ln(MV)	-0.003	0.035	0.031	-0.003	0.035	0.030
	(.720)	(.000)	(.000)	(.742)	(.000)	(.000)
Ln(Proceeds)	-0.001	0.000	-0.001	0.000	0.001	0.000
	(.598)	(.947)	(.545)	(.913)	(.384)	(.969)
InstHoldings	0.009	0.189	-0.188	0.014	0.170	-0.196
	(.980)	(.287)	(.244)	(.967)	(.340)	(.225)
Fixed Effects	Firm	Analyst	IB	Firm	Analyst	IB
Adjusted R^2	.068	.107	.060	.068	.107	.060
N	136,193	136,193	136,193	136,193	136,193	136,193

Table 5 – Relative Recommendations based on a 3-Tier System

This table provides the results from estimating regressions of relative recommendations on investment bank relationship measures, investment bank characteristics, analyst characteristics, and stock characteristics Results for the sub-periods before (1998-2001) and after (2003-2009) Global Settlement period are provided in Panels A and B, respectively. In this table, relative recommendations are measured based on a 3-tier system where a strong buy or buy recommendations are coded as 3 and strong sell or sell recommendations are coded as 1. Columns 1 through 3 respectively use equity, debt, and M&A investment banking relationship measures, while column 4 uses an overall relationship measure. *p*-values based on robust standard errors are presented in parentheses below the coefficients, where standard errors are clustered by firm. Each model contains year and firm fixed effects. Variable definitions are contained in Appendix Table A1.

	Equity Relationship	Debt Relationship	M&A Relationship	Overall Relationship
	Panel A: 1998 – 2001			
Intercept	0.088	0.102	0.086	0.093
	(.144)	(.086)	(.149)	(.120)
IB Relationship Measures:				
IBRel_GS	0.032	0.080	0.044	0.073
	(.037)	(.000)	(.011)	(.000)
IBRel_NonGS	0.011	0.011	0.075	0.035
	(.659)	(.724)	(.018)	(.049)
IB Characteristics:				
Ln(IB_Size)	0.009	0.001	0.009	0.006
	(.138)	(.847)	(.155)	(.295)
IB_MktShare	0.033	0.338	0.109	0.082
	(.631)	(.000)	(.104)	(.251)
IB_NonGS	-0.013	0.002	-0.010	-0.005
	(.076)	(.824)	(.199)	(.546)
Analyst Characteristics:				
RelAccuracy	0.071	0.074	0.072	0.072
	(.011)	(.008)	(.010)	(.011)
AllStar	-0.008	-0.018	-0.010	-0.014
	(.379)	(.038)	(.240)	(.113)
Ln(Seniority)	-0.003	-0.003	-0.003	-0.004
	(.677)	(.736)	(.678)	(.642)
Ln(Seasoning)	0.016	0.015	0.016	0.016
	(.016)	(.023)	(.018)	(.019)
Ln(NFollow)	-0.021	-0.017	-0.020	-0.019
	(.000)	(.001)	(.000)	(.000)
JobMove	-0.021	-0.021	-0.020	-0.019
	(.019)	(.017)	(.020)	(.034)
Stock Characteristics:				
Ln(ANF)	-0.023	-0.023	-0.024	-0.023
	(.012)	(.014)	(.012)	(.014)
Ln(MV)	-0.019	-0.019	-0.019	-0.019
	(.003)	(.003)	(.003)	(.002)
Ln(Proceeds)	0.001	-0.001	0.000	0.000
	(.626)	(.547)	(.931)	(.875)
InstHoldings	-0.758	-0.756	-0.753	-0.750
	(.004)	(.005)	(.005)	(.005)
Adjusted R^2	.057	.059	.057	.058
N	59,703	59,703	59,703	59,703

Table 5 – continued

	Equity Relationship	Debt Relationship	M&A Relationship	Overall Relationship
		Panel B: 2003 – 2009		
Intercept	0.519	0.508	0.515	0.489
	(.000)	(.000)	(.000)	(.000)
IB Relationship Measures:				
IBRel_GS	0.030	0.036	0.048	0.042
	(.057)	(.007)	(.007)	(.000)
IBRel_NonGS	0.086	0.096	0.145	0.113
	(.001)	(.000)	(.000)	(.000)
IB Characteristics:				
Ln(IB_Size)	-0.057	-0.069	-0.061	-0.052
	(.000)	(.000)	(.000)	(.000)
IB_MktShare	-1.207	-0.381	-1.090	-1.375
	(.000)	(.000)	(.000)	(.000)
IB_NonGS	-0.042	0.000	-0.042	-0.048
	(.000)	(.979)	(.000)	(.000)
Analyst Characteristics:				
RelAccuracy	-0.026	-0.027	-0.018	-0.018
	(.349)	(.328)	(.507)	(.514)
AllStar	-0.011	-0.018	-0.014	-0.013
	(.207)	(.044)	(.113)	(.143)
Ln(Seniority)	0.015	0.015	0.014	0.015
	(.009)	(.009)	(.015)	(.011)
Ln(Seasoning)	0.005	0.006	0.005	0.004
	(.382)	(.291)	(.425)	(.510)
Ln(NFollow)	-0.019	-0.020	-0.013	-0.015
	(.000)	(.000)	(.006)	(.002)
JobMove	0.007	0.006	0.002	0.003
	(.512)	(.576)	(.811)	(.728)
Stock Characteristics:				
Ln(ANF)	-0.008	-0.008	-0.009	-0.008
	(.344)	(.327)	(.241)	(.303)
Ln(MV)	-0.029	-0.028	-0.027	-0.027
	(.000)	(.000)	(.000)	(.000)
Ln(Proceeds)	-0.001	0.000	-0.001	-0.001
	(.734)	(.846)	(.396)	(.434)
InstHoldings	-0.214	-0.211	-0.224	-0.205
	(.440)	(.447)	(.420)	(.460)
Adjusted R^2	.050	.047	.052	.053
N	136,193	136,193	136,193	136,193

Table 6 – Logit Models for Buy/Sell Recommendations

This table provides the results from estimating logistic regressions of the probability that an analyst issues a buy or strong buy (sell or strong sell) recommendation on overall investment bank relationship measures, investment bank characteristics, analyst characteristics, and stock characteristics in columns 1 to 3 (4 to 6). Results for the full sample period from 1998 to 2009 are presented in columns 1 and 4. The remaining columns present results for the sub-periods before (1998-2001) and after (2003-2009) Global Settlement. *p*-values based on robust standard errors are presented in parentheses below the coefficients, where standard errors are clustered by firm. Each model contains year and firm fixed effects. Variable definitions are contained in Table A1 of Appendix 1.

	Buy or Strong Buy			Sell or Strong Sell		
	Full Period	1998-2001	2003-2009	Full Period	1998-2001	2003-2009
Post	-0.741	-	-	1.879	-	-
	(.000)			(.000)		
IB Relationship Measures:						
IBRel_GS	0.529	0.455	-	-0.786	-0.579	-
	(.000)	(.000)		(.000)	(.130)	
IBRel_GS*Post	-0.345	-	0.178	0.520	-	-0.261
	(.000)		(.000)	(.015)		(.000)
IBRel_NonGS	0.400	0.256	-	-1.313	-0.612	-
	(.000)	(.030)		(.000)	(.144)	
IBRel_NonGS*Post	-0.107	-	0.324	0.513	-	-0.809
	(.318)		(.000)	(.168)		(.000)
IB Characteristics:						
Ln(IB_Size)	-0.190	-0.125	-0.172	0.251	-1.155	0.355
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
IB_MktShare	-2.763	0.663	-4.712	5.931	-1.266	5.708
	(.000)	(.077)	(.000)	(.000)	(.558)	(.000)
IB_NonGS	-0.243	-0.046	-	1.277	0.166	-
	(.000)	(.278)		(.000)	(.415)	
IB_NonGS*Post	0.192	-	-0.136	-1.007	-	0.362
	(.000)		(.000)	(.000)		(.000)
Analyst Characteristics:						
RelAccuracy	0.228	0.583	0.049	0.178	-0.927	0.411
	(.004)	(.000)	(.630)	(.253)	(.141)	(.013)
AllStar	-0.021	-0.017	-0.021	0.178	-0.165	0.185
	(.409)	(.712)	(.499)	(.000)	(.476)	(.000)
Ln(Seniority)	0.08	0.008	0.057	-0.167	-0.367	-0.140
	(.000)	(.844)	(.006)	(.000)	(.036)	(.000)
Ln(Seasoning)	-0.108	-0.104	-0.066	0.130	0.548	0.112
	(.000)	(.003)	(.001)	(.000)	(.001)	(.001)
Ln(NFollow)	-0.116	-0.149	-0.071	0.115	0.127	0.071
	(.000)	(.000)	(.000)	(.000)	(.349)	(.015)
JobMove	0.071	0.026	0.099	-0.027	0.103	-0.054
	(.009)	(.588)	(.005)	(.648)	(.593)	(.408)
Stock Characteristics:						
Ln(ANF)	-0.430	-0.599	-0.286	0.143	0.021	0.172
	(.000)	(.000)	(.000)	(.002)	(.914)	(.002)
Ln(MV)	0.653	0.833	0.627	-0.650	-0.534	-0.591
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Ln(Proceeds)	0.005	-0.023	0.011	0.001	0.000	0.006
	(.365)	(.062)	(.072)	(.89)	(.991)	(.552)
InstHoldings	0.066	0.177	0.053	-0.037	-0.042	-0.022
	(.000)	(.000)	(.000)	(.016)	(.440)	(.217)

Table 6 – continued

Combined Post Effects:						
GS Banks	0.184	-	-	-0.266	-	-
	(.000)			(.000)		
NonGS Banks	0.293	-	-	-0.800	-	-
	(.000)			(.000)		
Pseudo R^2	.078	.060	.027	.112	.163	.034
N	212,107	54,219	133,483	171,542	11,111	109,467

Table 7 – Analyst Affiliation Effects and Lending

This table provides results related to the incremental effects of lending relationships on analyst affiliation bias. Panel A provides descriptive statistics for the lending variables. Panels B and C presents the results from regressions of relative recommendations on overall investment banking and lending relationship measures, and a set of control variables related to investment bank, analyst, and stock characteristics, with results for the sub-period before Global Settlement (1998-2001) in Panel B and results for the post period (2003-2009) in Panel C. *p*-values based on robust standard errors are presented in parentheses below the coefficients, where standard errors are clustered by firm. Coefficients on the control variables are not reported. Each model contains year and firm fixed effects. Variable definitions are contained in Table A1 of Appendix 1.

Panel A – Summary Statistics

	N	Mean	Median	Min	Max	Std. Dev.
IB Relationship Measures:						
IBRel_Lending (%)	216,242	2.82	0.00	0.00	1.00	14.16
IBRel_Overall (+loan) (%)	216,242	5.84	0.00	0.00	1.00	18.38
IB Characteristics:						
IB_MktShare_Lending (%)	216,242	4.56	0.74	0.00	35.92	8.29
IB_MktShare_Overall (+loan) (%)	216,242	4.58	2.05	0.00	23.83	5.50
Firm/Stock Characteristics:						
Proceeds_Lending	216,242	964.14	40.00	0.00	73,197.78	2,730.11
Proceeds_Overall (+loans)	216,242	2,538.37	375.00	0.00	251,207.45	8,315.22
Proceeds_Lending^{+}	114,659	1,818.33	675.00	0.50	73,197.78	3,536.08
Proceeds_Overall (+loans)$^{+}$	164,818	3,330.35	798.75	0.50	251,207.45	9,385.00

Panel B: Regression Results, 1998–2001

IBRel_GS$_{Overall}$	-	0.108 (.000)	0.101 (.000)	-
IBRel_NonGS$_{Overall}$	-	0.080 (.023)	0.077 (.042)	-
IBRel_GS$_{Lending}$	0.095 (.008)	0.154 (.000)	-	-
IBRel_NonGS$_{Lending}$	0.110 (.009)	0.234 (.000)	-	-
IBRel_GS$_{Overall}$*IBRel_GS$_{Lending}$	-	-	0.176 (.000)	-
IBRel_NonGS$_{Overall}$*IBRel_NonGS$_{Lending}$	-	-	0.207 (.040)	-
IBRel_GS$_{Overall+Lending}$	-	-	-	0.093 (.000)
IBRel_NonGS$_{Overall+Lending}$	-	-	-	0.135 (.000)
Adjusted R^2	.058	.050	.049	.052
N	59,703	59,703	59,703	59,703

Table 7 – continued

Panel C: Regressions Results, 2003–2009				
IBRel_$GS_{Overall}$	-	0.028 (.035)	0.026 (.068)	-
IBRel_$NonGS_{Overall}$	-	0.159 (.000)	0.152 (.000)	-
IBRel_$GS_{Lending}$	0.025 (.246)	0.072 (.001)	-	-
IBRel_$NonGS_{Lending}$	0.064 (.113)	0.069 (.109)	-	-
IBRel_$GS_{Overall}$*IBRel_$GS_{Lending}$	-	-	0.067 (.008)	-
IBRel_$NonGS_{Overall}$*IBRel_$NonGS_{Lending}$	-	-	0.082 (.201)	-
IBRel_$GS_{Overall+Lending}$	-	-	-	0.030 (.014)
IBRel_$NonGS_{Overall+Lending}$	-	-	-	0.121 (.000)
Adjusted R^2	.067	.069	.068	.067
N	136,193	136,193	136,193	136,193

APPENDIX

Table A1 – Variable Definitions

Variable		Definition
Analyst Recommendation and Global Settlement Variables:		
$RelRec_{ijkt}$	=	Relative Recommendation. The most recent recommendation issued by analyst *i* (from investment bank *j*) for firm *k* during the one-year window ending in quarter *t*, normalized by subtracting the consensus (median) recommendation across all analysts covering firm *k* (whether or not they are in our sample) in the same one-year window.
$Post_t$	=	Post Global Settlement. An indicator variable that equals one for all quarters after the Global Analyst Research Settlement and zero otherwise. Following Kadan et al. (2009), we define the beginning of the post Global Settlement period as September 2002.
IB Relationship Measures:		
$IBRelC_{jkt}$	=	Investment Bank Relationship (Continuous). The proportion of a firm *k*'s total transaction value over the 36 months ending in quarter *t* for which investment bank *j* acted as a lead or co-managing underwriter or an M&A advisor. This variable is calculated separately based on equity, debt, and M&A transactions, as well as the combined set of transactions across all three areas.
$IBRel_{jkt}$	=	Investment Bank Relationship (Dummy). A dummy variable equal to one if *IBREL* for a particular transaction category (equity, debt, M&A, lending, or overall) is positive and zero otherwise.
IB Characteristics:		
IB_Size_{jt}	=	Investment Bank Size. The number of analysts employed by investment bank *j* during quarter *t*, according to the I/B/E/S recommendations file.
$IBMktShare_{jt}$	=	Investment Bank Market Share. The proportion of total deal value in a particular transaction category (equity, debt, M&A, lending, or all four combined) during the previous 12 months for which investment bank *j* acted as lead underwriter or advisor.
IB_GS_j (IB_NonGS_j)	=	Global Settlement (Non-Global Settlement) Investment Bank. Indicator variables to identify whether or not investment bank *j* was one of the 12 investment banks included in the Global Analyst Research Settlement (including subsequent name variations as shown in Appendix Table A2). The twelve investment banks included in the Global Settlement are: Bear Stearns; Citigroup (Salomon Smith Barney); CS First Boston; Deutsche Bank; Goldman Sachs; JP Morgan; Lehman Brothers; Merrill Lynch; Morgan Stanley; Thomas Weisel, UBS Warburg; and U.S. Bancorp Piper Jaffray.
Analyst Characteristics:		
$RelAccuracy_{ijt}$	=	Relative Analyst Accuracy. The relative forecast accuracy of the analyst, as defined in Hong and Kubik (2003). For each analyst *i* following firm *k*, we first estimate the absolute value of the difference between the analyst's most recent forecast of fiscal-year earnings (issued between January 1 and July 1 of year *t*) and actual earnings, scaled by price (as of the end of year *t-1*). We then rescale such that the most accurate analyst following firm *k* scores 1 and the least accurate analyst scores 0. Finally, each analyst's relative forecast accuracy is defined as the mean score across all stocks followed by the analyst over years *t*-2 through *t*.

Table A1 continued

$AllStar_{ijt}$	=	All Star Analyst. An indicator variable that equals 1 if the analyst is a ranked as an All-Star by *Institutional Investor* magazine during year *t-1*, and 0 otherwise.
$Seniority_{ijt}$	=	Analyst Seniority. The number of years since analyst *i* first appeared in I/B/E/S.
$Seasoning_{ijt}$	=	Analyst Seasoning. The number of years since analyst *i* initiated coverage of firm *k*, according to I/B/E/S.
$NFollow_{ijt}$	=	Number of Firms Followed. The number of firms followed by analyst *i* during quarter *t*, according to I/B/E/S.
$JobMove_{ijt}$	=	Analyst Job Move. An indicator variable that equals 1 if analyst *i* changed employers during quarter *t*, according to I/B/E/S.
Stock Characteristics:		
ANF_{kt}	=	Analyst Following. The number of analysts issuing recommendations for firm *k* during the previous 12 months, according to the I/B/E/S recommendations file.
MV_{kt}	=	Market Value. The market value of equity for firm *k* at the end of year *t-1*, according to CRSP.
$DealValue_{kt}$	=	Aggregate Deal Value. The total deal value by firm *k* in a particular transaction category (equity, debt, M&A, lending, or all four combined) during the previous 36 months.
$InstHoldings_{kt}$	=	Institutional Holdings. The percentage of shares of firm *k* held by institutional investors at the end of quarter *t*, according to Thomson Reuters' 13F filings.

Table A2 – Sample Investment Banks

This table lists the investment banks included in our final sample, including all predecessor banks in the case of mergers. Investment Banks that were sanctioned in the Global Settlement and subsequent name variations that are also treated as sanctioned banks in our analysis are listed in bold type. Merrill Lynch and Lehman were included in the Global Settlement but are not included in our sample because they are missing from the I/B/E/S data for all or part of our sample period.

Ultimate IB Name	**Predecessor IBs**
Sanctioned Banks:	
Bank of America Merrill Lynch	Advest; Banc America; Bank of America; **Bank of America Merrill Lynch**
Citigroup Salomon Smith Barney	Schroder; **Salomon Smith Barney**; **Citigroup Salomon Smith Barney**
CS First Boston	DLJ; **CS First Boston**
Deutsche Alex Brown	**Deutsche Bank**; **Deutsche Alex Brown**
Goldman Sachs	**Goldman Sachs**
JP Morgan Chase	**Bear Stearns**; Chase HQ; Robert Flemming; **JP Morgan**; **JP Morgan Chase**
Morgan Stanley Dean Witter	**Morgan Stanley**; **Morgan Stanley Dean Witter**
Thomas Weisel	**Thomas Weisel**
UBS Paine Webber[a]	JC Bradford; **Paine Webber**; **UBS**; **UBS Warburg**; **UBS Paine Webber**
US Bancorp Piper Jaffray	US Bancorp; **Piper Jaffray**; **US Bancorp Piper Jaffray**
Non-Sanctioned Banks:	
ABN AMRO	ABN AMRO
BNP Paribas	Paribas; BNP Paribas
CIBC	CIBC
Commerzbank	Dresdner Kleinwort; Commerzbank
Friedman	Friedman
HSBC	HSBC
ING Barings Furman	ING Barings Furman
Lazard	Lazard
Needham	Needham
Prudential Securities	Vector Securities; Volpe Brown Whelan; Prudential Securities
Raymond James	Raymond James
RBC Capital Markets	Dain Rauscher Wessels; Ferris; Tucker Anthony Sutro; RBC Capital Markets
Robert Baird	Robert Baird
Scotia	Scotia
SG Cowen	Societe Generale; SG Cowen
Stephens	Stephens
Sun Trust Robinson	Sun Trust Equitable; Sun Trust Robinson
Wells Fargo	Black; JW Charles; Everen; First Union; First Van Kasper; Wachovia; Wachovia Corp; Wells Fargo
William Blair	William Blair

[a] In the case of UBS Paine Webber, occurrences of UBS, UBS Warburg, and Paine Webber prior to the UBS-Paine Webber merger are also classified as sanctioned banks. These three investment banks account for only 191 (0.09%) of the quarterly observations in our analysis.

Expectations and the Structure of Share Prices

John G. Cragg and
Burton G. Malkiel

The University of Chicago Press

Chicago and London

Table 2.16 **Analysis of Forecasts by Industrial Category: 1963 Predictions vs. 1963–68 Actual Earnings**

Pred.	Correlation	T	T^M	T^{BI}	T^{WI}	No. of Observations
1	.21	.75	.32	.23	.63	173
2	.25	.73	.31	.20	.62	171
3	.48	.66	.31	.18	.55	122
4	.75	.46	.05	.21	.41	59
5	.42	.62	.12	.17	.58	172
6	.69	.45	.07	.11	.43	37
7	.51	.58	.16	.22	.51	60
g_{p1}	.42	.65	.07	.26	.59	153
g_{p2}	.39	.71	.09	.32	.63	131
g_{p3}	.47	.66	.04	.19	.63	121
g_{p4}	.45	.77	.04	.17	.75	156

would be 4 for the most difficult industry (in years when there were four predictors compared), 8 for the next most difficult, and so on. In this case, the coefficient of concordance (Kendall's W) would be unity. The values of Kendall's W were significantly different from zero beyond the 0.05 level for most of the years as were differences between industries for the correlation coefficients for most of the predictors.[11] These findings indicate that there were industry differences. For the long-term predictions, correlation coefficients between forecasts and realizations tended to be highest in the oil, food and stores, and "cyclical" industries. For the short-term predictions, there was really no industry that was particularly easy to predict compared with the others; that is, prediction performances were uniformly mediocre across industries.

The electric utility industry turned out to be one of the more difficult industries for which to make long-term forecasts. This would come as a distinct surprise to the participating security analysts who claimed at the outset that they had some reservations about their abilities to predict earnings for the metals and other "cyclical" companies, but had confidence that they could make accurate predictions for the utilities.[12] It turned out that the long-term predictions for the utility industry were considerably worse than for the metals and "cyclicals."

In general, we had little success in associating forecasting performance with industry or company characteristics. Forecasting differences between industries were only moderately related to the average realized

11. The latter was tested on the basis of the asymptotic distribution of the correlation coefficient and the assumption that the data were distributed normally.

12. This confidence was also reflected in the fact that for the electric utility industry there was high agreement among the forecasters, whereas agreement was relatively low for the cyclical group.

WILEY-BLACKWELL

American Finance Association

The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings
Author(s): J. G. Cragg and Burton G. Malkiel
Reviewed work(s):
Source: *The Journal of Finance*, Vol. 23, No. 1 (Mar., 1968), pp. 67-84
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2325310
Accessed: 21/06/2012 10:36

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Blackwell Publishing and *American Finance Association* are collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Finance.*

STOR®
http://www.jstor.org

THE CONSENSUS AND ACCURACY OF SOME PREDICTIONS OF THE GROWTH OF CORPORATE EARNINGS

J. G. CRAGG* AND BURTON G. MALKIEL*

FOR YEARS ECONOMISTS HAVE EMPHASIZED the importance of expectations in a variety of problems.[1] The extent of agreement on the significance of expectations is almost matched, however, by the paucity of data that can be considered even reasonable proxies for these forecasts. One area in which expectations are highly important is the valuation of the common stock of a corporation. The price of a share is—or should be—determined primarily by investors' current expectations about the future values of variables that measure the relevant aspects of corporations' performance and profitability, particularly the anticipated growth rate of earnings per share.[2] This theoretical emphasis is matched by efforts in the financial community where security analysts spend considerable effort in forecasting the future earnings of companies they study. These forecasts are of particular interest because one can observe divergence of opinion among different individuals dealing with the same quantities. This paper is devoted to the analysis of a small sample of such predictions and certain related variables obtained from financial houses.[3]

I. NATURE AND SOURCES OF DATA

The principal data used in this study consisted of figures representing the expected growth of earnings per share for 185 corporations[4] as of the end of 1962 and 1963. These data were collected from five investment firms. The participants were recruited through requests to two organizations. One was a group of firms who used computers for financial analysis and who met periodically to discuss mutual problems, the other was the New York Society of

* University of British Columbia and Princeton University, respectively. This Research was supported by the Institute for Quantitative Research in Finance, the National Science Foundation, and the Graduate School of Business, University of Chicago. We are indebted to Paul Cootner for helpful comments.

1. A number of studies of anticipations data have been collected in two National Bureau Volumes [12] and [13]. Some more recent work on the assessment of expectations or forecasts has been done by Zarnowitz [16].

2. The classic theoretical statement of the anticipations view of the determination of share valuation may be found in J. B. Williams [15]. This position is also adopted in the standard textbook in the field [3]. The emphasis on the importance of earnings growth may also be found in [4], [5], and [19].

3. One of the few attempts to conduct a study of this type was made by the Continental Illinois Bank and Trust Company of Chicago [1] in 1963. The bank collected a sample of earnings estimates one year in advance from three investment firms. An analysis of these projections revealed that the financial firms tended to overestimate earnings and that over-all quality of the estimates tended to be poor.

4. The 185 companies for which the growth-rate estimates were made tended to be the large corporations in whose securities investment interest is centered. This selection was made on the basis of availability of data and was not chosen as a random sample.

Financial Analysts. As a result, eleven firms agreed to participate in the proposed study. From the original eleven, however, only five were able to supply comparable sets of long-term earnings forecasts for use in this study.[5] Even among these five there was not complete overlap in the corporations for which predictions were available. One of them had no data for 1962. For only two were data available for the full set of 185 companies.

Of the five participating firms, two are large New York City banks heavily involved in trust management, one is an investment banker and investment adviser doing mainly an institutional brokerage business, one is a mutual fund manager, and the remaining firm does a general brokerage and investment advisory business. We would not argue that these estimates give an accurate picture of general market expectations. It would, however, seem reasonable to suggest that they are representative of opinions of some of the largest professional investment institutions and that they may not be wholly unrepresentative of more general expectations. Since investors consult professional investment institutions in forming their own expectations, individuals' expectations may be strongly influenced—and so reflect—those of their advisers.[6] Also, insofar as investors follow the same sorts of procedures as those used by security analysts in forming expectations, the investors' expectations would resemble those of the analysts. It should be noted, however, that security analysts are not limited to published data in forming their expectations. They frequently visit the companies they study and discuss the corporations' prospects with their executives.

Each growth-rate figure was reported as an average annual rate of growth expected to occur in the next five years. At first thought, such a rate of growth depends on what earnings are expected to be in five years' time and on the base-year earnings figures. However, this dependence need not be very great if the growth rate is regarded more as a parameter of the process determining earnings than as an arithmetic quantity linking the current value to the expected future value. Discussion with the suppliers of the data indicated that all firms were attempting to predict the same future figure, the long-run average ("normalized") earnings level, abstracting from cyclical or special circumstances. The bases used were less clear. Some firms explicitly used their estimates of "normalized" earnings during the year in which the prediction was made. Others provided different figures as bases: in one case the firm estimated actual earnings, in another a prediction of earnings four years in the future was furnished. These differences did not seem to be reflected in the growth rates, however, since attempts to adjust the rates for differences in

5. We are deeply grateful to the participating firms, who wish to remain anonymous. Not all volunteers were able to supply data useful to this study, either because the actual supply of data would have been too burdensome (being kept for internal records in a form that made their extraction difficult) or because the data supplied were not comparable to data used here (either being of a short-term nature or being made at different dates). Because one of our main objectives is to examine differences and similarities in predictions of the same quantities, such data were not used in the present paper.

6. That several of our participating firms find it worthwhile to publish these projections and provide them to their customers provides *prima facie* evidence that a certain segment of the market places some reliance on such information in forming its own expectations.

base figures introduced rather than removed disparities among the predictions.

The growth rates were given as single numbers for each corporation. No indication was provided of the confidence with which these point estimates were held. One firm did provide an instability index of earnings which represented a measure of the past variability of earnings (around trend) adjusted by the security analyst to indicate potential future variability. Moreover, two firms provided quality ratings, which classified companies into three or four quality categories.

Two of the firms provided estimates of past growth rates as well as predictions. The figures represented perceived growth over the past 8-10 years, the past 4-5 years, the past 6 years, and the last year. It may seem unnecessary to rely on the participating firms for estimates of historic growth rates. However, the past growth of a company's earnings is not, in any meaningful sense, a well-defined concept. Earnings—being basically a small difference between two large quantities—can exhibit large year-to-year fluctuations. They also can be negative, which creates problems for most mechanical calculations. In addition, the accounting definition of earnings is not an exact conformity with the economically relevant concept of profits or return on investors' capital. For these reasons, calculated growth rates are sensitive to the particular method employed and the period chosen for the calculation. Consequently, such calculations may be a poor reflection of what growth is generally considered to have been, and may not be useful in assessing the past performance of corporations. Furthermore, it may be supposed that in assessing security analysts' predictions of growth their own estimates of past growth are more likely to be relevant than objectively calculated rates. The extent of agreement among the two types of measures is among the subjects considered in the next section.

Our participating firms also supplied an industrial classification. While other classifications are available, the concept of industry is not really precise enough to get a fixed, unquestionable assignment of corporations to industries. Particular problems are presented by conglomerate companies. Perceived industry may be more relevant than any other grouping when investigating anticipations. The classification we use represents a consensus about industry among our participants. Where disagreements occurred (as was often the case with conglomerates), the corporation was simply classified as "miscellaneous." The classification represented considerable aggregation over finer classifications and only eight industries were distinguished. These were:

1) Electricals and Electronics
2) Electric Utilities
3) Metals
4) Oils
5) Drugs and Specialty Chemicals
6) Foods and Stores
7) "Cyclical"—including companies such as automobile and aircraft manufacturers, and meat packers
8) "Miscellaneous"

II. Agreement among Predictors

The agreement among the growth-rate projections is described and summarized in this section. In the course of this description, the extent of agreement about base-earnings figures and the closeness of the projections to past, perceived, and calculated growth rates are also considered.

A. *Comparisons of Predictions of Future Growth Rates.*

The extent of agreement among the predictors about future growth rates is summarized in Table 1. Of the five predictors, the correlations among predictors A, B, C and E were all roughly of the same orders of magnitude.[7] Predictor D showed some tendency towards lower agreement. (Predictor D also had the highest average growth forecast and standard deviation for the companies for which it and others made forecasts.) Over-all agreement among

TABLE 1
Agreement Among Growth-Rate Predictions*

I. Correlation Coefficients

(Simple correlations in lower left portion, Spearman rank correlations in upper right portion)

	1962					1963				
	A	B	C	D		A	B	C	D	E
A	1.000	.768	.751	.388	A	1.000	.795	.717	.374	.709
B	.840	1.000	.728	.597	B	.832	1.000	.760	.518	.821
C	.889	.819	1.000	.690	C	.854	.764	1.000	.750	.746
D	.563	.621	.848	1.000	D	.537	.567	.898	1.000	.450
					E	.827	.835	.889	.704	1.000

II. Kendall's Coefficient of Concordance for Ranks of Companies by Different Predictors

Predictors	(A,B,C)	(A,B,D)	(A,B,C,D)	(A,B,C,D,E)
1962	.82	.73	.78	
1963	.83	.71	.81	.79

III. Proportions of Total Variance Due to Variance in Average Predictions

Predictors	(A,B,C)	(A,B,D)	(A,B,C,D)	(A,B,C,D,E)
1962	.87	.70	.79	
1963	.85	.68	.83	.87

* The numbers of observations on which this table and other tables are based varies between cells. For the correlations, the numbers of observations are reported below:

	1962				1963			
	A	B	C		A	B	C	D
B	185			B	185			
C	60	60		C	62	62		
D	178	178	58	D	182	182	61	
				E	125	125	39	124

For other comparisons, the number of observations is the minimum of the numbers of observations used to compute the correlations.

7. The analysis is presented mainly for the raw growth figures, but very similar impressions would be obtained from examining their logarithms.

the predictors is further summarized in the second and third parts of Table 1, which show the values of Kendall's coefficient of concordance and the proportion of total variance of the predictions that can be accounted for by differences in the mean prediction among companies.[8] It may be remarked that the entries in Table 1 are based on different numbers of observations. In each case, we used the maximum number of observations (companies) for which a comparison could be made. The impressions to be gained from Table 1 would be little changed, however, by basing all calculations only on the set for which all predictors provided data.

Though Table 1 suggests considerable agreement, the lack of agreement it also reveals can hardly be considered negligible. In addition to the lack of correlation, there were also some systematic differences among the predictors. For the matched set of observations the means and the standard deviations were of roughly the same sizes. However, the differences among the central tendencies were significant according to both parametric and nonparametric tests.

B. *Analysis of Predictions Within Industrial Classifications.*

One might suspect that the correlations among the predictors reflect little more than consensus about the industries that are expected to grow most rapidly rather than agreement about the relative rates of growth of firms within industries. This possibility was investigated by decomposing the correlation coefficients into two parts, one due to correlation within industries (r_w) and one due to correlation among the industry means (r_a).

$$r = r_w + r_a$$

where

$$r_w = \frac{\sum_{j=1}^{J}\sum_{i=1}^{N_j}(x_{ij}-\bar{x}_j)(y_{ij}-\bar{y}_j)}{\sqrt{\sum_{j=1}^{J}\sum_{i=1}^{N_j}(x_{ij}-\bar{x})^2\sum_{j=1}^{J}\sum_{i=1}^{N_j}(y_{ij}-\bar{y})^2}},$$

and

$$r_a = \frac{\sum_{j=1}^{J}N_j(\bar{x}_j-\bar{x})(\bar{y}_j-\bar{y})}{\sqrt{\sum_{j=1}^{J}\sum_{i=1}^{N_j}(x_{ij}-\bar{x})^2\sum_{j=1}^{J}\sum_{i=1}^{N_j}(y_{ij}-\bar{y})^2}}$$

with

8. The values shown in all parts of Table 1 are significant well beyond the conventionally used levels of significance. We may note that Tukey's test for interaction in a two-way analysis of variance [11, pp. 129-37]—the typical model in which the breakdown of variance used in Part 3 of Table 1 is employed—indicated a small but highly "significant" proportion of variance attributable to interaction. However, the usual analysis-of-variance model does not seem appropriate for this data, not only because of interactions, but also because of possible lack of homogeneity of variance.

x_{ij}, y_{ij}	being the i^{th} observations in the j^{th} class (industry),
N_j	being the number of observations in the j^{th} class,
J	being the number of classes,
$\bar{x}_j$, $\bar{y}_j$	being the averages within the classes, and
$\bar{x}$, $\bar{y}$	being the over-all averages.

This decomposition indicated that agreement concerning industry growth rates is not the major factor accounting for the correlations among the forecasts. The first part of Table 2 shows the values of r_a using the industrial classification obtained from the participating firms. As comparison with Table 1 shows, only a small part of the correlations among the predictions are due to correlations among the industry means. Further light can be shed on this question by calculating the partial correlations between the predictions, holding industry classification constant. The second panel of Table 2 reveals

TABLE 2
INDUSTRIAL CLASSIFICATION AND AGREEMENT AMONG PREDICTORS

I. Values of r_a

	1962				1963			
	A	B	C		A	B	C	D
B	.299			B	.305			
C	.285	.323		C	.230	.315		
D	.090	.184	.300	D	.057	.137	.317	
				E	.266	.348	.366	.194

II. Partial Correlations Holding Industrial Classification Constant

	1962				1963			
	A	B	C		A	B	C	D
B	.799			B	.786			
C	.861	.760		C	.838	.690		
D	.656	.665	.887	D	.657	.650	.861	
				E	.828	.790	.897	.777

that these partial correlations tended to be only slightly less than the simple correlations and, in the case of Predictor D, the partial correlations were actually higher.

It is also interesting to examine the extent to which the correlations among predictors' forecasts varied over the different industry groups. This should indicate whether certain industry groups are more difficult to forecast in an *ex ante* sense. The correlations among forecasters tended to be lowest in the oil and cyclical industry groups, and highest for electric utility companies. These differences were significant for all pairs of predictions considered. Ranking the correlations over industries, and then comparing these ranks among pairs of predictors, showed substantial concordance over the ordering of the correlations.[9]

9. The test for individual pairs of predictions was the likelihood-ratio test. Note that the ranking comparison is not based on independent observations so a statistical test of the concordance is not appropriate. This suggests that the "significance" of the over-all correlations mentioned earlier should really be treated only as descriptive indications of their sizes. The hypothesis that

C. *Comparisons of Predictions and Past Growth Rates.*

The extent of agreement among the predictors can usefully be evaluated by comparisons of the predicted growth rates with earlier predictions and with the past growth rates of earnings. The correlations of the 1963 predictions with the 1962 ones were: .94, .95, .96, and .88 for predictors A through D respectively. All of these are considerably higher than the correlations of the predictions with each other. On the other hand, changes in expected growth rates were not highly correlated among predictors.[10]

TABLE 3
PREDICTIONS AND PAST GROWTH RATES*
(CORRELATIONS OF PREDICTED WITH PAST GROWTH RATES)

	1962				1963				
	A	B	C	D	A	B	C	D	E
g_{p1}	.78	.68	.75	.41	.85	.73	.84	.56	.67
g_{p2}	.75	.67	.72	.51	.79	.69	.80	.58	.76
g_{p3}	.77	.71	.82	.61	.75	.72	.79	.70	.74
g_{p4}	.34	.37	.59	.44	.33	.45	.70	.75	.58
g_{c1}	.55	.46	.65	.32	.63	.52	.61	.30	.58
g_{c2}	.67	.60	.68	.18	.72	.58	.73	.20	.56
g_{c3}	.75	.63	.73	.17	.79	.66	.76	.17	.57
g_{c4}	.82	.68	.79	.24	.83	.69	.79	.29	.60

* g_{p1} is 8-10 year historic growth rate supplied by A
g_{p2} is 4-5 year historic growth rate supplied by A
g_{p3} is 6 year historic growth rate supplied by D
g_{p4} is preceding 1 year growth rate supplied by D
g_{c1} is log-regression trend fitted to last 4 years
g_{c2} is log-regression trend fitted to last 6 years
g_{c3} is log-regression trend fitted to last 8 years
g_{c4} is log-regression trend fitted to last 10 years.

Correlations of the predictions with eight past growth figures are shown in Table 3. Four of these past growth rates were supplied by the participating firms and represent the firms' perceptions of the growth of earnings per share that had occurred in different preceding periods. The others were calculated as the coefficient in the regression of the logarithms of earnings per share on time over the past 4, 6, 8, and 10 years. These correlations generally are not much lower than those found in comparing the predictions with each other. Among the perceived past growth rates, the correlations are apt to be lowest with the growth rates over the most recent year. With the calculated growth rates, there

the correlations are all zero within industries could, however, be rejected well beyond conventional significance levels. Predictor C was dropped from these tests due to paucity of data in many industries.

10. These correlations, for the participants supplying data in both years were:

	A	B	C
B	.19		
C	.04	.04	
D	.07	.11	.29

Only the two largest of these correlations would be significant at the .05 level.

was a tendency for the correlations to increase with the length of period over which the calculations were made.[11]

These comparisons of past with predicted growth rates suggest that the apparent agreement among the predictors may reflect little more than use by all of them of the historic figures. In investigating this possibility, the partial correlations among the predictions, holding constant past perceived growth rates, holding constant past calculated growth rates, and holding both sets constant were calculated. The first two sets of partial correlations were not much smaller than the simple correlations. Holding both sets constant produced the partial correlations shown in Table 4. These are considerably

TABLE 4
PARTIAL CORRELATIONS OF PREDICTIONS
HOLDING PAST GROWTH RATES CONSTANT

	1962				1963			
	A	B	C		A	B	C	D
B	.49			B	.49			
C	.49	.18		C	.25	.03		
D	.35	.39	.22	D	.56	.46	.40	
				E	.56	.62	−.11	.51

NUMBERS OF OBSERVATIONS

	1962				1963			
	A	B	C		A	B	C	D
B	111			B	112			
C	49	49		C	50	50		
D	111	111	49	D	112	112	50	
				E	78	78	36	78

smaller than the simple correlations, though all but the four smallest entries would be significant beyond the .05 level. Thus, while a substantial part of the agreement among predictors appears to result from their use of historic growth figures, there is also evidence that security analysts tend to make similar adjustments to the past growth rates.[12]

Examination of the correlations among past growth rates help both to evaluate the correlations among the predictions and to indicate the sensitivity of measurements of growth rates to the methods by which they were calculated. Table 5 presents correlations between 13 such past growth rates for our 1962 data. The correlations between the different measures of past growth are fairly low. When exactly the same data are used in the calculations, however, the

11. This effect was also found when the calculated growth rates were based on either 1) the regression of earnings per share on time; or, 2) the appropriate root of the ratio of earnings per share at the end of the period to earnings at the beginning.

12. The numbers of observations on which Table 4 is based are considerably smaller than those for which predictions were available. Only a small part of this loss was due to inability to calculate past growth rates due to negative earnings figures. Much more important was the fact that the predictors did not give numerical figures for past growth rates when these would be negative. One might think that the companies for which past growth rates were easily calculated would be ones with highest simple correlations among the predictors. However, the only cases for which this appeared to be true were the correlations of predictor D with A, B, and E.

correlations among the growth rates calculated by different methods are relatively high, though probably not so high that the choice of method of calculation would be a matter of no importance. Finally, the perceived growth rates furnished by the security firms tend to be more highly correlated with the growth rates calculated over longer periods. The increase in correlation coefficients did not continue, however, when calculations over more than ten years were made and, as shown in Table 5, it stopped before ten years in some cases. Correlations for other periods and for the 1963 data were of about the same magnitude as those in Table 5.

TABLE 5
PAST GROWTH CORRELATIONS, 1962*

	g_{p1}	g_{p2}	g_{p3}	g_{p4}	g_{c1}	g_{c2}	g_{c3}	g_{c4}	g_{c5}	g_{c6}	g_{c7}	g_{c8}
g_{p2}	.70											
g_{p3}	.82	.87										
g_{p4}	.49	.39	.37									
g_{c1}	.34	.47	.48	.15								
g_{c2}	.68	.74	.76	.05	.62							
g_{c3}	.81	.89	.97	.15	.49	.90						
g_{c4}	.93	.80	.87	.27	.41	.75	.93					
g_{c5}	.14	.19	.25	.39	.38	.24	.16	.15				
g_{c6}	.34	.46	.47	.14	.96	.59	.45	.37	.53			
g_{c7}	.92	.67	.78	.32	.48	.67	.83	.95	.33	.46		
g_{c8}	.36	.56	.49	.23	.99	.63	.50	.43	.40	.90	.51	
g_{c9}	.87	.75	.88	.18	.46	.77	.93	.99	.17	.40	.91	.43

* $g_{p1} - g_{p4}$, $g_{c1} - g_{c4}$ as defined in footnote to Table 3
g_{c5} is 1 year growth rate calculated from first differences of logarithm
g_{c6} is 4 year growth rate calculated from average of first differences of logs
g_{c7} is 10 year growth rate calculated from average of first differences of logs
g_{c8} is 4 year growth rate calculated from regression of earnings on time
g_{c9} is 10 year growth rate calculated from regression of earnings on time

D. *Comparisons of Predictions with Price-Earnings Ratios.*

Finally, we may examine the extent of agreement among predictors by comparing their forecasts with the price-earnings ratios of the corresponding securities. By utilizing a normative valuation model (see e.g., [4] or [8]) it is possible to calculate an implicit growth rate from the market-determined earnings multiple of a security. Thus, comparisons of the predictions with price-earnings ratios may be interpreted as examinations of the relationship between the forecasts and market-expected growth rates. Correlations with two versions of the price-earnings ratio are shown in Table 6. The prices used were the closing prices for the last day of the year. The earnings were either the actual earnings or the average of the base-earnings figures supplied by A and B for their growth rates. These latter figures represent "normalized" or trend-earnings figures. Specifically, they represent an attempt to estimate what earnings would be in the absence of cyclical or special factors. The correlation coefficients in the table are about the same as those obtained when the forecasts were compared with each other. Since price-earnings ratios are

TABLE 6
CORRELATIONS OF PREDICTIONS WITH PRICE-EARNINGS RATIOS*

	1962				
	A	B	C	D	
P/E	.76	.80	.86	.56	
P/NE	.82	.83	.83	.55	
	1963				
	A	B	C	D	E
P/E	.77	.74	.86	.67	.85
P/NE	.81	.76	.80	.60	.85

* P/E is the price/earnings ratio. P/NE is price/average of base (normalized) earnings of A and B.

affected by several variables other than expected growth rates, this exercise underscores the extent of disagreement among the forecasters.

III. Accuracy of Predictions

In assessing the forecasting abilities of the predictors, we encountered one major difficulty. The five years in the future for which the forecasts were made have not yet elapsed. As a result, we were forced to compare the forecasts with the realized growth of actual and normalized earnings (as estimated by Predictors A and B) through 1965. Since the latter figures represent what earnings are thought to be on their long-run growth path, perhaps not too much violence is done to the intentions of the forecasters by making these a standard of comparison.

A. *Method of Evaluation.*

The forecasts were evaluated by the use of simple correlations and by the inequality coefficient,[13]

$$U^2 = \frac{\Sigma(P_i - R_i)^2}{\Sigma R_i^2}, \tag{1}$$

where P_i is the predicted and R_i the realized growth rates for the i^{th} company. It will be noticed that the inequality coefficient, in effect, gives a comparison between perfect prediction ($U^2 = 0$) and a naive prediction of zero growth for all corporations ($U^2 = 1$).

We also investigated the extent to which errors in predictions were related to 1) errors in predicting the average over-all earnings growth of the sample firms; 2) errors in predicting the average growth rate of particular industries; and 3) errors in predicting the growth rates of firms within industries. To accomplish this, we decomposed the numerator of (1) into three parts. The first comes from the average prediction for all companies not being equal to the average realization. The second part arises from differences among the

13. Note that this is similar to the inequality coefficient introduced by Theil [14].

average industry predictions not being equal to the corresponding differences in industry realizations. The third arises from the differences in predictions for the corporations within an industry not being the same as the differences in realization.[14] The proportions of U^2 arising from these three sources will be called U^M, U^{BI}, and U^{WI} respectively for mean errors, between-industry errors, and within-industry errors.

B. *Over-all Accuracy of the Forecasts.*

Statistics summarizing the forecasting abilities of the predictors and the success of using perceived past growth rates to predict the future are presented in Table 7. By and large, the correlations of predicted and realized growth rates are low, though most of them are significantly greater than zero, and the inequality coefficients are large. The major exception to this is Predictor C's forecasts. However, this apparent superiority is largely illusory since C tended to concentrate on large, relatively stable companies and, we suspect, predictions were made only when there was *a priori* reason to believe that the forecasts would be reliable. That this conjecture has some validity is borne out by the fact that the set of companies for which C made forecasts had a lower average instability index than did our whole sample. Moreover, all the other forecasts, including the perceived past growth rates, did better for this set of companies than for the larger set.[15]

Several additional points about the over-all accuracy of the forecasts are worth mentioning. First, the forecasts based on perceived past growth rates, including even growth over the most recent year, do not perform much differently from the predictions. There seems to be no clear-cut forecasting advantage to the careful and involved procedures our predictors employed over their perceptions of past growth rates either in terms of correlation or of the inequality coefficient.

Second, all predictors had a better record than the no-growth forecast for each company. However, it is possible to find a single growth rate that would yield lower mean square errors than any of the predictions. This is a result of the average realized growth rates being considerably higher than the average

14. Letting P_{kj} and R_{kj} be the predicted and realized growth rates for the k^{th} company ($k = 1, \ldots, N_j$) in the j^{th} industry ($j = 1, \ldots, J$), we can write the numerator of (1) as:

$$\sum_{j=1}^{J}\sum_{k=1}^{N_j}(P_{kj}-R_{kj})^2 = \left[\sum_{j=1}^{J} N_j(\bar{P}-\bar{R})^2\right] + \left[\sum_{j=1}^{J} N_j\{(\bar{P}_j-\bar{P})-(\bar{R}_j-\bar{R})\}^2\right] + \left[\sum_{j=1}^{J}\sum_{i=1}^{N_j}\{(P_{kj}-\bar{P}_j)-(R_{kj}-\bar{R}_j)\}^2\right],$$

when $\bar{P}_j$, $\bar{R}_j$ are the averages for the j^{th} industry and $\bar{P}$ and $\bar{R}$ are the overall means. The three terms in square brackets are the ones referred to in the text.

15. For this smaller group of companies, the differences among predictors was far less than is suggested by Table 7. It is worth noting that C had a higher correlation and lower inequality index than the others in 1962 (with D a very close second), but both D and E were slightly better on the matched set in 1963.

TABLE 7

ACCURACY OF PREDICTIONS

I. 1962 Predictions Compared with Growth of Actual Earnings 1962-1965

Predictor	A	B	C	D	g_{p1}	g_{p2}	g_{p3}	g_{p4}
Correlation	.07	.16	.66	.45	.22	−.01	.23	.16
U	.80	.78	.57	.67	.74	.88	.74	.78
U^{M}	.31	.32	.20	.24	.17	.12	.10	.20
U^{BI}	.11	.10	.08	.06	.11	.04	.04	.12
U^{WI}	.58	.58	.71	.70	.73	.84	.75	.68
Number of Observations	185	185	60	178	168	140	140	145
II. 1962 Predictions Compared with Growth of Normalized Earnings 1962-1965								
Correlation	.26	.32	.68	.45	.23	.16	.38	.09
U	.74	.72	.57	.62	.72	.80	.67	.76
U^{M}	.25	.25	.08	.13	.09	.12	.09	.19
U^{BI}	.07	.06	.06	.08	.08	.07	.05	.08
U^{WI}	.68	.69	.86	.79	.83	.80	.86	.73
Number of Observations	180	180	59	175	164	136	138	142

III. 1963 Predictions Compared with Growth of Actual Earnings 1963-1965

Predictor	A	B	C	D	E	g_{p1}	g_{p2}	g_{p3}	g_{p4}
Correlation	.05	.16	.78	.47	.29	.20	.31	.22	.55
U	.85	.84	.59	.73	.81	.78	.75	.77	.62
U^{M}	.33	.34	.27	.28	.40	.20	.19	.16	.27
U^{BI}	.12	.11	.11	.07	.11	.09	.06	.06	.05
U^{WI}	.54	.55	.62	.66	.49	.70	.74	.79	.69
Number of Observations	185	185	62	182	125	167	143	138	169
IV. 1963 Predictions Compared with Growth of Normalized Earnings 1963-1965									
Correlation	.27	.29	.70	.34	.49	.36	.52	.41	.32
U	.78	.78	.61	.70	.74	.69	.64	.67	.69
U^{M}	.35	.35	.22	.23	.40	.22	.33	.23	.12
U^{BI}	.07	.06	.08	.09	.09	.08	.09	.05	.06
U^{WI}	.58	.59	.70	.68	.50	.70	.57	.72	.82
Number of Observations	180	180	61	177	123	163	139	136	165

expectation of each predictor. This may simply indicate a failure to anticipate the continuation of the expansion through the period considered, but it may also reflect the underestimation of change frequently found in investigating forecasts.[16]

Third, with the exception of the past growth rate in the year immediately preceding the forecast date, all predicted and perceived past growth rates were better at predicting the average normalized growth rates than the actual ones. However, whether this is because normalized earnings gave a better picture

16. See, for example, Zarnowitz [16]. Since almost all the actual growth rates were positive, we do not know whether underestimation of change would also characterize predictions when earnings were generally declining. No forecasters predicted a negative rate of growth.

of the true growth of corporations or because normalized earnings calculations are influenced by past growth-rate forecasts is open to question.

C. *Analysis of the Forecasts by Industrial Categories.*

Turning to the industry breakdown of the forecasts, we find that failure to forecast industry means (U^{BI}) accounted for only a very small proportion of the inequality coefficient. The main sources of inequality were the within-industry errors.

Looking at the correlations of predictions with future growth rates within industries permits us to assess which industries were most difficult to forecast in an *ex post* sense. The extent to which forecasters found the various industries difficult to predict is indicated in Table 8. To calculate the table, we first ranked each predictor's correlation coefficients between his forecasts and realizations over the eight industry groups. The industry for which the predictor had the most difficulty (worst correlation) was given a rank of one. In Table 8, we present the sums of the ranks for each industry over the four predictors.[17] If the difficulty ranking for all predictors was identical, the rank totals would be 4 for the most difficult industry (in 1963 when there are four predictors compared), 8 for the next most difficult, etc., and the coefficient of concordance (Kendall's W) would be unity. For each of the sets presented, the values of Kendall's W are significant (beyond the .05 level) as were the differences between industries for the correlation coefficients for each predictor.[18] Correlation coefficients between forecasts and realizations tended to

TABLE 8
RANK SCORES OF CORRELATIONS OF PREDICTIONS AND REALIZATIONS SUMMED OVER PREDICTORS*

	1962-65 Growth of Actual Earnings	1962-65 Growth of Normalized Earnings	1963-65 Growth of Actual Earnings	1963-65 Growth of Normalized Earnings	Total
Industry					
1)	20	23	20	28	91
2)	18	22	14	25	79
3)	9	11	24	14	58
4)	10	10	8	7	35
5)	5	7	24	26	62
6)	8	5	5	10	28
7)	14	15	20	20	69
8)	24	15	29	14	82
Kendall's W	.76	.74	.72	.65	.32

* Entries are sums of ranks over predictors for correlations of predictions with growth rates indicated in column headings.

17. Predictor C could not be included in this calculation because of a lack of observations in some industries.

18. The latter, however, was tested only on the basis of the asymptotic distribution of the correlation coefficient and the assumption that the data were distributed normally.

be highest in industries (1) electricals and electronics, (8) "miscellaneous," and (2) electric utilities; they were lowest in (6) foods and stores and (4) oils. Industry (5) drugs, showed very low correlations for the 1962 predictions and high ones for the 1963 predictions. Similar patterns emerged, though more weakly, when perceptions of past growth rates over more than one year were used as forecasts. It is interesting to note that certain industries which were "difficult to forecast" in an *ex ante* sense (see Section II. B) actually turned out to be difficult to predict, *ex post*. For example, there was high (low) agreement among predictors concerning the growth rates for the electric utilities (oils) and also high (low) correlation between predictions and realizations.

In general, we had little success in associating forecasting success with any industry or company characteristics. The differences between industries in forecasting success were only moderately related either to the average growth rates to be realized or to the variances of the realized growth rates. Two of the industries where the highest correlations were found, industries (1) and (2), had respectively the highest and the lowest average growth rates and variances. The third industry where success occurred, (8), fell in the middle range for both quantities. The rank-totals of the last column of Table 8 had a rank correlation with the rank-totals for average growth rates of .14 and of .37 with the rank-totals for the variances.

To further investigate how forecasting ability was related to company characteristics, the corporations were classified according to the quality ratings supplied by two of the predicting firms. There was a tendency for the correlations to be lowest (and negative) in the poorest-quality grouping, but they did not get systematically higher with quality, the highest correlations tending to occur in the middle classes. Similarly, classifying by high, low, or medium values of the instability index showed no pronounced differences in performance. The forecasting performances were again worst for the lowest-quality corporations and best in the middle category. When the corporations were classified by high, medium, or low price-earnings multiple, or past growth rate of earnings, or future growth rates of earnings, sales or assets, no pronounced or significant patterns emerged.

IV. An Appraisal of the Forecasts

The rather poor over-all forecasting performances of the predictors and the fact that their past perceptions of growth rates were about as reliable forecasts as their explicit predictions raises two questions: 1) Does any naive forecasting device based on historic data yield as good forecasts as the painstaking efforts of security analysts? 2) Is it the basically volatile nature of earnings that explains our results and would the predictions appear more accurate if they were taken to be forecasts of more stable measures of the growth of corporations?

To investigate the first of these questions, past growth rates calculated on the basis of arithmetic and logarithmic regressions and on the geometric means of first ratios, calculated over periods up to 14 years, were compared with

TABLE 9

CORRELATIONS OF CALCULATED PAST GROWTH RATES ON REALIZATIONS*

	Growth of Actual Earnings 1962-65	Growth of Normalized Earnings 1962-65	Growth of Actual Earnings 1963-65	Growth of Normalized Earnings 1963-65
	I. Correlations			
g_{c1}	.03	.42	.01	.26
g_{c2}	—.15	.19	—.15	.06
g_{c3}	—.13	.15	—.16	.02
g_{c4}	—.10	.09	—.11	—.02
g_{c5}	.22	.62	.18	.46
g_{c6}	.12	.51	.06	.34
g_{c7}	.01	.24	—.01	.12
g_{c8}	—.02	.37	—.03	.23
g_{c9}	—.12	.09	—.14	—.01
	II. Inequality Coefficients			
g_{1c}	.93	.79	.93	.85
g_{c2}	1.03	.95	1.01	.96
g_{c3}	.95	.88	.96	.91
g_{c4}	.88	.82	.90	.86
g_{c5}	1.27	1.22	1.11	1.08
g_{c6}	.89	.73	.90	.80
g_{c7}	.83	.75	.86	.80
g_{c8}	.98	.85	.96	.87
g_{c9}	.89	.83	.91	.86

* For definition of g's see footnote to Table 5.

the realized growth rates through 1965. A selection of these comparisons based on data ending in 1962 is found in Table 9.[19]

It is interesting to note first that the calculated growth rates tend to be more closely correlated with the growth rates of normalized earnings than with the growth rates of actual earnings. This is an even more pronounced feature of the calculated growth rates than of the data considered earlier. Second, while the correlations of the calculated growth rates with the realized growth rates tended to be lower than those found for the predictions and perceptions, and fewer of them differed significantly from zero, these differences are not pronounced. However, unlike the earlier data, the calculations seem to have almost no forecasting ability, a finding similar to that of I. M. D. Little [7] for British corporations. Among the calculated rates, those for shorter periods of time tend to be somewhat better in terms of correlation than those for longer ones, a feature highlighted by the strong showing of the growth rates calculated over only one year (g_{c5}). Third, while one would have expected that extrapolations using as the last year for the calculation the same year that is used for the first year in calculation of the realization would have a lower correlation than extrapolations where the data ended a year earlier, in

19. The figures there are typical both of what was found when other periods were used and of the comparisons of calculations ending in 1961 and 1963 with the perceived growth after 1962 and 1963 respectively.

fact the reverse tendency manifested itself. Finally, among the possible ways of calculating growth rates, those based on the geometric means of the first ratios surpassed those based on regressions.

The superiority of the past perceived growth rates over the calculated ones should not be taken too seriously, however, for it was largely due to the fact that negative perceived growth rates were not reported by our participants. The survey respondents only indicated that the rates were negative. As a result, companies for which this was true had to be dropped from the sample when correlations of realized with perceived past growth rates were made. When we dropped the companies whose past calculated growth rates were negative (in order to put the calculated and perceived growth rates on a similar basis), the correlation coefficients of the calculated with the realized growth rates were raised. For example, with this change the first row of Table 9 would read

.30 .53 .17 .42

which compares favorably with the data in Table 7. Similar improvements occurred using the other types of calculated growth rates.

The possibilities of obtaining useful forecasts from simple extrapolation were also examined by calculating growth rates over the four preceding years[20] for (1) earnings plus depreciation, (2) earnings before taxes, (3) sales, (4) assets, and (5) share prices. The correlations of these growth rates calculated to the end of 1962, both with 1962-1965 and 1963-1965 earnings growth and the growth rates of the same variables, are shown in the first five rows of Table 10. It will be noticed that both the levels and the variation of these correlation coefficients are quite similar to those found for the predictions and perceptions of past growth and the equivalently calculated past growth rates of earnings. There was also no marked tendency for the extrapolations to do better at predicting their own growth rates than the growth rates of normalized earnings, but they tended to be better at predicting their own rates than the growth of actual earnings.

The last two rows of Table 10 show the correlations of the price-earnings ratio and the price-to-normalized-earnings ratio with the actual future growth of earnings. As mentioned earlier, these ratios have implicit in them a forecast of the rate of growth anticipated by the market. We find that, in terms of correlation, the market-determined earnings multiples perform no differently from the other predictors we have considered.

A similar picture emerged when the predictions and perceptions of growth rates of earnings were used to predict the growth that would occur in these same variables through the end of 1965. With the exception of the growth of price, the performance of the predictions and perceptions were about the same in terms of correlation as those shown when they were used to forecast the growth of normalized earnings. The inequality coefficients were, if anything, slightly lower. For price growth, however, these forecasts had virtually

20. Other periods and methods of calculating growth rates were also used. The ones presented tended to be very slightly better than the others and are comparable to the most successful of the longer-term earnings extrapolations.

TABLE 10
EXTRAPOLATIONS FROM OTHER SERIES AS PREDICTORS OF EARNINGS AND OWN GROWTH RATES*
(CORRELATION COEFFICIENTS)

	Growth of Actual Earnings 1962-65	Growth of Normalized Earnings 1962-65	Growth of Actual Earnings 1963-65	Growth of Normalized Earnings 1963-65	Growth Rate of Corresponding Variable 1962-65	Growth Rate of Corresponding Variable 1963-65
g_{e1}	.11	.39	.05	.27	.28	.20
g_{e2}	.29	.21	.42	.30	.24	.38
g_{e3}	.23	.37	.15	.29	.39	.31
g_{e4}	.29	.46	.47	.60	.63	.27
g_{e5}	.04	.34	—.03	.20	—.06	.05
P/E	.21	.25	.13	.18	—	—
P/NE	.14	.35	.08	.21	—	—

* g_{e1} is growth of earnings plus depreciation
g_{e2} is growth of earnings plus taxes
g_{e3} is growth of sales
g_{e4} is growth of assets
g_{e5} is growth of price of stock
P/E is price-earnings ratio at end of 1962
P/NE is price-normalized earnings ratio at end of 1962
The period used for the calculations of the growth rates was 1958-62 and the rates were calculated as

$$g = \sqrt[4]{V_{62} / V_{58}} \text{ where } V_{62} \text{ and } V_{58} \text{ are the values of the variables.}$$

no merit, with even poorer performance than they had for the growth of actual earnings.

V. CONCLUSION

In this paper, we have examined the characteristics of a small sample of security analysts' predictions of the long-run earnings growth of corporations. The extent of agreement among the different predictors was considered and their forecasting abilities assessed. Evidence has recently accumulated [7] that earnings growth in past periods is not a useful predictor of future earnings growth. The remarkable conclusion of the present study is that the careful estimates of the security analysts participating in our survey, the bases of which are not limited to public information, perform little better than these past growth rates. Moreover, the market price-earnings ratios themselves were not better than either the analysts' forecasts or the past growth rates in forecasting future earnings growth.

We must be cautious, however, in overgeneralizing these results. We did not have data to investigate directly whether the performance of the predictions of growth in the period considered were atypical of the usual forecasting abilities of such forecasts. The question is important, however, since it can be argued that the peculiarities of the expansion that occurred after the date of the forecasts made the period especially difficult to forecast. Moreover, our work is hampered by the fact that only a few firms were able to participate in our survey. It may also be that shorter-term earnings predictions are considerably more successful relative to naive forecasting methods. Fortunately, we are presently collecting additional data that will help shed light on these conjectures and permit a study of the generation of earnings forecasts and their usefulness in security evaluation.

REFERENCES

1. John W. Buckman. *The Reliability of Earnings Forecasts,* prepared for the Trust Investment Committee of the Continental Illinois Bank (Chicago: September 13, 1963), mimeographed.
2. John M. Culbertson. "The Term Structure of Interest Rates," *Quarterly Journal of Economics,* November, 1957, 485-517.
3. Benjamin Graham, David L. Dodd and Sidney Cottle. *Security Analysis, Principles and Technique,* 4th ed. (New York: McGraw-Hill, 1962).
4. Myron J. Gordon. *The Investment, Financing, and Valuation of the Corporation* (Homewood: Richard D. Irwin, 1962).
5. C. C. Holt. "The Influence of Growth Duration on Share Prices," *Journal of Finance,* September, 1962, 465-75.
6. Reuben Kessel. *The Cyclical Behavior of the Term Structure of Interest Rates* (New York: National Bureau of Economic Research, 1962).
7. I. M. D. Little. "Higgledy Piggledy Growth," *Oxford Institute of Statistics Bulletin,* November, 1962, 387-412.
8. Burton G. Malkiel. "Equity Yields, Growth, and the Structure of Share Prices," *American Economic Review,* December, 1963, 1004-31.
9. David Meiselman. *The Term Structure of Interest Rates* (Englewood Cliffs: Prentice-Hall, 1962).
10. Franco Modigliani and Richard Sutch. "Innovations in Interest Rate Policy," *American Economic Review: Papers and Proceedings,* May 1966, 178-197.
11. Henry Scheffé. *The Analysis of Variance* (New York: John Wiley & Sons, 1959).
12. *Short-Term Economic Forecasting.* Studies in Income and Wealth (Princeton: Princeton University Press for NBER, 1955).
13. *The Quality and Economic Significance of Anticipations Data* (Princeton: Princeton University Press for NBER, 1960).
14. Henri Theil. *Applied Economic Forecasting* (Chicago: Rand McNally, 1966).
15. J. B. Williams. *The Theory of Investment Value* (Cambridge: Harvard University Press, 1938).
16. Victor Zarnowitz. *An Appraisal of Short-Term Economic Forecasts* (New York: National Bureau of Economic Research, 1967).

DISCOUNTED CASHFLOW MODELS: WHAT THEY ARE AND HOW TO CHOOSE THE RIGHT ONE..

THE FUNDAMENTAL CHOICES FOR DCF VALUATION

- **Cashflows to Discount**
    - Dividends
    - Free Cash Flows to Equity
    - Free Cash Flows to Firm
- **Expected Growth**
    - Stable Growth
    - Two Stages of Growth: High Growth -> Stable Growth
    - Three Stages of Growth: High Growth -> Transition Period -> Stable Growth
- **Discount Rate**
    - Cost of Equity
    - Cost of Capital
- **Base Year Numbers**
    - Current Earnings / Cash Flows
    - Normalized Earnings / Cash Flows

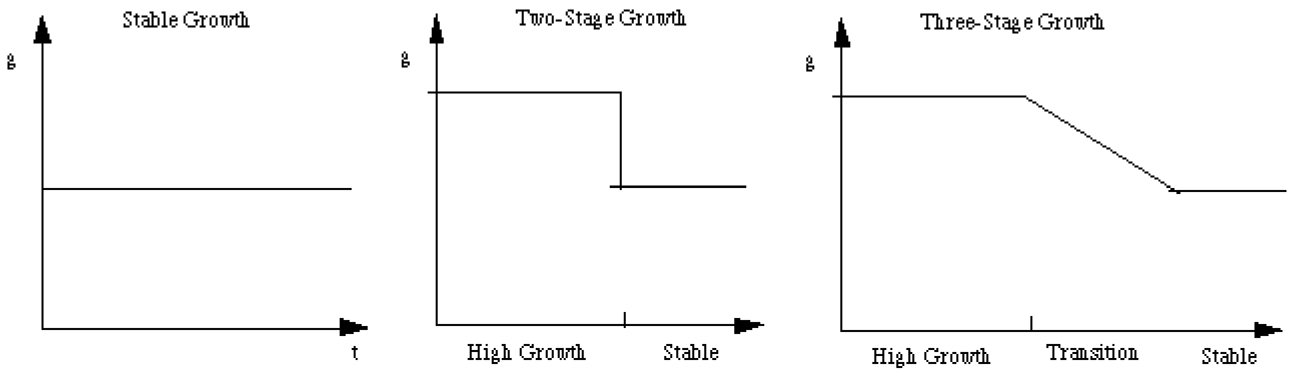
WHICH CASH FLOW TO DISCOUNT...

- **The Discount Rate should be consistent with the cash flow being discounted**
    - Cash Flow to Equity -> Cost of Equity
    - Cash Flow to Firm -> Cost of Capital
- **Should you discount Cash Flow to Equity or Cash Flow to Firm?**
    - *Use Equity Valuation*
        - (a) for firms which have stable leverage, whether high or not, and
        - (b) if equity (stock) is being valued
    - *Use Firm Valuation*
        - (a) for firms which have high leverage, and expect to lower the leverage over time, because
            - debt payments do not have to be factored in
            - the discount rate (cost of capital) does not change dramatically over time.
        - (b) for firms for which you have partial information on leverage (eg: interest expenses are missing..)
        - (c) in all other cases, where you are more interested in valuing the firm than the equity. (Value Consulting?)
- **Given that you discount cash flow to equity, should you discount dividends or Free Cash Flow to Equity?**
    - *Use the Dividend Discount Model*
        - (a) For firms which pay dividends (and repurchase stock) which are close to the Free Cash Flow to Equity (over a extended period)

- (b)For firms where FCFE are difficult to estimate (Example: Banks and Financial Service companies)
- *Use the FCFE Model*
  - (a) For firms which pay dividends which are significantly higher or lower than the Free Cash Flow to Equity. (What is significant? ... As a rule of thumb, if dividends are less than 75% of FCFE or dividends are greater than FCFE)
  - (b) For firms where dividends are not available (Example: Private Companies, IPOs)

WHAT IS THE RIGHT GROWTH PATTERN...

- **The Choices**



THE PRESENT VALUE FORMULAE

- For Stable Firm: $V_0 = \frac{CF_1}{r - g_n}$
- For two stage growth: $V_0 = \frac{CF_0 * (1+g) * \left(1 - \frac{(1+g)^n}{(1+r)^n}\right)}{r - g} + \frac{CF_{n+1}}{(r - g_n)(1+r)^n}$
- For three stage growth: $V_0 = \sum_{t=1}^{t=n1} \frac{CF_0 * (1+g_a)^t}{(1+r)^t} + \sum_{t=n1+1}^{t=n2} \frac{CF_t}{(1+r)^t} + \frac{CF_{n2+1}}{(r - g_n)(1+r)^n}$

Definitions of Terms

V_0= Value of Equity (if cash flows to equity are discounted) or Firm (if cash flows to firm are discounted)

CF_t = Cash Flow in period t; *Dividends* or *FCFE* if valuing equity or *FCFF* if valuing firm.

r = Cost of Equity (if discounting Dividends or FCFE) or Cost of Capital (if discounting FCFF)

g = Expected growth rate in Cash Flow being discounted

g_a= Expected growth in Cash Flow being discounted in first stage of three stage growth model

g_n= Expected growth in Cash Flow being discounted in stable period

n = Length of the high growth period in two-stage model

n1 = Length of the first high growth period in three-stage model

n2 - n1 = Transition period in three-stage model

WHICH MODEL SHOULD I USE?

- Use the growth model only if cash flows are positive
- *Use the stable growth model, if*
    - the firm is growing at a rate which is below or close (within 1-2%) to the growth rate of the economy
- *Use the two-stage growth model if*
    - the firm is growing at a moderate rate (... within 8% of the stable growth rate)
- *Use the three-stage growth model if*
    - the firm is growing at a high rate (... more than 8% higher than the stable growth rate)

SUMMARIZING THE MODEL CHOICES

	Dividend Discount Model	**FCFE Model**	**FCFF Model**
Stable Growth Model	• Growth rate in firmís earnings is stable. (g of $\text{firm}_{economy}$+1%) • Dividends are close to FCFE (or) FCFE is difficult to compute. • Leverage is stable	• Growth rate in firmís earnings is stable. ($g_{firmeconomy}$+1%) • Dividends are very different from FCFE (or) Dividends not available (Private firm) • Leverage is stable	• Growth rate in firmís earnings is stable. ($g_{firmeconomy}$+1%) • Leverage is high and expected to change over time (unstable).
Two-Stage Model	• Growth rate in firmís earnings is moderate. • Dividends are close to FCFE (or) FCFE is difficult to compute. • Leverage is stable	• Growth rate in firmís earnings is moderate. • Dividends are very different from FCFE (or) Dividends not available (Private firm) • Leverage is stable	• Growth rate in firmís earnings is moderate. • Leverage is high and expected to change over time (unstable).
Three-Stage Model	• Growth rate in firmís earnings is high.	• Growth rate in firmís earnings is high.	• Growth rate in firmís earnings is high.

	• Dividends are close to FCFE (or) FCFE is difficult to compute. • Leverage is stable	• Dividends are very different from FCFE (or) Dividends not available (Private firm) • Leverage is stable	• Leverage is high and expected to change over time (unstable).

GROWTH AND FIRM CHARACTERISTICS

	Dividend Discount Model	*FCFE Discount Model*	*FCFF Discount Model*
High growth firms generally	• Pay no or low dividends • Earn high returns on projects (ROA) • Have low leverage (D/E) • Have high risk (high betas)	• Have high capital expenditures relative to depreciation. • Earn high returns on projects • Have low leverage • Have high risk	• Have high capital expenditures relative to depreciation. • Earn high returns on projects • Have low leverage • Have high risk
Stable growth firms generally	• Pay large dividends relative to earnings (high payout) • Earn moderate returns on projects (ROA is closer to market or industry average) • Have higher leverage • Have average risk (betas are closer to one.)	• narrow the difference between cap ex and depreciation. (Sometimes they offset each other) • Earn moderate returns on projects (ROA is closer to market or industry average) • Have higher leverage • Have average risk (betas are closer to one.)	• narrow the difference between cap ex and depreciation. (Sometimes they offset each other) • Earn moderate returns on projects (ROA is closer to market or industry average) • Have higher leverage • Have average risk (betas are closer to one.)

SHOULD I NORMALIZE EARNINGS?

- *Why normalize earnings?*
  - The firm may have had an exceptionally good or bad year (which is not expected to be sustainable)
  - The firm is in financial trouble, and its current earnings are below normal or negative.
- *What types of firms can I normalize earnings for?*
  - The firms used to be financially healthy, and the current problems are viewed as temporary.

- The firm is a small upstart firm in an established industry, where the average firm is profitable.

HOW DO I NORMALIZE EARNINGS?

- If the firm is in trouble because of a recession, and its size has not changed significantly over time,
- *Use average earnings over an extended time period for the firm*

Normalized Earnings = Average Earnings from past period (5 or 10 years)

- If the firm is in trouble because of a recession, and its size has changed significantly over time,
- *Use average Return on Equity over an extended time period for the firm*

Normalized Earnings = Current Book Value of Equity * Average Return on Equity (Firm)

- If the firm is in trouble because of firm-specific factors, and the rest of the industry is healthy,
- *Use average Return on Equity for comparable firms*

Normalized Earnings = Current Book Value of Equity * Average Return on Equity (Comparables)

Valuation Approaches and Metrics: A Survey of the Theory and Evidence

Aswath Damodaran
Stern School of Business

November 2006

Valuation Approaches and Metrics: A Survey Article

Valuation lies at the heart of much of what we do in finance, whether it is the study of market efficiency and questions about corporate governance or the comparison of different investment decision rules in capital budgeting. In this paper, we consider the theory and evidence on valuation approaches. We begin by surveying the literature on discounted cash flow valuation models, ranging from the first mentions of the dividend discount model to value stocks to the use of excess return models in more recent years. In the second part of the paper, we examine relative valuation models and, in particular, the use of multiples and comparables in valuation and evaluate whether relative valuation models yield more or less precise estimates of value than discounted cash flow models. In the final part of the paper, we set the stage for further research in valuation by noting the estimation challenges we face as companies globalize and become exposed to risk in multiple countries.

Valuation can be considered the heart of finance. In corporate finance, we consider how best to increase firm value by changing its investment, financing and dividend decisions. In portfolio management, we expend resources trying to find firms that trade at less than their true value and then hope to generate profits as prices converge on value. In studying whether markets are efficient, we analyze whether market prices deviate from value, and if so, how quickly they revert back. Understanding what determines the value of a firm and how to estimate that value seems to be a prerequisite for making sensible decisions.

Given the centrality of its role, you would think that the question of how best to value a business, private or public, would have been well researched. As we will show in this paper, the research into valuation models and metrics in finance is surprisingly spotty, with some aspects of valuation, such as risk assessment, being deeply analyzed and others, such as how best to estimate cash flows and reconciling different versions of models, not receiving the attention that they deserve.

Overview of Valuation

Analysts use a wide spectrum of models, ranging from the simple to the sophisticated. These models often make very different assumptions about the fundamentals that determine value, but they do share some common characteristics and can be classified in broader terms. There are several advantages to such a classification -- it makes it is easier to understand where individual models fit in to the big picture, why they provide different results and when they have fundamental errors in logic.

In general terms, there are four approaches to valuation. The first, discounted cashflow valuation, relates the value of an asset to the present value of expected future cashflows on that asset. The second, liquidation and accounting valuation, is built around valuing the existing assets of a firm, with accounting estimates of value or book value often used as a starting point. The third, relative valuation, estimates the value of an asset by looking at the pricing of 'comparable' assets relative to a common variable like earnings, cashflows, book value or sales. The final approach, contingent claim valuation, uses option pricing models to measure the value of assets that share option characteristics. This is what generally falls under the rubric of real options.

Since almost everything in finance can be categorized as a subset of valuation and we run the risk of ranging far from our mission, we will keep a narrow focus in this paper. In particular, we will steer away any work done on real options, since it merits its own survey article. In addition, we will keep our focus on papers that have examined the theory and practice of valuation of companies and stocks, rather than on questions of assessing risk and estimating discount rates that have consumed a great deal of attention in the literature.

Discounted Cash flow Valuation

In discounted cashflows valuation, the value of an asset is the present value of the expected cashflows on the asset, discounted back at a rate that reflects the riskiness of these cashflows. This approach gets the most play in academia and comes with the best theoretical credentials. In this section, we will look at the foundations of the approach and some of the preliminary details on how we estimate its inputs.

Essence of Discounted Cashflow Valuation

We buy most assets because we expect them to generate cash flows for us in the future. In discounted cash flow valuation, we begin with a simple proposition. The value of an asset is not what someone perceives it to be worth but it is a function of the expected cash flows on that asset. Put simply, assets with high and predictable cash flows should have higher values than assets with low and volatile cash flows.

The notion that the value of an asset is the present value of the cash flows that you expect to generate by holding it is neither new nor revolutionary. While knowledge of compound interest goes back thousands of years[1], the concrete analysts of present value was stymied for centuries by religious bans on charging interest on loans, which was treated as usury. In a survey article on the use of discounted cash flow in history, Parker (1968) notes that the earliest interest rate tables date back to 1340 and were prepared by Francesco Balducci Pegolotti, a Florentine merchant and politician, as part of his manuscript titled *Practica della Mercatura*, which was not officially published until

[1] Neugebauer, O.E.H., 1951, The Exact Sciences in Antiquity, Copenhagen, Ejnar Munksgaard. He notes that interest tables existed in Mesopotamia.

1766.[2] The development of insurance and actuarial sciences in the next few centuries provided an impetus for a more thorough study of present value. Simon Stevin, a Flemish mathematician, wrote one of the first textbooks on financial mathematics in 1582 and laid out the basis for the present value rule in an appendix.[3]

The extension of present value from insurance and lending to corporate finance and valuation can be traced to both commercial and intellectual impulses. On the commercial side, the growth of railroads in the United States in the second half of the nineteenth century created a demand for new tools to analyze long-term investments with significant cash outflows in the earlier years being offset by positive cash flows in the later years. A civil engineer, A.M. Wellington, noted not only the importance of the time value of money but argued that the present value of future cash flows should be compared to the cost of up-front investment.[4] He was followed by Walter O. Pennell, an engineer of Southwestern Bell, who developed present value equations for annuities, to examine whether to install new machinery or retain old equipment.[5]

The intellectual basis for discounted cash flow valuation were laid by Alfred Marshall and Bohm-Bawerk, who discussed the concept of present value in their works in the early part of the twentieth century.[6] In fact, Bohm-Bawerk (1903) provided an explicit example of present value calculations using the example of a house purchase with twenty annual installment payments. However, the principles of modern valuation were developed by Irving Fisher in two books that he published – *The Rate of Interest* in 1907 and *The Theory of Interest* in 1930.[7] In these books, he suggested four alternative approaches for analyzing investments, that he claimed would yield the same results. He argued that when confronted with multiple investments, you should pick the investment (a) that has the highest present value at the market interest rate; (b) where the present

---

[2] Parker, R.H., 1968, Discounted Cash Flow in Historical Perspective, Journal of Accounting Research, v6, 58-71.
[3] Stevin, S., 1582, Tables of Interest.
[4] Wellington, A.M., 1887, The Economic Theory of the Location of Railways, Wiley, New York.
[5] Pennell, W.O., 1914, Present Worth Calculations in Engineering Studies, Journal of the Association of Engineering Societies.
[6] Marshall, A., 1907, Principles of Economics, Macmillan, London; Bohm-Bawerk, A. V., 1903, Recent Literature on Interest, Macmillan.
[7] Fisher, I., 1907, The Rate of Interest, Macmillan, New York; Fisher, I., 1930, The Theory of Interest, Macmillan, New York.

value of the benefits exceeded the present value of the costs the most; (c) with the "rate of return on sacrifice" that most exceeds the market interest rate or (d) that, when compared to the next most costly investment, yields a rate of return over cost that exceeds the market interest rate. Note that the first two approaches represent the net present value rule, the third is a variant of the IRR approach and the last is the marginal rate of return approach. While Fisher did not delve too deeply into the notion of the rate of return, other economists did. Looking at a single investment, Boulding (1935) derived the internal rate of return for an investment from its expected cash flows and an initial investment.[8] Keynes (1936) argued that the "marginal efficiency of capital" could be computed as the discount rate that makes the present value of the returns on an asset equal to its current price and that it was equivalent to Fisher's rate of return on an investment.[9] Samuelson (1937) examined the differences between the internal rate of return and net present value approaches and argued that rational investors should maximize the latter and not the former.[10] In the last 50 years, we have seen discounted cash flow models extend their reach into security and business valuation, and the growth has been aided and abetted by developments in portfolio theory.

Using discounted cash flow models is in some sense an act of faith. We believe that every asset has an intrinsic value and we try to estimate that intrinsic value by looking at an asset's fundamentals. What is intrinsic value? Consider it the value that would be attached to an asset by an all-knowing analyst with access to all information available right now and a perfect valuation model. No such analyst exists, of course, but we all aspire to be as close as we can to this perfect analyst. The problem lies in the fact that none of us ever gets to see what the true intrinsic value of an asset is and we therefore have no way of knowing whether our discounted cash flow valuations are close to the mark or not.

There are four variants of discounted cash flow models in practice, and theorists have long argued about the advantages and disadvantages of each. In the first, we discount expected cash flows on an asset (or a business) at a risk-adjusted discount rate to

---

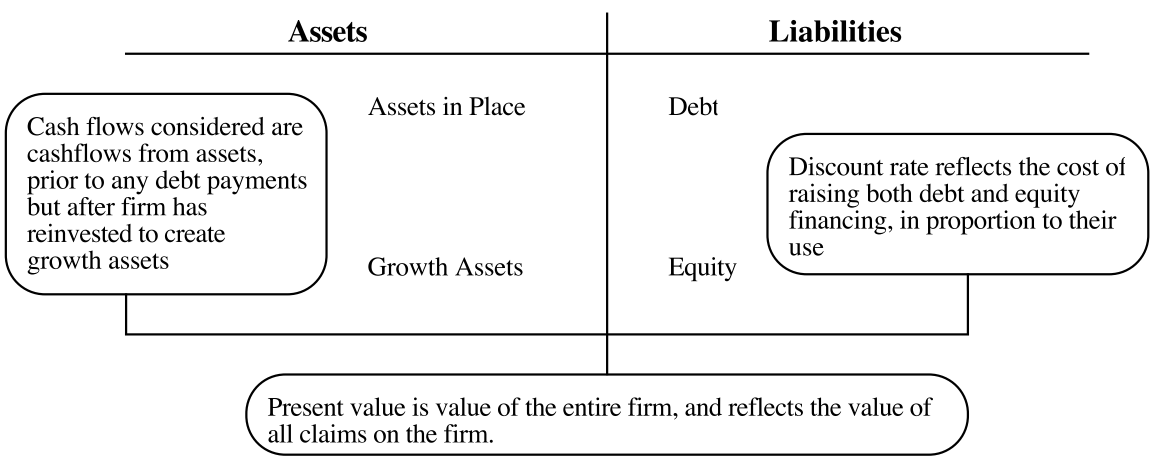
[8] Boulding, K.E., 1935, The Theory of a Single Investment, Quarterly Journal of Economics, v49, 479-494.
[9] Keynes, J.M., 1936, The General Theory of Employment, Macmillan, London.

arrive at the value of the asset. In the second, we adjust the expected cash flows for risk to arrive at what are termed risk-adjusted or certainty equivalent cash flows which we discount at the riskfree rate to estimate the value of a risky asset. In the third, we value a business first, without the effects of debt, and then consider the marginal effects on value, positive and negative, of borrowing money. This approach is termed the adjusted present value approach. Finally, we can value a business as a function of the excess returns we expect it to generate on its investments. As we will show in the following section, there are common assumptions that bind these approaches together, but there are variants in assumptions in practice that result in different values.

Discount Rate Adjustment Models

Of the approaches for adjusting for risk in discounted cash flow valuation, the most common one is the risk adjusted discount rate approach, where we use higher discount rates to discount expected cash flows when valuing riskier assets, and lower discount rates when valuing safer assets. There are two ways in which we can approach discounted cash flow valuation. The first is to value the entire business, with both assets-in-place and growth assets; this is often termed firm or enterprise valuation.

Firm Valuation

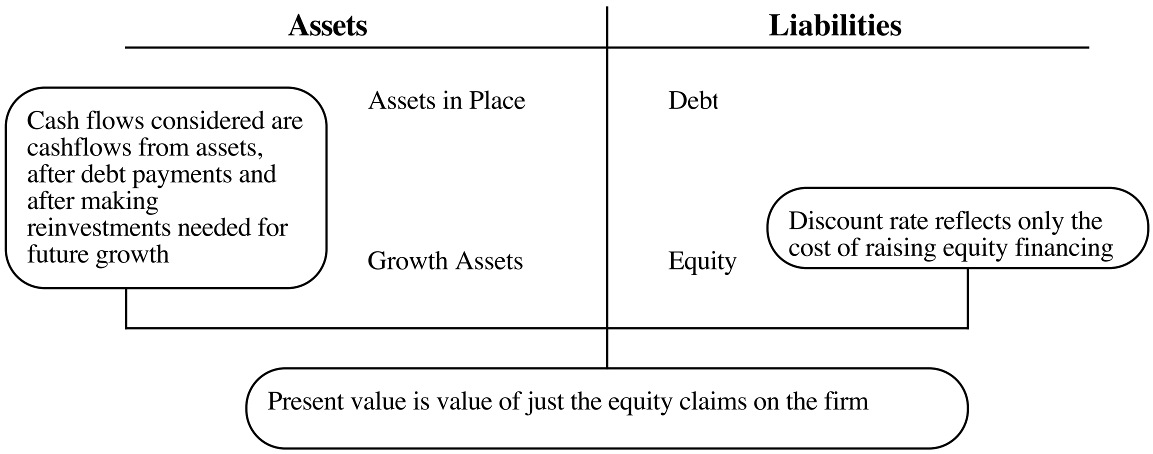


The cash flows before debt payments and after reinvestment needs are termed free cash flows to the firm, and the discount rate that reflects the composite cost of financing from all sources of capital is the cost of capital.

[10] Samuelson, P., 1937, Some Aspects of the Pure Theory of Capital, Quarterly Journal of Economics, v51,

The second way is to just value the equity stake in the business, and this is called equity valuation.

Equity Valuation



The cash flows after debt payments and reinvestment needs are called free cash flows to equity, and the discount rate that reflects just the cost of equity financing is the cost of equity.

Note also that we can always get from the former (firm value) to the latter (equity value) by netting out the value of all non-equity claims from firm value. Done right, the value of equity should be the same whether it is valued directly (by discounting cash flows to equity a the cost of equity) or indirectly (by valuing the firm and subtracting out the value of all non-equity claims).

1. Equity DCF Models

In equity valuation models, we focus our attention of the equity investors in a business and value their stake by discounting the expected cash flows to these investors at a rate of return that is appropriate for the equity risk in the company. The first set of models examined take a strict view of equity cash flows and consider only dividends to be cashflows to equity. These dividend discount models represent the oldest variant of discounted cashflow models. We then consider broader definitions of cash flows to equity, by first including stock buybacks in cash flows to equity and by then expanding out analysis to cover potential dividends or free cash flows to equity.

469-496.

a. Dividend Discount Model

The oldest discounted cash flow models in practice tend to be dividend discount models. While many analysts have turned away from dividend discount models on the premise that they yield estimates of value that are far too conservative, many of the fundamental principles that come through with dividend discount models apply when we look at other discounted cash flow models.

Basis for Dividend Discount Models

When investors buy stock in publicly traded companies, they generally expect to get two types of cashflows - dividends during the holding period and an expected price at the end of the holding period. Since this expected price is itself determined by future dividends, the value of a stock is the present value of dividends through infinity.

$$\text{Value per share of stock} = \sum_{t=1}^{t=\infty} \frac{E(DPS_t)}{(1+k_e)^t}$$

where,

$E(DPS_t)$ = Expected dividends per share in period t

k_e = Cost of equity

The rationale for the model lies in the present value rule - the value of any asset is the present value of expected future cash flows discounted at a rate appropriate to the riskiness of the cash flows. There are two basic inputs to the model - expected dividends and the cost on equity. To obtain the expected dividends, we make assumptions about expected future growth rates in earnings and payout ratios. The required rate of return on a stock is determined by its riskiness, measured differently in different models - the market beta in the CAPM, and the factor betas in the arbitrage and multi-factor models. The model is flexible enough to allow for time-varying discount rates, where the time variation is caused by expected changes in interest rates or risk across time.

While explicit mention of dividend discount models did not show up in research until the last few decades, investors and analysts have long linked equity values to dividends. Perhaps the first book to explicitly connect the present value concept with dividends was *The Theory of Investment Value* by John Burr Williams (1938), where he stated the following:

> *"A stock is worth the present value of all the dividends ever to be paid upon it, no more, no less... Present earnings, outlook, financial condition, and capitalization should bear upon the price of a stock only as they assist buyers and sellers in estimating future dividends."*

Williams also laid the basis for forecasting pro forma financial statements and drew a distinction between valuing mature and growth companies.[11] While much of his work has become shrouded with myth, Ben Graham (1934) also made the connection between dividends and stock values, but not through a discounted valuation model. He chose to develop instead a series of screening measures that including low PE, high dividend yields, reasonable growth and low risk that highlighted stocks that would be under valued using a dividend discount model.[12]

Variations on the Dividend Discount Model

Since projections of dollar dividends cannot be made in perpetuity and publicly traded firms, at least in theory, can last forever, several versions of the dividend discount model have been developed based upon different assumptions about future growth. We will begin with the simplest – a model designed to value stock in a stable-growth firm that pays out what it can afford to in dividends. The value of the stock can then be written as a function of its expected dividends in the next time period, the cost of equity and the expected growth rate in dividends.

$$\text{Value of Stock} = \frac{\text{Expected Dividends next period}}{\text{(Cost of equity - Expected growth rate in perpetuity}}$$

Though this model has made the transition into every valuation textbook, its origins are relatively recent and can be traced to early work by David Durand and Myron Gordon. It was Durand (1957) who noted that valuing a stock with dividends growing at a constant rate forever was a variation of The Petersburg Paradox, a seminal problem in utility theory for which a solution was provided by Bernoulli in the eighteenth century.[13] It was Gordon, though, who popularized the model in subsequent articles and a book, thus

---

[11] Williams, J.B., 1938, Theory of Investment Value, Fraser Publishing company (reprint).

[12] Dodd, D. and B. Graham, 1934, Security Analysis, McGraw Hill, New York; Graham, B., 1949, The Intelligent Investor, Collins (reprint).

[13] Durand, D., 1957, Growth Stocks and the St. Petersburg Paradox, Journal of Finance, v12, 348-363.

giving it the title of the Gordon growth model.[14] While the Gordon growth model is a simple approach to valuing equity, its use is limited to firms that are growing at stable rates that can be sustained forever. There are two insights worth keeping in mind when estimating a 'stable' growth rate. First, since the growth rate in the firm's dividends is expected to last forever, it cannot exceed the growth rate of the economy in which the firm operates. The second is that the firm's other measures of performance (including earnings) can also be expected to grow at the same rate as dividends. To see why, consider the consequences in the long term of a firm whose earnings grow 3% a year forever, while its dividends grow at 4%. Over time, the dividends will exceed earnings. On the other hand, if a firm's earnings grow at a faster rate than dividends in the long term, the payout ratio, in the long term, will converge towards zero, which is also not a steady state. Thus, though the model's requirement is for the expected growth rate in dividends, analysts should be able to substitute in the expected growth rate in earnings and get precisely the same result, if the firm is truly in steady state.

In response to the demand for more flexibility when faced with higher growth companies, a number of variations on the dividend discount model were developed over time in practice. The simplest extension is a two-stage growth model allows for an initial phase where the growth rate is not a stable growth rate and a subsequent steady state where the growth rate is stable and is expected to remain so for the long term. While, in most cases, the growth rate during the initial phase will be higher than the stable growth rate, the model can be adapted to value companies that are expected to post low or even negative growth rates for a few years and then revert back to stable growth. The value of equity can be written as the present value of expected dividends during the non-stable growth phase and the present value of the price at the end of the high growth phase, usually computed using the Gordon growth model:

$$P_0 = \sum_{t=1}^{t=n} \frac{E(DPS_t)}{(1+\text{ Cost of Equity})^t} + \frac{P_n}{(1+\text{ Cost of Equity})^n} \text{ where } P_n = \frac{E(DPS_{n+1})}{(\text{Cost of Equity - g})}$$

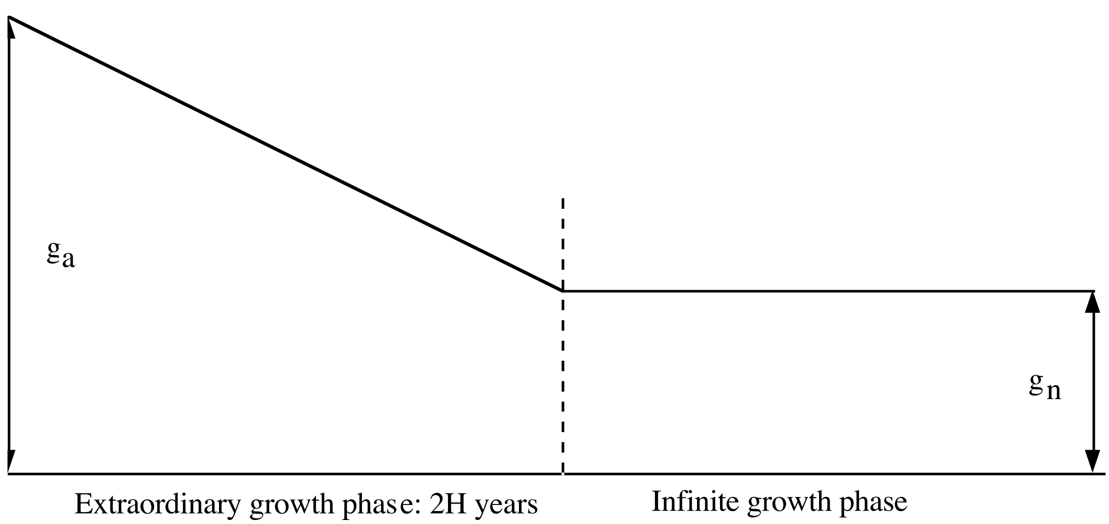
where $E(DPS_t)$ is the expected dividends per share in period t and g is the stable growth rate after n years. More complicated variants of this model allow for more than two

[14] Gordon, M.J., 1962, The Investment, Financing and Valuation of the Corporation, Homewood, Illinois:

stages of growth, with a concurrent increase in the number of inputs that have to be estimated to value a company, but no real change in the underlying principle that the value of a stock is the present value of the expected dividends.[15]

To allow for computational simplicity with higher growth models, some researchers added constraints on other aspects of firm behavior including risk and dividend payout to derive "simpler" high growth models. For instance, the H model is a two-stage model for growth, but unlike the classical two-stage model, the growth rate in the initial growth phase is not constant but declines linearly over time to reach the stable growth rate in steady state. This model was presented in Fuller and Hsia (1984) and is based upon the assumption that the earnings growth rate starts at a high initial rate (g_a) and declines linearly over the extraordinary growth period (which is assumed to last 2H periods) to a stable growth rate (g_n).[16] It also assumes that the dividend payout and cost of equity are constant over time and are not affected by the shifting growth rates. Figure 1 graphs the expected growth over time in the H Model.

Figure 1: Expected Growth in the H Model



Richard D. Irwin, Inc.

[15] The development of multi-stage dividend discount models can be attributed more to practitioners than academic researchers. For instance, Sanford Bernstein, an investment firm founded in 1967, has used a proprietary two-stage dividend discount model to analyze stocks for decades. An extensive categorization of multi-stage models is provided in Damodaran, A., 1994, Damodaran on Valuation, John Wiley, New York.

[16] Fuller, R.J. and C. Hsia, 1984, A Simplified Common Stock Valuation Model, Financial Analysts Journal, v40, 49-56.

The value of expected dividends in the H Model can be written as:

$$P_0 = \frac{DPS_0 * (1+g_n)}{(r - g_n)} + \frac{DPS_0 * H * (g_a - g_n)}{(r - g_n)}$$

where DPS_0 is the current dividend per share and growth is expected to decline linearly over the next 2H years to a stable growth rate of g_n. This model avoids the problems associated with the growth rate dropping precipitously from the high growth to the stable growth phase, but it does so at a cost. First, the decline in the growth rate is expected to follow the strict structure laid out in the model -- it drops in linear increments each year based upon the initial growth rate, the stable growth rate and the length of the extraordinary growth period. While small deviations from this assumption do not affect the value significantly, large deviations can cause problems. Second, the assumption that the payout ratio is constant through both phases of growth exposes the analyst to an inconsistency -- as growth rates decline the payout ratio usually increases. The allowance for a gradual decrease in growth rates over time may make this a useful model for firms which are growing rapidly right now, but where the growth is expected to decline gradually over time as the firms get larger and the differential advantage they have over their competitors declines. The assumption that the payout ratio is constant, however, makes this an inappropriate model to use for any firm that has low or no dividends currently. Thus, the model, by requiring a combination of high growth and high payout, may be quite limited in its applicability [17].

Applicability of the Dividend Discount Model

While many analysts have abandoned the dividend discount model, arguing that its focus on dividends is too narrow, the model does have its proponents. The dividend discount model's primary attraction is its simplicity and its intuitive logic. After all, dividends represent the only cash flow from the firm that is tangible to investors. Estimates of free cash flows to equity and the firm remain estimates and conservative investors can reasonably argue that they cannot lay claim on these cash flows. The second advantage of using the dividend discount model is that we need fewer

---

[17] Proponents of the model would argue that using a steady state payout ratio for firms that pay little or no dividends is likely to cause only small errors in the valuation.

assumptions to get to forecasted dividends than to forecasted free cashflows. To get to the latter, we have to make assumptions about capital expenditures, depreciation and working capital. To get to the former, we can begin with dividends paid last year and estimate a growth rate in these dividends. Finally, it can be argued that managers set their dividends at levels that they can sustain even with volatile earnings. Unlike cash flows that ebb and flow with a company's earnings and reinvestments, dividends remain stable for most firms. Thus, valuations based upon dividends will be less volatile over time than cash flow based valuations.

The dividend discount model's strict adherence to dividends as cash flows does expose it to a serious problem. Many firms choose to hold back cash that they can pay out to stockholders. As a consequence, the free cash flows to equity at these firms exceed dividends and large cash balances build up. While stockholders may not have a direct claim on the cash balances, they do own a share of these cash balances and their equity values should reflect them. In the dividend discount model, we essentially abandon equity claims on cash balances and under value companies with large and increasing cash balances. At the other end of the spectrum, there are also firms that pay far more in dividends than they have available in cash flows, often funding the difference with new debt or equity issues. With these firms, using the dividend discount model can generate value estimates that are too optimistic because we are assuming that firms can continue to draw on external funding to meet the dividend deficits in perpetuity.

Notwithstanding its limitations, the dividend discount model can be useful in three scenarios.

- It establishes a <u>baseline or floor value</u> for firms that have cash flows to equity that exceed dividends. For these firms, the dividend discount model will yield a conservative estimate of value, on the assumption that the cash not paid out by managers will be wasted n poor investments or acquisitions.
- It yields realistic estimates of value per share for <u>firms that do pay out their free cash flow to equity as dividends</u>, at least on average over time. There are firms, especially in mature businesses, with stable earnings, that try to calibrate their dividends to available cashflows. At least until very recently, regulated utility companies in the United States, such as phone and power, were good examples of such firms.

- In sectors where cash flow estimation is difficult or impossible, dividends are the only cash flows that can be estimated with any degree of precision. There are two reasons why dividend discount model remain widely used to value financial service companies. The first is that estimating capital expenditures and working capital for a bank, an investment bank or an insurance company is difficult to do.[18] The second is that retained earnings and book equity have real consequences for financial service companies since their regulatory capital ratios are computed on the basis of book value of equity.

In summary, then, the dividend discount model has far more applicability than its critics concede. Even the conventional wisdom that the dividend discount model cannot be used to value a stock that pays low or no dividends is wrong. If the dividend payout ratio is adjusted to reflect changes in the expected growth rate, a reasonable value can be obtained even for non-dividend paying firms. Thus, a high-growth firm, paying no dividends currently, can still be valued based upon dividends that it is expected to pay out when the growth rate declines. In practice, Michaud and Davis (1981) note that the dividend discount model is biased towards finding stocks with high dividend yields and low P/E ratios to be under valued.[19] They argue that the anti-growth bias of the dividend discount model can be traced to the use of fixed and often arbitrary risk premiums and costs of equity, and suggest that the bias can be reduced or even eliminated with the use of market implied risk premiums and returns.

How well does the dividend discount model work?

The true measure of a valuation model is how well it works in (i) explaining differences in the pricing of assets at any point in time and across time and (ii) how quickly differences between model and market prices get resolved.

Researchers have come to mixed conclusions on the first question, especially at it relates to the aggregate equity market. Shiller (1981) presents evidence that the volatility

---

[18] This is true for any firm whose primary asset is human capital. Accounting conventions have generally treated expenditure on human capital (training, recruiting etc.) as operating expenditures. Working capital is meaningless for a bank, at least in its conventional form since current assets and liabilities comprise much of what is on the balance sheet.

[19] Michaud, R.O. and P.L. Davis, 1981, Valuation Model Bias and the Scale Structure of Dividend Discount Returns, Journal of Finance, v37, 563-573.

in stock prices is far too high to be explained by variance in dividends over time; in other words, market prices vary far more than the present value of dividends.[20] In attempts to explain the excess market volatility, Poterba and Summers (1988) argue that risk premiums can change over time[21] and Fama and French (1988) note that dividend yields are much more variable than dividends.[22] Looking at a much longer time period (1871-2003), Foerster and Sapp (2005) find that the dividend discount model does a reasonably good job of explaining variations in the S&P 500 index, though there are systematic differences over time in how investors value future dividends.[23]

To answer the second question, Sorensen and Williamson (1985) valued 150 stocks from the S&P 400 in December 1980, using the dividend discount model.[24] They used the difference between the market price at that time and the model value to form five portfolios based upon the degree of under or over valuation. They made fairly broad assumptions in using the dividend discount model:

(a) The average of the earnings per share between 1976 and 1980 was used as the current earnings per share.

(b) The cost of equity was estimated using the CAPM.

(c) The extraordinary growth period was assumed to be five years for all stocks and the I/B/E/S consensus analyst forecast of earnings growth was used as the growth rate for this period.

(d) The stable growth rate, after the extraordinary growth period, was assumed to be 8% for all stocks.

(e) The payout ratio was assumed to be 45% for all stocks.

The returns on these five portfolios were estimated for the following two years (January 1981-January 1983) and excess returns were estimated relative to the S&P 500 Index using the betas estimated at the first stage. Figure 2 illustrates the excess returns earned

---

[20] Shiller, R., 1981, Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? American Economic Review, v71, 421-436.

[21] Poterba, J., and L. Summers, 1988, Mean reversion in stock prices: evidence and implications, Journal of Financial Economics, v22, 27-59.

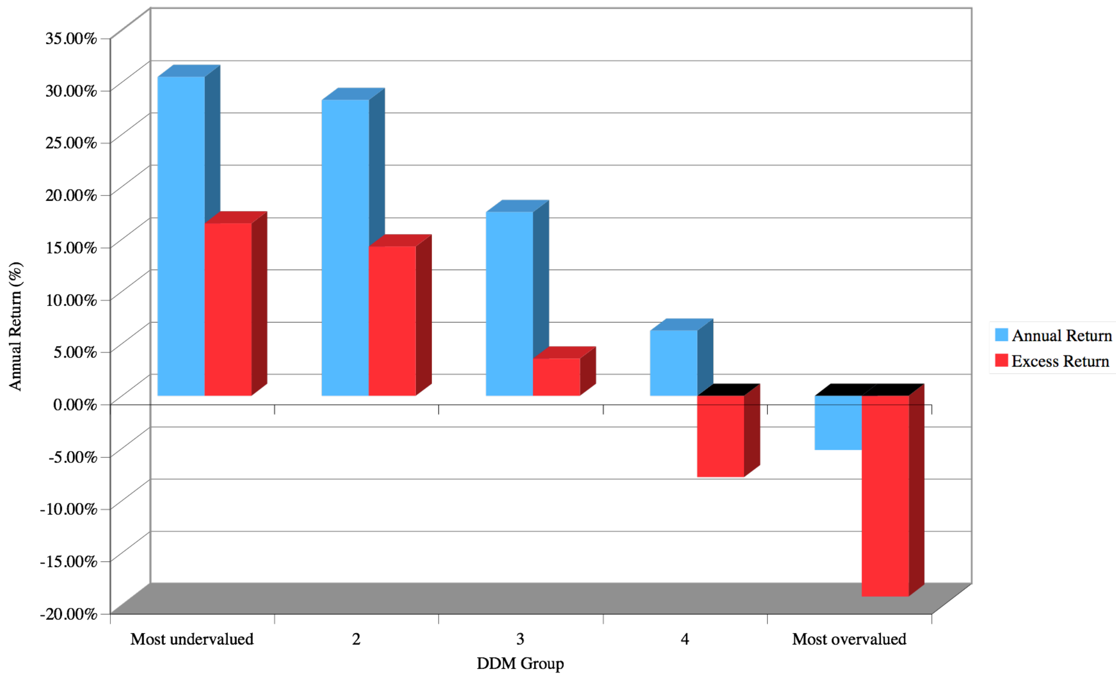
[22] Fama, E. and K. French, 1988, Dividend Yields and Expected Stock Returns, Journal of Financial Economics 22, 3-25.

[23] Foerster, S.R. and S.G. Sapp, 2005, Dividends and Stock Valuation: A Study of the Nineteenth to the Twenty-first Century, Working Paper, University of Western Ontario.

[24] Sorensen, E.H. and D.A. Williamson, 1985, Some Evidence on the Value of the Dividend Discount

by the portfolio that was undervalued by the dividend discount model relative to both the market and the overvalued portfolio.

Figure 2: Performance of Dividend Discount Model



The undervalued portfolio had a positive excess return of 16% per annum between 1981 and 1983, while the overvalued portfolio had a negative excess return of almost 20% per annum during the same time period. In the long term, undervalued (overvalued) stocks from the dividend discount model outperform (under perform) the market index on a risk-adjusted basis. However, this result should be taken with a grain of salt, given that the dividend discount model tends to find stocks with high dividend yields and low PE ratios to be under valued, and there is well established empirical evidence showing that stocks with those characteristics generate excess returns, relative to established risk and return models in finance. In other words, it is unclear how much of the superior performance attributed to the dividend discount model could have been replicated with a far simpler strategy of buying low PE stocks with high dividend yields.

Model, Financial Analysts Journal, v41, 60-69.

b. Extended Equity Valuation Models

In the dividend discount model, we implicitly assume that firms pay out what they can afford to as dividends. In reality, though, firms often choose not to do so. In some cases, they accumulate cash in the hope of making investments in the future. In other cases, they find other ways, including buybacks, of returning cash to stockholders. Extended equity valuation models try to capture this cash build-up in value by considering the cash that could have been paid out in dividends rather than the actual dividends.

Dividends versus Potential Dividends

Fama and French (2001) report that only 20.8% of firms paid dividends in 1999, compared with 66.5% in 1978 and find that only a portion of the decline can be attributed to changes in firm characteristics; there were more small cap, high growth firms in 1999 than in 1978. After controlling for differences, they conclude that firms became less likely to pay dividends over the period.[25]

The decline in dividends over time has been attributed to a variety of factors. DeAngelo, DeAngelo and Skinner (2004) argue that aggregate dividends paid by companies has not decreased and that the decreasing dividends can be traced to smaller firms that are uninterested in paying dividends.[26] Baker and Wurgler (2004) provide a behavioral rationale by pointing out that the decrease in dividends over time can be attributed to an increasing segment of investors who do not want dividends.[27] Hoberg and Prabhala (2005) posit that the decrease in dividends is because of an increase in risk, by noting that increases in idiosyncratic risk (rather than dividend clientele) explain the drop in dividends.[28] Notwithstanding the reasons, the gap between dividends paid and potential dividends has increased over time both in the aggregate and for individual firms, creating a challenge to those who use dividend discount models.

---

[25] Fama, E.F. and K.R. French, 2001, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay?, Journal of Financial Economics 60, 3–44.

[26] DeAngelo, H., L. DeAngelo, and D. Skinner, 2004, Are dividends disappearing? Dividend concentration and the consolidation of earnings, Journal of Financial Economics, v72, 425–456.

[27] Baker, M., and J. Wurgler, 2004a, Appearing and disappearing dividends: The link to catering incentives, Journal of Financial Economics 73, 271–288. Baker, M., and J. Wurgler 2004b, A catering theory of dividends, The Journal of Finance 59, 1125–1165.

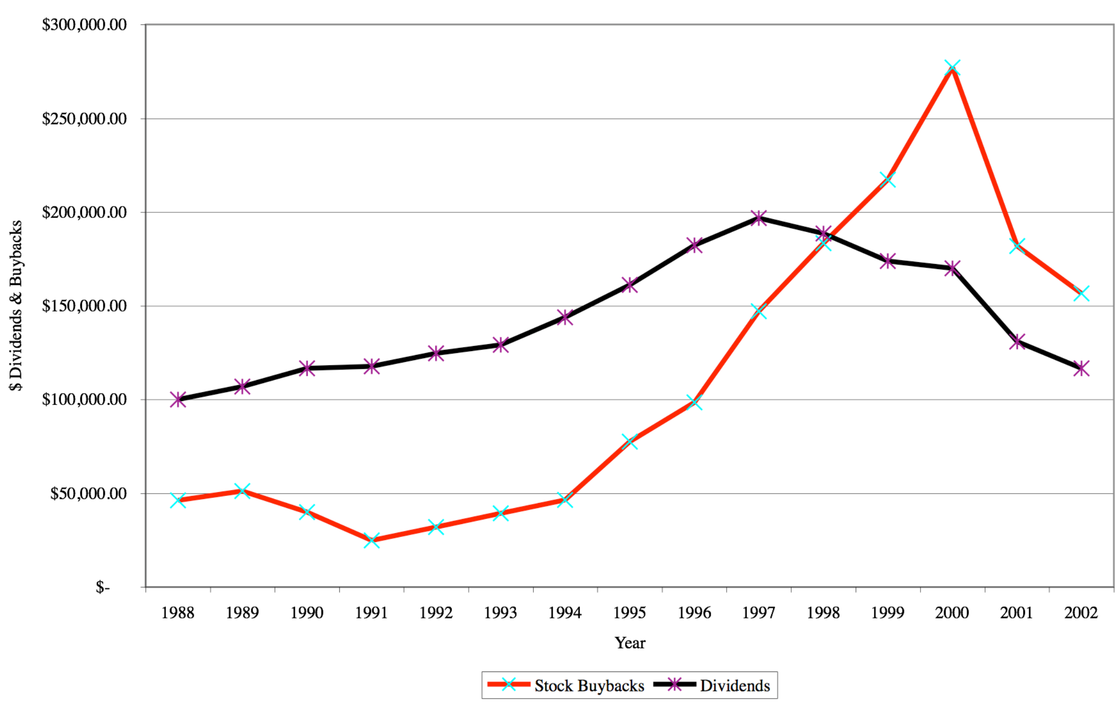
[28] Hoberg, G. and N.R. Prabhala, 2005, Disappearing Dividends: The Importance of idiosyncratic risk and the irrelevance of catering, Working Paper, University of Maryland.

One fix for this problem is to replace dividends in the dividend discount models with potential dividends, but that raises an estimation question: How do we best estimate potential dividends? There are three suggested variants. In the first, we extend our definition of cash returned to stockholders to include stock buybacks, thus implicitly assuming that firms that accumulate cash by not paying dividends return use them to buy back stock. In the second, we try to compute the cash that could have been paid out as dividends by estimating the residual cash flow after meeting reinvestment needs and making debt payments. In the third, we either accounting earnings or variants of earnings as proxies for potential dividends.

Buybacks as Dividends

One reason for the fall of the dividend discount model from favor has been the increased use of stock buybacks as a way of returning cash to stockholders. A simple response to this trend is to expand the definition of dividends to include stock buybacks and to value stocks based on this composite number. In recent years, firms in the United States have increasingly turned to stock buybacks as a way of returning cash to stockholders. Figure 3 presents the cumulative amounts paid out by firms in the form of dividends and stock buybacks from 1989 to 2002.

Figure 3: Stock Buybacks and Dividends: Aggregate for US Firms - 1989-2002



The trend towards stock buybacks is very strong, especially in the 1990s. By early 2000, more cash was being returned to stockholders in stock buybacks than in conventional dividends.

What are the implications for the dividend discount model? Focusing strictly on dividends paid as the only cash returned to stockholders exposes us to the risk that we might be missing significant cash returned to stockholders in the form of stock buybacks. The simplest way to incorporate stock buybacks into a dividend discount model is to add them on to the dividends and compute a modified payout ratio:

$$\text{Modified dividend payout ratio} = \frac{\text{Dividends} + \text{Stock Buybacks}}{\text{Net Income}}$$

While this adjustment is straightforward, the resulting ratio for any year can be skewed by the fact that stock buybacks, unlike dividends, are not smoothed out. In other words, a firm may buy back $ 3 billion in stock in one year and not buy back stock for the next 3 years. Consequently, a much better estimate of the modified payout ratio can be obtained by looking at the average value over a four or five year period. In addition, firms may sometimes buy back stock as a way of increasing financial leverage. If this is a concern, we could adjust for this by netting out new debt issued from the calculation above:

$$\text{Modified dividend payout} = \frac{\text{Dividends} + \text{Stock Buybacks} - \text{Long Term Debt issues}}{\text{Net Income}}$$

Damodaran (2006) presents this extension to the basic dividend discount model and argues that it works well in explaining the market prices of companies that follow a policy of returning cash over regular intervals in the form of stock buybacks.[29]

Free Cash Flow to Equity (FCFE) Model

The free cash flow to equity model does not represent a radical departure from the traditional dividend discount model. In fact, one way to describe a free cash flow to equity model is that it represents a model where we discount potential dividends rather than actual dividends. Damodaran (1994) a measure of free cash flow to equity that captures the cash flow left over all reinvestment needs and debt payments:

FCFE = Net Income + Depreciation - Capital Expenditures – Change in non-cash Working Capital – (New Debt Issued – Debt repayments)

Practitioners have long used variants of free cash flow to equity to judge the attractiveness of companies as investments. Buffett, for instance, has argued that investors should judge companies based upon what he called “owner’s earnings”, which he defined to be cash flows left over after capital expenditures and working capital needs, a measure of free cash flow to equity that ignores cash flows from debt.[30]

When we replace the dividends with FCFE to value equity, we are doing more than substituting one cash flow for another. We are implicitly assuming that the FCFE will be paid out to stockholders. There are two consequences.

1. There will be no future cash build-up in the firm, since the cash that is available after debt payments and reinvestment needs is paid out to stockholders each period.
2. The expected growth in FCFE will include growth in income from operating assets and not growth in income from increases in marketable securities. This follows directly from the last point.

How does discounting free cashflows to equity compare with the modified dividend discount model, where stock buybacks are added back to dividends and discounted? You can consider stock buybacks to be the return of excess cash accumulated largely as a consequence of not paying out their FCFE as dividends. Thus, FCFE represent a smoothed out measure of what companies can return to their stockholders over time in the form of dividends and stock buybacks.

The FCFE model treats the stockholder in a publicly traded firm as the equivalent of the owner in a private business. The latter can lay claim on all cash flows left over in the business after taxes, debt payments and reinvestment needs have been met. Since the free cash flow to equity measures the same for a publicly traded firm, we are assuming that stockholders are entitled to these cash flows, even if managers do not choose to pay them out. In essence, the FCFE model, when used in a publicly traded firm, implicitly assumes that there is a strong corporate governance system in place. Even if stockholders cannot force managers to return free cash flows to equity as dividends, they can put pressure on managers to ensure that the cash that does not get paid out is not wasted.

---

[29] Damodaran, A. 2006, Damodaran on Valuation, Second Edition, John Wiley and Sons, New York.
[30] Hagstrom, R., 2004, The Warren Buffett Way, John Wiley, New York.

As with the dividend discount model, there are variations on the free cashflow to equity model, revolving around assumptions about future growth and reinvestment needs. The constant growth FCFE model is designed to value firms that are growing at a stable rate and are hence in steady state. The value of equity, under the constant growth model, is a function of the expected FCFE in the next period, the stable growth rate and the required rate of return.

$$P_0 = \frac{\text{Expected FCFE}_1}{\text{Cost of Equity} - \text{Stable Growth Rate}}$$

The model is very similar to the Gordon growth model in its underlying assumptions and works under some of the same constraints. The growth rate used in the model has to be less than or equal to the expected nominal growth rate in the economy in which the firm operates. The assumption that a firm is in steady state also implies that it possesses other characteristics shared by stable firms. This would mean, for instance, that capital expenditures, relative to depreciation, are not disproportionately large and the firm is of 'average' risk. Damodaran (1994, 2002) examines two-stage and multi-stage versions of these models with the estimation adjustments that have to be made as growth decreases over time. The assumptions about growth are similar to the ones made by the multi-stage dividend discount model, but the focus is on FCFE instead of dividends, making it more suited to value firms whose dividends are significantly higher or lower than the FCFE. In particular, it gives more realistic estimates of value for equity for high growth firms that are expected to have negative cash flows to equity in the near future. The discounted value of these negative cash flows, in effect, captures the effect of the new shares that will be issued to fund the growth during the period, and thus indirectly captures the dilution effect of value of equity per share today.

Earnings Models

The failure of companies to pay out what they can afford to in dividends and the difficulties associated with estimating cash flows has led some to argue that firms are best valued by discounting earnings or variants of earnings. Ohlson (1995) starts with the dividend discount model but adds an overlay of what he terms a “clean surplus” relation, where the goodwill on the balance sheet represents the present value of future abnormal earnings. He goes on to show that the value of a stock can be written in terms of its book value and capitalized current earnings, adjusted for dividends.[31] Feltham and Ohlson (1995) build on the same argument to establish a relationship between value and earnings.[32] Penman and Sougiannis (1997) also argue that GAAP earnings can be substituted for dividends in equity valuation, as long as analysts reduce future earnings and book value to reflect dividend payments.[33] Since these models are built as much on book value as they are on earnings, we will return to consider them in the context of accounting valuation models.

While it is possible, on paper, to establish the equivalence of earnings-based and dividend discount models, if done right, the potential for double counting remains high with the former. In particular, discounting earnings as if they were cash flows paid out to stockholders while also counting the growth that is created by reinvesting those earnings will lead to the systematic overvaluation of stocks. In one of the more egregious examples of this double counting, Glassman and Hassett (2000) assumed that equity was close to risk free in the long term and discounted earnings as cash flows, while counting on long term earnings growth set equal to nominal GDP growth, to arrive at the conclusion that the Dow Jones should be trading at three times its then prevailing level.[34]

Potential Dividend versus Dividend Discount Models

The FCFE model can be viewed as an alternative to the dividend discount model. Since the two approaches sometimes provide different estimates of value for equity, it is worth examining when they provide similar estimates of value, when they provide different estimates of value and what the difference tells us about the firm.

There are two conditions under which the value from using the FCFE in discounted cashflow valuation will be the same as the value obtained from using the dividend discount model. The first is the obvious one, where the dividends are equal to the FCFE. There are firms that maintain a policy of paying out excess cash as dividends

---

[31]Ohlson, J. 1995, Earnings, Book values and Dividends in Security Valuation, Contemporary Accounting Research, v11, 661-687.

[32]Feltham, G. and J. Ohlson. 1995. Valuation and Clean Surplus Accounting of Operation and Financial Activities, Contemporary Accounting Research, v11, 689-731.

[33] Penman, S. and T. Sougiannis, 1997. The Dividend Displacement Property and the Substitution of Anticipated Earnings for Dividends in Equity Valuation, The Accounting Review, v72, 1-21.

[34] Glassman, J. and K. Hassett, 2000, Dow 36,000: The New Strategy for Profiting from the Coming Rise in the Stock Market, Three Rivers Press.

either because they have pre-committed to doing so or because they have investors who expect this policy of them. The second condition is more subtle, where the FCFE is greater than dividends, but the excess cash (FCFE - Dividends) is invested in fairly priced assets (i.e. assets that earn a fair rate of return and thus have zero net present value). For instance, investing in financial assets that are fairly priced should yield a net present value of zero. To get equivalent values from the two approaches, though, we have to keep track of accumulating cash in the dividend discount model and add it to the value of equity. Damodaran (2006) provides an illustration of this equivalence.[35]

There are several cases where the two models will provide different estimates of value. First, when the FCFE is greater than the dividend and the excess cash either earns below-market interest rates or is invested in negative net present value assets, the value from the FCFE model will be greater than the value from the dividend discount model. There is reason to believe that this is not as unusual as it would seem at the outset. There are numerous case studies of firms, which having accumulated large cash balances by paying out low dividends relative to FCFE, have chosen to use this cash to overpay on acquisitions. Second, the payment of dividends less than FCFE lowers debt-equity ratios and may lead the firm to become under levered, causing a loss in value. In the cases where dividends are greater than FCFE, the firm will have to issue either new stock or debt to pay these dividends or cut back on its investments, leading to at least one of three negative consequences for value. If the firm issues new equity to fund dividends, it will face substantial issuance costs that decrease value. If the firm borrows the money to pay the dividends, the firm may become over levered (relative to the optimal) leading to a loss in value. Finally, if paying too much in dividends leads to capital rationing constraints where good projects are rejected, there will be a loss of value (captured by the net present value of the rejected projects). There is a third possibility and it reflects different assumptions about reinvestment and growth in the two models. If the same growth rate used in the dividend discount and FCFE models, the FCFE model will give a higher value than the dividend discount model whenever FCFE ar
e higher than dividends and a lower value when dividends exceed FCFE. In reality, the growth rate in FCFE should be different from the growth rate in dividends, because the free cash flow to equity is assumed to be paid out to stockholders. In general, when firms

[35] Damnodaran, A,, 2006, Damodaran on Valuation (Second edition), John Wiley & Sons, New York.

pay out much less in dividends than they have available in FCFE, the expected growth rate and terminal value will be higher in the dividend discount model, but the year-to-year cash flows will be higher in the FCFE model.

When the value using the FCFE model is different from the value using the dividend discount model, with consistent growth assumptions, there are two questions that need to be addressed - What does the difference between the two models tell us? Which of the two models is the appropriate one to use in evaluating the market price? The more common occurrence is for the value from the FCFE model to exceed the value from the dividend discount model. The difference between the value from the FCFE model and the value using the dividend discount model can be considered one component of the value of controlling a firm - it measures the value of controlling dividend policy. In a hostile takeover, the bidder can expect to control the firm and change the dividend policy (to reflect FCFE), thus capturing the higher FCFE value. As for which of the two values is the more appropriate one for use in evaluating the market price, the answer lies in the openness of the market for corporate control. If there is a sizable probability that a firm can be taken over or its management changed, the market price will reflect that likelihood and the appropriate benchmark to use is the value from the FCFE model. As changes in corporate control become more difficult, either because of a firm's size and/or legal or market restrictions on takeovers, the value from the dividend discount model will provide the appropriate benchmark for comparison.

2. Firm DCF Models

The alternative to equity valuation is to value the entire business. The value of the firm is obtained by discounting the free cashflow to the firm at the weighted average cost of capital. Embedded in this value are the tax benefits of debt (in the use of the after-tax cost of debt in the cost of capital) and expected additional risk associated with debt (in the form of higher costs of equity and debt at higher debt ratios).

Basis for Firm Valuation Models

In the cost of capital approach, we begin by valuing the firm, rather than the equity. Netting out the market value of the non-equity claims from this estimate yields the value of equity in the firm. Implicit in the cost of capital approach is the assumption that the cost of capital captures both the tax benefits of borrowing and the expected bankruptcy costs. The cash flows discounted are the cash flows to the firm, computed as if the firm had no debt and no tax benefits from interest expenses.

The origins of the firm valuation model lie in one of corporate finance's most cited papers by Miller and Modigliani (1958) where they note that the value of a firm can be written as the present value of its after-tax operating cash flows:[36]

$$\text{Value of firm} = \sum_{t=1}^{t=\infty} \frac{E(X_t - I_t)}{(1+\ \text{Cost of Capital})^t}$$

where X_t is the after-tax operating earnings and I_t is the investment made back into the firm's assets in year t. The focus of that paper was on capital structure, with the argument being that the cost of capital would remain unchanged as debt ratio changed in a world with no taxes, default risk and agency issues. While there are varying definitions of the expected after-tax operating cash flow in use, the most common one is the free cash flow to the firm, defined as follows:

Free Cash Flow to Firm = After-tax Operating Income – (Capital Expenditures – Depreciation) – Change in non-cash Working Capital

In essence, this is a cash flow after taxes and reinvestment needs but before any debt payments, thus providing a contrast to free cashflows to equity that are after interest payments and debt cash flows.

There are two things to note about this model. The first is that it is general enough to survive the relaxing of the assuming of financing irrelevance; in other words, the value of the firm is still the present value of the after-tax operating cash flows in a world where the cost of capital changes as the debt ratio changes. Second, while it is a widely held preconception that the cost of capital approach requires the assumption of a constant debt ratio, the approach is flexible enough to allow for debt ratios that change over time. In fact, one of the biggest strengths of the model is the ease with which changes in the financing mix can be built into the valuation through the discount rate rather than through the cash flows.

The most revolutionary and counter intuitive idea behind firm valuation is the notion that equity investors and lenders to a firm are ultimately partners who supply capital to the firm and share in its success. The primary difference between equity and

---

[36]Modigliani, F. and M. Miller, 1958, The Cost of Capital, Corporation Finance and the Theory of Investment, American Economic Review, v48, 261-297.

debt holders in firm valuation models lies in the nature of their cash flow claims – lenders get prior claims to fixed cash flows and equity investors get residual claims to remaining cash flows.

Variations on firm valuation models

As with the dividend discount and FCFE models, the FCFF model comes in different forms, largely as the result of assumptions about how high the expected growth is and how long it is likely to continue. As with the dividend discount and FCFE models, a firm that is growing at a rate that it can sustain in perpetuity – a stable growth rate – can be valued using a stable growth mode using the following equation:

$$\text{Value of firm} = \frac{FCFF_1}{WACC - g_n}$$

where,

$FCFF_1$ = Expected FCFF next year

WACC = Weighted average cost of capital

g_n = Growth rate in the FCFF (forever)

There are two conditions that need to be met in using this model, both of which mirror conditions imposed in the dividend discount and FCFE models. First, the growth rate used in the model has to be less than or equal to the growth rate in the economy – nominal growth if the cost of capital is in nominal terms, or real growth if the cost of capital is a real cost of capital. Second, the characteristics of the firm have to be consistent with assumptions of stable growth. In particular, the reinvestment rate used to estimate free cash flows to the firm should be consistent with the stable growth rate. Implicit in the use of a constant cost of capital for a growing firm is the assumption that the debt ratio of the firm is held constant over time. The implications of this assumption were examined in Miles and Ezzel (1980), who noted that the approach not only assumed tax savings that would grow in perpetuity but that these tax savings were, in effect, being discounted as the unlevered cost of equity to arrive at value.[37]

Like all stable growth models, this one is sensitive to assumptions about the expected growth rate. This sensitivity is accentuated, however, by the fact that the

---

[37] Miles, J. and J.R. Ezzell, 1980, The weighted average cost of capital, perfect capital markets and project life: A clarification, Journal of Financial and Quantitative Analysis, v40, 1485-1492.

discount rate used in valuation is the WACC, which is lower than the cost of equity for most firms. Furthermore, the model is sensitive to assumptions made about capital expenditures relative to depreciation. If the inputs for reinvestment are not a function of expected growth, the free cashflow to the firm can be inflated (deflated) by reducing (increasing) capital expenditures relative to depreciation. If the reinvestment rate is estimated from the return on capital, changes in the return on capital can have significant effects on firm value.

Rather than break the free cash flow model into two-stage and three-stage models and risk repeating what was said earlier, we present the general version of the model in this section. The value of the firm, in the most general case, can be written as the present value of expected free cashflows to the firm.

$$\text{Value of Firm} = \sum_{t=1}^{t=\infty} \frac{\text{FCFF}_t}{(1+\text{WACC})^t}$$

where,

FCFF_t = Free Cashflow to firm in year t

WACC = Weighted average cost of capital

If the firm reaches steady state after n years and starts growing at a stable growth rate g_n after that, the value of the firm can be written as:

$$\text{Value of Operating Assets of the firm} = \sum_{t=1}^{t=n} \frac{\text{FCFF}_t}{(1+\text{WACC})^t} + \frac{[\text{FCFF}_{n+1}/(\text{WACC}-g_n)]}{(1+\text{WACC})^n}$$

Since the cash flows used are cash flows from the operating assets, the cost of capital that is used should reflect only the operating risk of the company. It also follows that the present value of the cash flows obtained by discounting the cash flows at the cost of capital will measure the value of only the operating assets of the firm (which contribute to the operating income). Any assets whose earnings are not part of operating income have not been valued yet. The McKinsey books on valuation have provided extensive coverage both of the estimation questions associated with discounted cash flow valuation and the link between value and corporate financial decisions.[38]

To get from the value of operating assets to the value of equity, we have to first incorporate the value of non-operating assets that are owned by the firm and then subtract out all non-equity claims that may be outstanding against the firm. Non-operating assets include all assets whose earnings are not counted as part of the operating income. The most common of the non-operating assets is cash and marketable securities, which can often amount to billions at large corporations and the value of these assets should be added on to the value of the operating assets. In addition, the operating income from minority holdings in other companies is not included in the operating income and FCFF; we therefore need to value these holdings and add them on to the value of the operating assets. Finally, the firm may own idle and unutilized assets that do not generate earnings or cash flows. These assets can still have value and should be added on to the value of the operating assets. The non-equity claims that have to be subtracted out include not only all debt, but all capitalized leases as well as unfunded pension plan and health care obligations. Damodaran (2006) contains extensive discussions of the adjustments that have to be made to arrive at equity value and further still at equity value per share.[39]

Firm versus Equity Valuation Models

This firm valuation model, unlike the dividend discount model or the FCFE model, values the firm rather than equity. The value of equity, however, can be extracted from the value of the firm by subtracting out the market value of outstanding debt. Since this model can be viewed as an alternative way of valuing equity, two questions arise - Why value the firm rather than equity? Will the values for equity obtained from the firm valuation approach be consistent with the values obtained from the equity valuation approaches described in the previous section?

The advantage of using the firm valuation approach is that cashflows relating to debt do not have to be considered explicitly, since the FCFF is a pre-debt cashflow, while they have to be taken into account in estimating FCFE. In cases where the leverage is

---

[38] Copeland, T.E., T. Koller and J. Murrin, 1990, Valuation: Measuring and Managing the Value of Companies, John Wiley and Sons (first three editions) and Koller, T., M. Goedhart and D. Wessels, 2005, Valuation: Measuring and Managing the Value of Companies, John Wiley and Sons (Fourth Edition).
[39] Damodaran, A., 2006, Damodaran on Valuation, Second Edition, John Wiley and Sons, New York.

expected to change significantly over time, this is a significant saving, since estimating new debt issues and debt repayments when leverage is changing can become increasingly difficult, the further into the future you go. The firm valuation approach does, however, require information about debt ratios and interest rates to estimate the weighted average cost of capital.

The value for equity obtained from the firm valuation and equity valuation approaches will be the same if you make consistent assumptions about financial leverage. Getting them to converge in practice is much more difficult. Let us begin with the simplest case – a no-growth, perpetual firm. Assume that the firm has $166.67 million in earnings before interest and taxes and a tax rate of 40%. Assume that the firm has equity with a market value of $600 million, with a cost of equity of 13.87% debt of $400 million and with a pre-tax cost of debt of 7%. The firm's cost of capital can be estimated.

$$\text{Cost of capital} = (13.87\%)\left(\frac{600}{1000}\right) + (7\%)(1 - 0.4)\left(\frac{400}{1000}\right) = 10\%$$

$$\text{Value of the firm} = \frac{\text{EBIT}(1 - t)}{\text{Cost of capital}} = \frac{166.67(1 - 0.4)}{0.10} = \$1{,}000$$

Note that the firm has no reinvestment and no growth. We can value equity in this firm by subtracting out the value of debt.

Value of equity = Value of firm – Value of debt = $ 1,000 - $400 = $ 600 million

Now let us value the equity directly by estimating the net income:

Net Income = (EBIT – Pre-tax cost of debt * Debt) (1-t) = (166.67 - 0.07*400) (1-0.4) = 83.202 million

The value of equity can be obtained by discounting this net income at the cost of equity:

$$\text{Value of equity} = \frac{\text{Net Income}}{\text{Cost of equity}} = \frac{83.202}{0.1387} = \$\,600 \text{ million}$$

Even this simple example works because of the following assumptions that we made implicitly or explicitly during the valuation.

1. The values for debt and equity used to compute the cost of capital were equal to the values that we obtained in the valuation. Notwithstanding the circularity in reasoning – you need the cost of capital to obtain the values in the first place – it indicates that a cost of capital based upon market value weights will not yield the

same value for equity as an equity valuation model, if the firm is not fairly priced in the first place.

2. There are no extraordinary or non-operating items that affect net income but not operating income. Thus, to get from operating to net income, all we do is subtract out interest expenses and taxes.
3. The interest expenses are equal to the pre-tax cost of debt multiplied by the market value of debt. If a firm has old debt on its books, with interest expenses that are different from this value, the two approaches will diverge.

If there is expected growth, the potential for inconsistency multiplies. We have to ensure that we borrow enough money to fund new investments to keep our debt ratio at a level consistent with what we are assuming when we compute the cost of capital.

Certainty Equivalent Models

While most analysts adjust the discount rate for risk in DCF valuation, there are some who prefer to adjust the expected cash flows for risk. In the process, they are replacing the uncertain expected cash flows with the certainty equivalent cashflows, using a risk adjustment process akin to the one used to adjust discount rates.

Misunderstanding Risk Adjustment

At the outset of this section, it should be emphasized that many analysts misunderstand what risk adjusting the cash flows requires them to do. There are some who consider the cash flows of an asset under a variety of scenarios, ranging from best case to catastrophic, assign probabilities to each one, take an expected value of the cash flows and consider it risk adjusted. While it is true that bad outcomes have been weighted in to arrive at this cash flow, it is still an expected cash flow and is not risk adjusted. To see why, assume that you were given a choice between two alternatives. In the first one, you are offered \$ 95 with certainty and in the second, you will receive \$ 100 with probability 90% and only \$50 the rest of the time. The expected values of both alternatives is \$95 but risk averse investors would pick the first investment with guaranteed cash flows over the second one.

If this argument sounds familiar, it is because it is a throwback to the very beginnings of utility theory. In one of the most widely cited thought experiments in economics, Nicholas Bernoulli proposed a hypothetical gamble that updated would look something like this: He would flip a coin once and would pay you a dollar if the coin came up tails on the first flip; the experiment would stop if it came up heads. If you won the dollar on the first flip, though, you would be offered a second flip where you could double your winnings if the coin came up tails again. The game would thus continue, with the prize doubling at each stage, until you lost. How much, he wanted to know, would you be willing to pay to partake in this gamble? This gamble, called the St. Petersburg Paradox, has an expected value of infinity but no person would be willing to pay that much. In fact, most of us would pay only a few dollars to play this game. In that context, Bernoulli unveiled the notion of a certainty equivalent, a guaranteed cash flow that we would accept instead of an uncertain cash flow and argued that more risk averse investors would settle for lower certainty equivalents for a given set of uncertain cash flows than less risk averse investors. In the example given in the last paragraph, a risk averse investor would have settled for a guaranteed cash flow of well below \$95 for the second alternative with an expected cash flow of \$95.[40]

The practical question that we will address in this section is how best to convert uncertain expected cash flows into guaranteed certainty equivalents. While we do not disagree with the notion that it should be a function of risk aversion, the estimation challenges remain daunting.

Utility Models: Bernoulli revisited

The first (and oldest) approach to computing certainty equivalents is rooted in the utility functions for individuals. If we can specify the utility function of wealth for an individual, we are well set to convert risky cash flows to certainty equivalents for that individual. For instance, an individual with a log utility function would have demanded a certainty equivalent of \$79.43 for the risky gamble presented in the last section (90% chance of \$ 100 and 10% chance of \$ 50):

Utility from gamble = .90 ln(100) + .10 ln(50) = 4.5359

Certainty Equivalent = $\exp^{4.5359}$ = \$93.30

[40] Bernoulli, D., 1738, Exposition of a New Theory on the Measurement of Risk. Translated into English in Econometrica, January 1954.

The certainty equivalent of $93.30 delivers the same utility as the uncertain gamble with an expected value of $95. This process can be repeated for more complicated assets, and each expected cash flow can be converted into a certainty equivalent.[41]

One quirk of using utility models to estimate certainty equivalents is that the certainty equivalent of a positive expected cash flow can be negative. Consider, for instance, an investment where you can make $ 2000 with probability 50% and lose $ 1500 with probability 50%. The expected value of this investment is $ 250 but the certainty equivalent may very well be negative, with the effect depending upon the utility function assumed.

There are two problems with using this approach in practice. The first is that specifying a utility function for an individual or analyst is very difficult, if not impossible, to do with any degree of precision. In fact, most utility functions that are well behaved (mathematically) do not seem to explain actual behavior very well. The second is that, even if we were able to specify a utility function, this approach requires us to lay out all of the scenarios that can unfold for an asset (with corresponding probabilities) for every time period. Not surprisingly, certainty equivalents from utility functions have been largely restricted to analyzing simple gambles in classrooms.

Risk and Return Models

A more practical approach to converting uncertain cash flows into certainty equivalents is offered by risk and return models. In fact, we would use the same approach to estimating risk premiums that we employ while computing risk adjusted discount rates but we would use the premiums to estimate certainty equivalents instead.

Certainty Equivalent Cash flow = Expected Cash flow/ (1 + Risk Premium in Risk-adjusted Discount Rate)

Assume, for instance, that Google has a risk-adjusted discount rate of 13.45%, based upon its market risk exposure and current market conditions; the riskfree rate used was 4.25%. Instead of discounting the expected cash flows on the stock at 13.45%, we would

[41] Gregory, D.D., 1978, Multiplicative Risk Premiums, Journal of Financial and Quantitative Analysis, v13, 947-963. This paper derives certainty equivalent functions for quadratic, exponential and gamma distributed utility functions and examines their behavior.

decompose the expected return into a risk free rate of 4.25% and a compounded risk premium of 8.825%.[42]

$$\text{Compounded Risk Premium} = \frac{(1+\ \text{Risk adjusted Discount Rate})}{(1+\ \text{Riskfree Rate})} - 1 = \frac{(1.1345)}{(1.0425)} - 1 = .08825$$

If the expected cash flow in years 1 and 2 are \$ 100 million and \$ 120 million respectively, we can compute the certainty equivalent cash flows in those years:

Certainty Equivalent Cash flow in year 1 = \$ 100 million/1.08825 = \$ 91.89 million

Certainty Equivalent Cash flow in year 2 = \$120 million/ 1.08825^2 = \$ 101.33 million

This process would be repeated for all of the expected cash flows and it has two effects. Formally, the adjustment process for certainty equivalents can be then written more formally as follows (where the risk adjusted return is r and the riskfree rate is r_f):[43]

$$CE\ (CF_t) = \alpha_t\ E(CF_t) = \frac{(1+r_f)^t}{(1+r\ )^t} E(CF_t)$$

This adjustment has two effects. The first is that expected cash flows with higher uncertainty associated with them have lower certainty equivalents than more predictable cash flows at the same point in time. The second is that the effect of uncertainty compounds over time, making the certainty equivalents of uncertain cash flows further into the future lower than uncertain cash flows that will occur sooner.

Cashflow Haircuts

A far more common approach to adjusting cash flows for uncertainty is to "haircut" the uncertain cash flows subjectively. Thus, an analyst, faced with uncertainty, will replace uncertain cash flows with conservative or lowball estimates. This is a weapon commonly employed by analysts, who are forced to use the same discount rate for projects of different risk levels, and want to even the playing field. They will haircut the cash flows of riskier projects to make them lower, thus hoping to compensate for the failure to adjust the discount rate for the additional risk.

---

[42] A more common approximation used by many analysts is the difference between the risk adjusted discount rate and the risk free rate. In this case, that would have yielded a risk premium of 9.2% (13.45% - 4.25% = 9.20%)

[43] Robichek, A.A. and S. C. Myers, 1966, Conceptual Problems in the Use of Risk Adjusted Discount Rates, Journal of Finance, v21, 727-730.

In a variant of this approach, there are some investors who will consider only those cashflows on an asset that are predictable and ignore risky or speculative cash flows when valuing the asset. When Warren Buffet expresses his disdain for the CAPM and other risk and return models, and claims to use the riskfree rate as the discount rate, we suspect that he can get away with doing so because of a combination of the types of companies he chooses to invest in and his inherent conservatism when it comes to estimating the cash flows.

While cash flow haircuts retain their intuitive appeal, we should be wary of their usage. After all, gut feelings about risk can vary widely across analysts looking at the same asset; more risk averse analysts will tend to haircut the cashflows on the same asset more than less risk averse analysts. Furthermore, the distinction we drew between diversifiable and market risk when developing risk and return models can be completely lost when analysts are making intuitive adjustments for risk. In other words, the cash flows may be adjusted downwards for risk that will be eliminated in a portfolio. The absence of transparency about the risk adjustment can also lead to the double counting of risk, especially when the analysis passes through multiple layers of analysis. To provide an illustration, after the first analyst looking at a risky investment decides to use conservative estimates of the cash flows, the analysis may pass to a second stage, where his superior may decide to make an additional risk adjustment to the already risk adjusted cash flows.

Risk Adjusted Discount Rate or Certainty Equivalent Cash Flow

Adjusting the discount rate for risk or replacing uncertain expected cash flows with certainty equivalents are alternative approaches to adjusting for risk, but do they yield different values, and if so, which one is more precise? The answer lies in how we compute certainty equivalents. If we use the risk premiums from risk and return models to compute certainty equivalents, the values obtained from the two approaches will be the same. After all, adjusting the cash flow, using the certainty equivalent, and then discounting the cash flow at the riskfree rate is equivalent to discounting the cash flow at a risk adjusted discount rate. To see this, consider an asset with a single cash flow in one year and assume that r is the risk-adjusted cash flow, r_f is the riskfree rate and RP is the compounded risk premium computed as described earlier in this section.

Certainty Equivalent Value = $\frac{CE}{(1+r_f)} = \frac{E(CF)}{(1+RP)(1+r_f)} = \frac{E(CF)}{\frac{(1+r)}{(1+r_f)}(1+r_f)} = \frac{E(CF)}{(1+r)}$

This analysis can be extended to multiple time periods and will still hold.[44] Note, though, that if the approximation for the risk premium, computed as the difference between the risk-adjusted return and the risk free rate, had been used, this equivalence will no longer hold. In that case, the certainty equivalent approach will give lower values for any risky asset and the difference will increase with the size of the risk premium.

Are there other scenarios where the two approaches will yield different values for the same risky asset? The first is when the risk free rates and risk premiums change from time period to time period; the risk-adjusted discount rate will also then change from period to period. Robichek and Myers, in the paper we referenced earlier, argue that the certainty equivalent approach yields more precise estimates of value in this case. The other is when the certainty equivalents are computed from utility functions or subjectively, whereas the risk-adjusted discount rate comes from a risk and return model. The two approaches can yield different estimates of value for a risky asset. Finally, the two approaches deal with negative cash flows differently. The risk-adjusted discount rate discounts negative cash flows at a higher rate and the present value becomes less negative as the risk increases. If certainty equivalents are computed from utility functions, they can yield certainty equivalents that are negative and become more negative as you increase risk, a finding that is more consistent with intuition.[45]

The biggest dangers arise when analysts use an amalgam of approaches, where the cash flows are adjusted partially for risk, usually subjectively and the discount rate is also adjusted for risk. It is easy to double count risk in these cases and the risk adjustment to value often becomes difficult to decipher.

---

[44] The proposition that risk adjusted discount rates and certainty equivalents yield identical net present values is shown in the following paper: Stapleton, R.C., 1971, Portfolio Analysis, Stock Valuation and Capital Budgeting Decision Rules for Risky Projects, Journal of Finance, v26, 95-117.

[45] Beedles, W.L., 1978, Evaluating Negative Benefits, Journal of Financial and Quantitative Analysis, v13, 173-176.

Excess Return Models

The model that we have presented in this section, where expected cash flows are discounted back at a risk-adjusted discount rate is the most commonly used discounted cash flow approach but there are variants. In the excess return valuation approach, we separate the cash flows into excess return cash flows and normal return cash flows. Earning the risk-adjusted required return (cost of capital or equity) is considered a normal return cash flow but any cash flows above or below this number are categorized as excess returns; excess returns can therefore be either positive or negative. With the *excess return valuation* framework, the value of a business can be written as the sum of two components:

Value of business = Capital Invested in firm today + Present value of excess return cash flows from both existing and future projects

If we make the assumption that the accounting measure of capital invested (book value of capital) is a good measure of capital invested in assets today, this approach implies that firms that earn positive excess return cash flows will trade at market values higher than their book values and that the reverse will be true for firms that earn negative excess return cash flows.

Basis for Models

Excess return models have their roots in capital budgeting and the net present value rule. In effect, an investment adds value to a business only if it has positive net present value, no matter how profitable it may seem on the surface. This would also imply that earnings and cash flow growth have value only when it is accompanied by excess returns, i.e., returns on equity (capital) that exceed the cost of equity (capital). Excess return models take this conclusion to the logical next step and compute the value of a firm as a function of expected excess returns.

While there are numerous versions of excess return models, we will consider one widely used variant, which is economic value added (EVA) in this section. The economic value added (EVA) is a measure of the surplus value created by an investment or a portfolio of investments. It is computed as the product of the "excess return" made on an investment or investments and the capital invested in that investment or investments.

Economic Value Added = (Return on Capital Invested – Cost of Capital) (Capital Invested) = After-tax operating income – (Cost of Capital) (Capital Invested)

Economic value added is a simple extension of the net present value rule. The net present value of the project is the present value of the economic value added by that project over its life.[46]

$$\text{NPV} = \sum_{t=1}^{t=n} \frac{\text{EVA}_t}{\left(1+k_c\right)^t}$$

where EVA_t is the economic value added by the project in year t and the project has a life of n years and k_c is the cost of capital.

This connection between economic value added and NPV allows us to link the value of a firm to the economic value added by that firm. To see this, let us begin with a simple formulation of firm value in terms of the value of assets in place and expected future growth.[47]

Firm Value = Value of Assets in Place + Value of Expected Future Growth

Note that in a discounted cash flow model, the values of both assets in place and expected future growth can be written in terms of the net present value created by each component.

$$\text{Firm Value} = \text{Capital Invested}_{\text{Assets in Place}} + \text{NPV}_{\text{Assets in Place}} + \sum_{t=1}^{t=\infty} \text{NPV}_{\text{Future Projects, t}}$$

Substituting the economic value added version of net present value into this equation, we get:

$$\text{Firm Value} = \text{Capital Invested}_{\text{Assets in Place}} + \sum_{t=1}^{t=\infty} \frac{\text{EVA}_{\text{t, Assets in Place}}}{\left(1+k_c\right)} + \sum_{t=1}^{t=\infty} \frac{\text{EVA}_{\text{t, Future Projects}}}{\left(1+k_c\right)}$$

Thus, the value of a firm can be written as the sum of three components, the capital invested in assets in place, the present value of the economic value added by these assets and the expected present value of the economic value that will be added by future investments.[48]

---

[46] This is true, though, only if the expected present value of the cash flows from depreciation is assumed to be equal to the present value of the return of the capital invested in the project. A proof of this equality can be found in Damodaran, A, 1999, Value Enhancement: Back to Basics, Contemporary Finance Digest, v2, 5-51.

[47] Brealey, R.A. and S. C. Myers, 2003, Principles of Corporate Finance (Seventh Edition), McGraw-Hill Irwin.

[48] Brealery, A., 2004, Investment Valuation, Second Edition, John Wiley & Sons, New York.

Measuring Economic Value Added

The definition of EVA outlines three basic inputs we need for its computation - the return on capital earned on investments, the cost of capital for those investments and the capital invested in them. In measuring each of these, we will make many of the same adjustments we discussed in the context of discounted cash flow valuation. Stewart (1991) and Young and O'Byrne (2000) extensively cover the computation of economic value added in their books on the topic.[49]

How much *capital is invested* in existing assets? One obvious answer is to use the market value of the firm, but market value includes capital invested not just in assets in place but in expected future growth[50]. Since we want to evaluate the quality of assets in place, we need a measure of the capital invested in these assets. Given the difficulty of estimating the value of assets in place, it is not surprising that we turn to the book value of capital as a proxy for the capital invested in assets in place. The book value, however, is a number that reflects not just the accounting choices made in the current period, but also accounting decisions made over time on how to depreciate assets, value inventory and deal with acquisitions. The older the firm, the more extensive the adjustments that have to be made to book value of capital to get to a reasonable estimate of the market value of capital invested in assets in place. Since this requires that we know and take into account every accounting decision over time, there are cases where the book value of capital is too flawed to be fixable. Here, it is best to estimate the capital invested from the ground up, starting with the assets owned by the firm, estimating the market value of these assets and cumulating this market value. To evaluate the return on this invested capital, we need an estimate of the *after-tax operating income* earned by a firm on these investments. Again, the accounting measure of operating income has to be adjusted for operating leases, R&D expenses and one-time charges to compute the return on capital. The third and final component needed to estimate the economic value added is the *cost of capital*. In keeping with arguments both in the investment analysis and the discounted cash flow valuation sections, the cost of capital should be estimated based upon the

---

[49] Stewart , G. B. (1991), The Quest for Value. The EVA Management Guide. Harper Business; Young, S.D and S.F. OByrne, 2000, EVA and Value-Based Management, McGraw Hill,

[50] As an illustration, computing the return on capital at Google using the market value of the firm, instead of book value, results in a return on capital of about 1%. It would be a mistake to view this as a sign of poor investments on the part of the firm's managers.

market values of debt and equity in the firm, rather than book values. There is no contradiction between using book value for purposes of estimating capital invested and using market value for estimating cost of capital, since a firm has to earn more than its market value cost of capital to generate value. From a practical standpoint, using the book value cost of capital will tend to understate cost of capital for most firms and will understate it more for more highly levered firms than for lightly levered firms. Understating the cost of capital will lead to overstating the economic value added.

In a survey of practices of firms that used economic value added, Weaver (2001) notes that firms make several adjustments to operating income and book capital in computing EVA, and that the typical EVA calculation involves 19 adjustments from a menu of between 9 and 34 adjustments. In particular, firms adjust book value of capital and operating income for goodwill, R&D and leases, before computing return on capital.[51]

Variants on Economic Value Added

There are several variants on economic value added that build on excess returns. While they share the same basic foundation – that value is created by generating excess returns on investments – they vary in how excess returns are computed.

- In Economic Profit, the excess return is defined from the perspective of equity investors and thus is based on net income and cost of equity, rather than after-tax operating income and cost of capital

  Economic Profit = Net Income – Cost of Equity * Book Value of Equity

  Many of the papers that we referenced in the context of earnings-based valuation models, especially by Ohlson, are built on this theme. We will examine these models in the context of accounting based valuations later in this paper.[52]
- In Cash Flow Return on Investment or CFROI models, there are two significant differences. The first is that the return earned on investments is computed not based on accounting earnings but on after-tax cash flow. The second is that both returns and the cost of capital are computed in real terms rather than nominal terms. Madden

[51] Weaver, S. C., 2001, Measuring Economic Value Added: A Survey of the Practices of EVA Proponents, Journal of Applied Finance, Fall/Winter, pp. 7-17.

(1998) provides an extensive analysis of the CFROI approach and what he perceives as its advantages over conventional accounting-based measures.[53]

While proponents of each measure claim its superiority, they agree on far more than they disagree on. Furthermore, the disagreements are primarily in which approach computes the excess return earned by a firm best, rather than on the basic premise that the value of a firm can be written in terms of its capital invested and the present value of its excess return cash flows.

Equivalence of Excess Return and DCF Valuation Models

It is relatively simple to show that the discounted cash flow value of a firm should match the value that you obtain from an excess return model, if you are consistent in your assumptions about growth and reinvestment. In particular, excess return models are built around a link between reinvestment and growth; in other words, a firm can generate higher earnings in the future only by reinvesting in new assets or using existing assets more efficiently. Discounted cash flow models often do not make this linkage explicit, even though you can argue that they should. Thus, analysts will often estimate growth rates and reinvestment as separate inputs and not make explicit links between the two.

Illustrating that discounted cash flow models and excess return models converge when we are consistent about growth and reinvestment is simple. The equivalence of discounted cash flow firm valuations and EVA valuations is shown in several papers: Fernandez (2002), Hartman (2000) and Shrieves and Wachowicz (2000).[54] In a similar vein, Feltham and Ohlson (1995), Penman (1998) and Lundholm and O'Keefe (2001) all provide proof that equity excess return models converge on equity discounted cash flow models.[55]

---

[52] Ohlson, J. 1995, Earnings, Book values and Dividends in Security Valuation, Contemporary Accounting Research, v11, 661-687.

[53] Madden. B.L., 1998, CFROI Cash Flow Return on Investment Valuation: A Total System Approach to Valuing a Firm, Butterworth-Heinemann.

[54] Fernandez, P., 2002, Three Residual Income Valuation Models and Discounted Cash Flow Valuation, Working Paper, IESE Business School; Hartman, J. C., 2000, On the Equivalence of Net Present Value and Economic Value Added as Measures of a Project's Economic Worth, *The Engineering Economist,* v45, 158-165.*;* Shrieves, R.E. and J.M. Wachowicz, 2000, Free Cash Flow, Economic Value Added and Net Present Value: A Reconciliation of Variations of Discounted Cash Flow Valuation, Working Paper, University of Tennessee.

[55] Feltham, G. and J. Ohlson, 1995, Valuation and Clean Surplus Accounting of Operation and Financial

The model values can diverge because of differences in assumptions and ease of estimation. Penman and Sourgiannis (1998) compared the dividend discount model to excess return models and concluded that the valuation errors in a discounted cash flow model, with a ten-year horizon, significantly exceeded the errors in an excess return model.[56] They attributed the difference to GAAP accrual earnings being more informative than either cash flows or dividends. Francis, Olson and Oswald (1999) concurred with Penman and also found that excess return models outperform dividend discount models.[57] Courteau, Kao and Richardson (2001) argue that the superiority of excess return models in these studies can be attributed entirely to differences in the terminal value calculation and that using a terminal price estimated by Value Line (instead of estimating one) results in dividend discount models outperforming excess return models.[58]

Adjusted Present Value Models

In the *adjusted present value (APV) approach*, we separate the effects on value of debt financing from the value of the assets of a business. In contrast to the conventional approach, where the effects of debt financing are captured in the discount rate, the APV approach attempts to estimate the expected dollar value of debt benefits and costs separately from the value of the operating assets.

Basis for APV Approach

In the APV approach, we begin with the value of the firm without debt. As we add debt to the firm, we consider the net effect on value by considering both the benefits and the costs of borrowing. In general, using debt to fund a firm's operations creates tax

---

Activities, Contemporary Accounting Research, v11, 689-731; Penman, S.H., 1998, A Synthesis of Equity Valuation Techniques and the Terminal Value Calculation for the Dividend Discount Model, Review of Accounting Studies, *v*2, 303-323; Lundholm, R., and T. O'Keefe. 2001. Reconciling value estimates from the discounted cash flow model and the residual income model. Contemporary Accounting Research, *v*18, 311-35.

[56] Penman, S. and T. Sougiannis. 1998. A Comparison of Dividend, Cash Flow, and Earnings Approaches to Equity Valuation, Contemporary Accounting Research, v15, 343-383.

[57] Francis, J., P. Olsson, and D. Oswald. 2000. Comparing the Accuracy and Explainability of Dividend, Free Cash Flow and Abnormal Earnings Equity Value Estimates. Journal of Accounting Research, v38, 45-70.

[58] Courteau, L., J. Kao and G.D. Richardson, 2001, The Equivalence of Dividend, Cash Flow and Residual Earnings Approaches to Equity Valuation Employing Ideal Terminal Value Calculations, Contemporary Accounting Research, v18 ,625–661.

benefits (because interest expenses are tax deductible) on the plus side and increases bankruptcy risk (and expected bankruptcy costs) on the minus side. The value of a firm can be written as follows:

Value of business = Value of business with 100% equity financing + Present value of Expected Tax Benefits of Debt – Expected Bankruptcy Costs

The first attempt to isolate the effect of tax benefits from borrowing was in Miller and Modigliani (1963), where they valued the present value of the tax savings in debt as a perpetuity using the cost of debt as the discount rate.[59] The adjusted present value approach, in its current form, was first presented in Myers (1974) in the context of examining the interrelationship between investment and financing decisions. [60]

Implicitly, the adjusted present value approach is built on the presumption that it is easier and more precise to compute the valuation impact of debt in absolute terms rather than in proportional terms. Firms, it is argued, do not state target debt as a ratio of market value (as implied by the cost of capital approach) but in dollar value terms.

Measuring Adjusted Present Value

In the adjusted present value approach, we estimate the value of the firm in three steps. We begin by estimating the value of the firm with no leverage. We then consider the present value of the interest tax savings generated by borrowing a given amount of money. Finally, we evaluate the effect of borrowing the amount on the probability that the firm will go bankrupt, and the expected cost of bankruptcy.

The first step in this approach is the estimation of the value of the unlevered firm. This can be accomplished by valuing the firm as if it had no debt, i.e., by discounting the expected free cash flow to the firm at the unlevered cost of equity. In the special case where cash flows grow at a constant rate in perpetuity, the value of the firm is easily computed.

$$\text{Value of Unlevered Firm} = \frac{\text{FCFF}_o(1+g)}{\rho_u - g}$$

---

[59] Modigliani, F. and M. Miller (1963), Corporate Income Taxes and the Cost of Capital: A Correction, American Economic Review, v53, 433-443.

[60] Myers, S., 1974, Interactions in Corporate Financing and Investment Decisions—Implications for Capital Budgeting, Journal of Finance, v29,1-25.

where $FCFF_0$ is the current after-tax operating cash flow to the firm, ρ_u is the unlevered cost of equity and g is the expected growth rate. In the more general case, we can value the firm using any set of growth assumptions we believe are reasonable for the firm. The inputs needed for this valuation are the expected cashflows, growth rates and the unlevered cost of equity.

The second step in this approach is the calculation of the expected tax benefit from a given level of debt. This tax benefit is a function of the tax rate of the firm and is discounted to reflect the riskiness of this cash flow.

$$\text{Value of Tax Benefits} = \sum_{t=1}^{t=\infty} \frac{\text{Tax Rate}_t \ * \ \text{Interest Rate}_t * \text{Debt}_t}{(1+\text{r})^{\text{t}}}$$

There are three estimation questions that we have to address here. The first is what tax rate to use in computing the tax benefit and whether than rate can change over time. The second is the dollar debt to use in computing the tax savings and whether that amount can vary across time. The final issue relates to what discount rate to use to compute the present value of the tax benefits. In the early iterations of APV, the tax rate and dollar debt were viewed as constants (resulting in tax savings as a perpetuity) and the pre-tax cost of debt was used as the discount rate leading to a simplification of the tax benefit value:

$$\begin{aligned} &= \frac{(\text{Tax Rate})(\text{Cost of Debt})(\text{Debt})}{\text{Cost of Debt}} \\ \text{Value of Tax Benefits} &= (\text{Tax Rate})(\text{Debt}) \\ &= t_c D \end{aligned}$$

Subsequent adaptations of the approach allowed for variations in both the tax rate and the dollar debt level, and raised questions about whether it was appropriate to use the cost of debt as the discount rate. Fernandez (2004) argued that the value of tax benefits should be computed as the difference between the value of the levered firm, with the interest tax savings, and the value of the same firm without leverage.[61] Consequently, he arrives at a much higher value for the tax savings than the conventional approach, by a multiple of the unlevered firm's cost of equity to the cost of debt. Cooper and Nyborg (2006) argue

[61] Fernandez, P., P., 2004, The value of tax shields is not equal to the present value of the tax shields, Journal of Financial Economics, v73, 145-165.

that Fernandez is wrong and that the value of the tax shield is the present value of the interest tax savings, discounted back at the cost of debt.[62]

The third step is to evaluate the effect of the given level of debt on the default risk of the firm and on expected bankruptcy costs. In theory, at least, this requires the estimation of the probability of default with the additional debt and the direct and indirect cost of bankruptcy. If π_a is the probability of default after the additional debt and BC is the present value of the bankruptcy cost, the present value of expected bankruptcy cost can be estimated.

$$\text{PV of Expected Bankruptcy cost} \begin{aligned} &= (\text{Probability of Bankruptcy})(\text{PV of Bankruptcy Cost}) \\ &= \pi_a BC \end{aligned}$$

This step of the adjusted present value approach poses the most significant estimation problem, since neither the probability of bankruptcy nor the bankruptcy cost can be estimated directly. There are two basic ways in which the probability of bankruptcy can be estimated indirectly. One is to estimate a bond rating, as we did in the cost of capital approach, at each level of debt and use the empirical estimates of default probabilities for each rating. The other is to use a statistical approach to estimate the probability of default, based upon the firm's observable characteristics, at each level of debt. The bankruptcy cost can be estimated, albeit with considerable error, from studies that have looked at the magnitude of this cost in actual bankruptcies. Research that has looked at the direct cost of bankruptcy concludes that they are small[63], relative to firm value. In fact, the costs of distress stretch far beyond the conventional costs of bankruptcy and liquidation. The perception of distress can do serious damage to a firm's operations, as employees, customers, suppliers and lenders react. Firms that are viewed as distressed lose customers (and sales), have higher employee turnover and have to accept much tighter restrictions from suppliers than healthy firms. These indirect bankruptcy costs can be catastrophic for many firms and essentially make the perception of distress into a

---

[62] Cooper, I.A. and K.G. Nyborg, 2006, The value of tax shields is equal to the present value of the tax shields, Journal of Financial Economics, v81, 215-225.

[63] Warner, J.N., 1977, Bankruptcy Costs: Some Evidence, Journal of Finance, v32, 337-347. In this study of railroad bankruptcies, the direct cost of bankruptcy was estimated to be about 5%.

reality. The magnitude of these costs has been examined in studies and can range from 10-25% of firm value.[64]

Variants on APV

While the original version of the adjusted present value model was fairly rigid in its treatment of the tax benefits of debt and expected bankruptcy costs, subsequent variations allow for more flexibility in the treatment of both. Some of these changes can be attributed to pragmatic considerations, primarily because of the absence of information, whereas others represented theoretical corrections.

One adaptation of the model was suggested by Luehrman (1997), where he presents an example where the dollar debt level, rather than remain fixed as it does in conventional APV, changes over time as a fraction of book value.[65] The interest tax savings reflect the changing debt but the present value of the tax savings is still computed using the cost of debt.

Another variation on adjusted present value was presented by Kaplan and Ruback (1995) in a paper where they compared the discounted cash flow valuations of companies to the prices paid in leveraged transactions.[66] They first estimated what they termed capital cash flows which they defined to be cash flows to both debt and equity investors and thus inclusive of the tax benefits from interest payments on debt. This is in contrast with the conventional unlevered firm valuation, which uses only operating cash flows and does not include interest tax savings. These capital cash flows are discounted back at the unlevered cost of equity to arrive at firm value. In effect, the compressed adjusted present value approach differs from the conventional adjusted present value approach on two dimensions. First, the tax savings from debt are discounted back at the unlevered cost of equity rather than the cost of debt. Second, the expected bankruptcy costs are effectively

---

[64] For an examination of the theory behind indirect bankruptcy costs, see Opler, T. and S. Titman, 1994, Financial Distress and Corporate Performance. Journal of Finance 49, 1015-1040. For an estimate on how large these indirect bankruptcy costs are in the real world, see Andrade, G. and S. Kaplan, 1998, How Costly is Financial (not Economic) Distress? Evidence from Highly Leveraged Transactions that Become Distressed. Journal of Finance. 53, 1443-1493. They look at highly levered transactions that subsequently became distressed snd conclude that the magnitude of these costs ranges from 10% to 23% of firm value.

[65] Luehrman, T. A., 1997, Using APV: A Better Tool for Valuing Operations, *Harvard Business Review*, May-June, 145-154.

ignored in the computation. Kaplan and Ruback argue that this approach is simpler to use than the conventional cost of capital approach in levered transactions because the leverage changes over time, which will result in time-varying costs of capital. In effect, they are arguing that it is easier to reflect the effects of changing leverage in the cash flows than it is in debt ratios. Gilson, Hotchkiss and Ruback (2000) use the compressed APV approach to value bankrupt firms that are reorganized and conclude that while the approach yields unbiased estimates of value, the valuation errors remain large.[67] The key limitation of the compressed APV approach, notwithstanding its simplicity, is that it ignores expected bankruptcy costs. In fact, using the compressed adjusted present value approach will lead to the conclusion that a firm is always worth more with a higher debt ratio than with a lower one. Kaplan and Ruback justify their approach by noting that the values that they arrive at are very similar to the values obtained using comparable firms, but this cannot be viewed as vindication.

Ruback (2000) provides a more extensive justification of the capital cash flow approach to valuation.[68] He notes that the conventional APV's assumption that interest tax savings have the same risk as the debt (and thus get discounted back at the cost of debt) may be justifiable for a fixed dollar debt but that it is more reasonable to assume that interest tax savings share the same risk as the operating assets, when dollar debt is expected to change over time. He also notes that the capital cash flow approach assumes that debt grows with firm value and is thus closer to the cost of capital approach, where free cash flows to the firm are discounted back at a cost of capital. In fact, he shows that when the dollar debt raised each year is such that the debt ratio stays constant, the cost of capital approach and the capital cash flows approach yield identical results.

---

[66] Kaplan, S.N. and R.S. Ruback, 1995, The Valuation of Cash Flow Forecasts, Journal of Finance, v50, 1059-1093.

[67] Gilson, S.C., E. S. Hotchkiss and R. Ruback, 1998, Valuation of Bankrupt Firms, Review of Financial Studies, v13, 43-74. The one modification they introduce is that the tax savings from net operating loss carryforwards are discounted back at the cost of debt.

[68] Ruback, R.S., 2000, Capital Cash Flows: A Simple Approach to Valuing Risky Cash Flows, Working Paper, Harvard Business School.

Cost of Capital versus APV Valuation

To understand when the cost of capital approach, the adjusted present value approach and the modified adjusted present value approach (with capital cash flows) yield similar and different results, we consider the mechanics of each approach in table 1:

Table 1: Cost of Capital, APV and Compressed APV

	Cost of Capital	*Conventional APV*	*Compressed APV*
Cash flow discounted	Free Cash Flow to Firm (prior to all debt payments)	Free Cash Flow to Firm (prior to debt payments)	Free Cash Flow to Firm + Tax Savings from Interest Payments
Discount Rate used	Weighted average of cost of equity and after-tax cost of debt = Cost of capital	Unlevered cost of equity	Weighted average of cost of equity and pre-tax cost of debt = Unlevered cost of equity
Tax Savings from Debt	Shows up through the discount rate	Added on separately as present value of tax savings (using cost of debt as discount rate)	Shows up through cash flow
Dollar debt levels	Determined by debt ratios used in cost of capital. If debt ratio stays fixed, dollar debt increases with firm value	Fixed dollar debt	Dollar debt can change over time – increase or decrease.
Discount rate for tax benefits from interest expenses	Discounted at unlevered cost of equity	Discounted at pre-tax cost of debt	Discounted at unlevered cost of equity
Bankruptcy Costs	Reflected as higher costs of equity and debt, as default risk increases.	Can be computed separately, based upon likelihood of distress and the cost of such distress. (In practice, often ignored)	Can be computed separately, based upon likelihood of distress and the cost of such distress. (In practice, often ignored)

In an APV valuation, the value of a levered firm is obtained by adding the net effect of debt to the unlevered firm value.

$$\text{Value of Levered Firm} = \frac{\text{FCFF}_o(1+g)}{\rho_u - g} + t_c D - \pi_a BC$$

The tax savings from debt are discounted back at the cost of debt. In the cost of capital approach, the effects of leverage show up in the cost of capital, with the tax benefit incorporated in the after-tax cost of debt and the bankruptcy costs in both the levered beta and the pre-tax cost of debt. Inselbag and Kaufold (1997) provide examples where they get identical values using the APV and Cost of Capital approaches, but only because they infer the costs of equity to use in the latter.[69]

Will the approaches yield the same value? Not necessarily. The first reason for the differences is that the models consider bankruptcy costs very differently, with the adjusted present value approach providing more flexibility in allowing you to consider indirect bankruptcy costs. To the extent that these costs do not show up or show up inadequately in the pre-tax cost of debt, the APV approach will yield a more conservative estimate of value. The second reason is that the conventional APV approach considers the tax benefit from a fixed dollar debt value, usually based upon existing debt. The cost of capital and compressed APV approaches estimate the tax benefit from a debt ratio that may require the firm to borrow increasing amounts in the future. For instance, assuming a market debt to capital ratio of 30% in perpetuity for a growing firm will require it to borrow more in the future and the tax benefit from expected future borrowings is incorporated into value today. Finally, the discount rate used to compute the present value of tax benefits is the pre-tax cost of debt in the conventional APV approach and the unlevered cost of equity in the compressed APV and the cost of capital approaches. As we noted earlier, the compressed APV approach yields equivalent values to the cost of capital approach, if we allow dollar debt to reflect changing firm value (and debt ratio assumptions) and ignore the effect of indirect bankruptcy costs. The conventional APV approach yields a higher value than either of the other two approaches because it views the tax savings from debt as less risky and assigns a higher value to it.

Which approach will yield more reasonable estimates of value? The dollar debt assumption in the APV approach is a more conservative one but the fundamental flaw with the APV model lies in the difficulties associated with estimating expected bankruptcy costs. As long as that cost cannot be estimated, the APV approach will

---

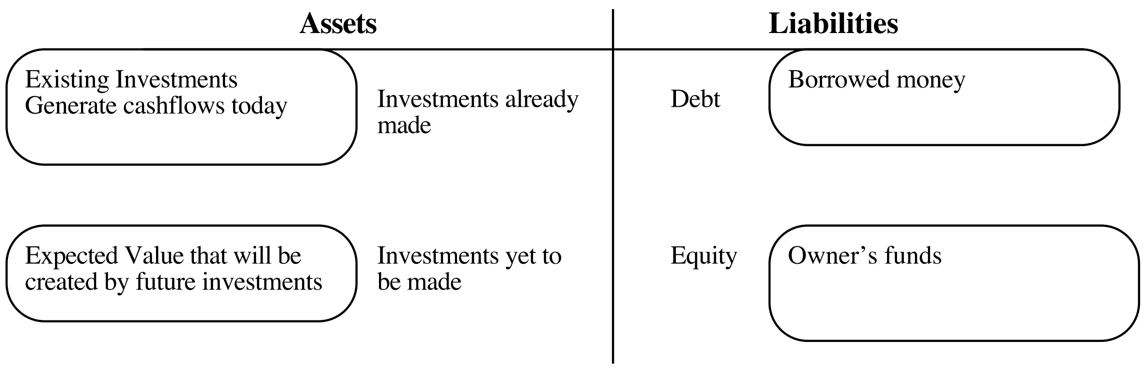
[69] Inselbag, I. and H. Kaufold, 1997, Two DCF approaches for valuing companies under alternative financing strategies and how to choose between them, Journal of Applied Corporate Finance, v10, 114-122.

continue to be used in half-baked form where the present value of tax benefits will be added to the unlevered firm value to arrive at total firm value.

Liquidation and Accounting Valuation

The value of an asset in the discounted cash flow framework is the present value of the expected cash flows on that asset. Extending this proposition to valuing a business, it can be argued that the value of a business is the sum of the values of the individual assets owned by the business. While this may be technically right, there is a key difference between valuing a collection of assets and a business. A business or a company is an on-going entity with assets that it already owns and assets it expects to invest in the future. This can be best seen when we look at the financial balance sheet (as opposed to an accounting balance sheet) for an ongoing company in figure 4:

Figure 4: A Simple View of a Firm



Note that investments that have already been made are categorized as assets in place, but investments that we expect the business to make in the future are growth assets.

A financial balance sheet provides a good framework to draw out the differences between valuing a business as a going concern and valuing it as a collection of assets. In a going concern valuation, we have to make our best judgments not only on existing investments but also on expected future investments and their profitability. While this may seem to be foolhardy, a large proportion of the market value of growth companies comes from their growth assets. In an asset-based valuation, we focus primarily on the assets in place and estimate the value of each asset separately. Adding the asset values together yields the value of the business. For companies with lucrative growth opportunities, asset-based valuations will yield lower values than going concern valuations.

Book Value Based Valuation

There are some who contend that the accounting estimate of the value of a business, as embodied by the book value of the assets and equity on a balance sheet, represents a more reliable estimate of value than valuation models based on shaky assumptions about the future. In this section, we examine book value as a measure of the value of going concern and then extend the analysis to look at book value based valuation models that are also use forecasted earnings to estimate value. We end the section with a short discussion of fair value accounting, a movement that has acquired momentum in recent years.

Book Value

The original ideals for accounting statements were that the income statements would provide a measure of the true earnings potential of a firm and that the balance sheet would yield a reliable estimate of the value of the assets and equity in the firm. Daniels (1934), for instance, lays out these ideals thus:[70]

> *"In short the lay reader of financial statements usually believes that the total asset figure of the balance sheet is indicative, and is intended to be so, of the value of the company. He probably understanding this "value" as what the business could be sold for, market value – the classic meeting of the minds between a willing buyer and seller."*

In the years since, accountants have wrestled with how put this ideal into practice. In the process, they have had the weigh how much importance to give the historical cost of an asset relative to its estimated value today and have settled on different rules. For fixed assets, they have largely concluded that the book value should be reflective of the original cost of the asset and subsequent depletion in and additions to that asset. For current assets, they have been much more willing to consider the alternative of market value. Finally, they have discovered new categories for assets such as brand name where neither the original cost nor the current value is easily accessible.

While there are few accountants who would still contend that the book value of a company is a good measure of its market value, this has not stopped some investors from implicitly making that assumption. In fact, the notion that a stock is under valued if is market price falls below its book value is deeply entrenched in investing. It is one of the screens that Ben Graham proposed for finding undervalued stocks[71] and it remains a rough proxy for what is loosely called value investing.[72] Academics have fed into this belief by presenting evidence that low price to book value stocks do earn higher returns than the rest of the market.[73]

Is it possible for book value to be a reasonable proxy for the true value of a business? For mature firms with predominantly fixed assets, little or no growth opportunities and no potential for excess returns, the book value of the assets may yield a reasonable measure of the true value of these firms. For firms with significant growth opportunities in businesses where they can generate excess returns, book values will be very different from true value.

Book Value plus Earnings

In the context of equity valuation models, we considered earnings based models that have been developed in recent years, primarily in the accounting community. Most of these models are built on a combination of book values and expected future earnings and trace their antecedents to Ohlson (1995) and Feltham and Ohlson (1995), both works that we referenced earlier in the context of earnings based valuation models.[74] Ohlson's basic model states the true value of equity as a function of its book value of equity and the excess equity returns that the firm can generate in the future. As a consequence, it is termed a residual income model and can be derived from a simple dividend discount model:

$$\text{Value of equity} = \sum_{t=1}^{t=\infty} \frac{\text{E(Dividends}_t)}{(1+\ \text{Cost of Equity})^t}$$

---

[70] Daniels, M.B., 1934, Principles of Asset Valuation, The Accounting Review, v9, 114-121.
[71] Graham, B., 1949, The Intelligent Investor, HarperCollins,
[72] Morningstar categorizes mutual funds into growth and value, based upon the types of stocks that they invest in. Funds that invest in low price to book stocks are categorized as value funds.
[73] Fama, E.F. and K.R. French, 1992, *The Cross-Section of Expected Returns*, Journal of Finance, v47, 427-466.
[74] Ohlson, J. 1995, Earnings, Book values and Dividends in Security Valuation, Contemporary Accounting Research, v11, 661-687.; Feltham and Ohlson, 1995, Valuation and Clean Surplus Accounting for Operating and Financial Activities, Contemporary Accounting Research, v11, 689-731.

Now substitute in the full equation for book value (BV) of equity as a function of the starting book equity and earnings and dividends during a period (clean surplus relationship):

$$\text{Book Value of Equity}_t = \text{BV of Equity}_{t-1} + \text{Net Income}_t - \text{Dividends}_t$$

Substituting back into the dividend discount model, we get

$$\text{Value of Equity}_0 = \text{BV of Equity}_0 + \frac{\sum_{t=1}^{t=\infty} (\text{Net Income}_t - \text{Cost of Equity}_t * \text{BV of Equity}_{t-1})}{(1 + \text{Cost of Equity}_t)^t}$$

Thus the value of equity in a firm is the sum of the current book value of equity and the present value of the expected excess returns to equity investors in perpetuity

The enthusiasm with which the Ohlson residual income model has been received by accounting researchers is puzzling, given that it is neither new nor revolutionary. Walter(1966)[75] and Mao (1974)[76] extended the dividend discount model to incorporate excess returns earned on future investment opportunities. In fact, we used exactly the same rationale to relate enterprise value to EVA earlier in the paper. The only real difference is that the Ohlson model is an extension of the more limiting dividend discount model, whereas the EVA model is an extension of a more general firm valuation model. In fact, Lundholm and O'Keefe (2001) show that discounted cash flow models and residual income models yield identical valuations of companies, if we make consistent assumptions.[77] One explanation for the enthusiasm is that the Ohlson model has allowed accountants to argue that accounting numbers are still relevant to value. After all, Lev (1989) had presented evidence on the declining significance of accounting earnings

---

[7575] Walter, J.E., 1966, Dividend Policies and Common Stock Prices, Journal of Finance, v11, 29-41. Walters modified the dividend discount model as follows: $P = \frac{D + \frac{ROE}{k_e}(E - D)}{k_e}$, where E and D are the expected earnings and dividends in the next period, ROE is the expected return on equity in perpetuity on retained earnings and k_e is the cost of equity. Note that the second term in the numerator is the excess return generated on an annual basis and that dividing by the cost of equity yields its present value in perpetuity.

[76] Mao, J.C.T., 1974, The Valuation of Growth Stocks: The Investment Opportunities Approach, Journal of Finance, v21, 95-102. The key difference is that rather than build off book value of equity, as Ohlson did, Mao capitalized current earnings (as a perpetuity) and added the present value of future excess returns to this value.

[77] Lundholm, R., and T. O'Keefe. 2001. Reconciling value estimates from the discounted cash flow model and the residual income model. *Contemporary Accounting Research, v*18, 311-35.

numbers by noting a drop in the correlation between market value and earnings.[78] In the years since, a number of studies have claimed to find strong evidence to back up the Ohlson model. For instance, Frankel and Lee (1996)[79], Hand and Landsman (1998)[80] and Dechow, Hutton and Sloan (1999)[81] all find that the residual income model explains 70-80% of variation in prices across stocks. The high R-squared in these studies is deceptive since they are not testing an equation as much as a truism: the total market value of equity should be highly correlated with the total book value of equity and total net income. Firms with higher market capitalization will tend to have higher book value of equity and higher net income, reflecting their scale and this has little relevance for whether the Ohlson model actually works.[82] A far stronger and more effective test of the model is whether changes in equity value are correlated with changes in book value of equity and net income and the model does no better on these tests than established models.

Fair Value Accounting

In the last decade, there has been a strong push from both accounting rule makers and regulators towards "fair value accounting". Presumably, the impetus for this push has been a return to the original ideal that the book value of the assets on a balance sheet and the resulting net worth for companies be good measures of the fair value of these assets and equity.

The move towards fair value accounting has not been universally welcomed even within the accounting community. On the one hand, there are some who believe that this is a positive development increasing the connection of accounting statements to value and

---

[78] Lev, B. 1989. On the usefulness of earnings: Lessons and directions from two decades of empirical research, Journal of Accounting Research, v
27 (Supplement): 153-192.
[79] Frankel, R. and C.M.C. Lee. 1998. Accounting Valuation, Market Expectations, and Crosssectional Stock Returns. Journal of Accounting Economics, v25: 283-319.
[80] Hand, J.R.M. and W.R. Landsman. 1999. Testing the Ohlson Model: v or not v, that is the Question. Working Paper, University of North Carolina at Chapel Hill.
[81] Dechow, P., A. Hutton, R. Sloan, 1999. An Empirical Assessment of the Residual Income Valuation Model. Journal of Accounting and Economics 26 (1-3)1-34.
[82] Lo, K. and Lys, T., 2005, The Ohlson Model: Contribution to Valuation Theory, Limitations and Empirical Applications, Working Paper, Kellogg School of Management, Northwestern University.

providing useful information to financial markets.[83] There are others who believe that fair value accounting increases the potential for accounting manipulation, and that financial statements will become less informative as a result.[84] In fact, it used to be common place for firms in the United States to revalue their assets at fair market value until 1934, and the SEC discouraged this practice after 1934 to prevent the widespread manipulation that was prevalent.[85] While this debate rages on, the accounting standards boards have adopted a number of rules that favor fair value accounting, from the elimination of purchase accounting in acquisitions to the requirement that more assets be marked to market on the balance sheet.

The question then becomes an empirical one. Do fair value judgments made by accountants provide information to financial markets or do they just muddy up the waters? In a series of articles, Barth concluded that fair value accounting provided useful information to markets in a variety of contexts.[86] In contrast, Nelson (1996) examines fair value accounting in banking, where marking to market has been convention for a much longer period, and finds the reported fair values of investment securities have little incremental explanatory power when it comes to market values.[87] In an interesting test of the effects of fair value accounting, researchers have begun looking at market reactions in the aftermath of the adoption of SFAS 141 and 142, which together eliminated pooling, while also requiring that firms estimate "fair-value impairments" of goodwill rather than amortizing goodwill. Chen, Kohlbeck and Warfield (2004) find that stock prices react negatively to goodwill impairments, which they construe to indicate that there is

---

[83] Barth, M., W. Beaver and W. Landsman. 2001. The relevance of the value-relevance literature for financial accounting standard setting: another view. *Journal of Accounting and Economics* 31: 77-104

[84] Holthausen, R. and R. Watts. 2001. The relevance of the value-relevance literature for financial accounting standard setting. Journal of Accounting and Economics, v31, 3-75.

[85] Fabricant, S. 1938. Capital Consumption and Adjustment, National Bureau of Economic Research.

[86] Barth, M.E., 1994. Fair Value Accounting: Evidence from Investment Securities and theMarket Valuation of Banks, *Accounting Review*, v69, No. 1 (January): 1–25; Barth, M.E., W. R. Landsman, and J. M. Whalen. 1995. Fair value accounting: effects on banks' earnings volatility, regulatory capital, and value of contractual cash flows, *Journal of Banking and Finance* v19, No.3-4 (June): 577–605; Barth, M.E., W.H. Beaver, and W.R. Landsman. 1996. Value relevance of banks fair value disclosures under SFAS 107, *The Accounting Review,* v71, No.4 (October): 513–37; Barth, M.E. and G. Clinch. 1998. Revalued financial, tangible, and intangible assets: Associations with share prices and non-market-based value estimates, *Journal of Accounting Research*, v36 (Supplement): 199–233.

[87] Nelson, K.K., 1996, Fair Value Accounting for Commercial Banks: An Empirical Analysis of SFAS 107, The Accounting Review, v71, 161-182.

information in these accounting assessments.[88] Note, though, that this price reaction can be consistent with a number of other interpretations as well and can be regarded, at best, as weak evidence that fair value accounting assessments convey information to markets.

We believe that fair value accounting, at best, will provide a delayed reflection of what happens in the market. In other words, goodwill be impaired (as it was in many technology companies in 2000 and 2001) after the market value has dropped and fair value adjustments will convey little, if any, information to financial markets. If in the process of marking to market, some of the raw data that is now provided to investors is replaced or held back, we will end up with accounting statements that neither reflect market value nor invested capital.

Liquidation Valuation

One special case of asset-based valuation is liquidation valuation, where we value assets based upon the presumption that they have to be sold now. In theory, this should be equal to the value obtained from discounted cash flow valuations of individual assets but the urgency associated with liquidating assets quickly may result in a discount on the value. The magnitude of the discount will depend upon the number of potential buyers for the assets, the asset characteristics and the state of the economy.

The research on liquidation value can be categorized into two groups. The first group of studies examines the relationship between liquidation value and the book value of assets, whereas the second takes apart the deviations of liquidation value from discounted cash flow value and addresses directly the question of how much of a cost you bear when you have to liquidate assets rather than sell a going concern.

While it may seem naïve to assume that liquidation value is equal or close to book value, a number of liquidation rules of thumb are structured around book value. For instance, it is not uncommon to see analysts assume that liquidation value will be a specified percentage of book value. Berger, Ofek and Swary (1996) argue and provide evidence that book value operates as a proxy for abandonment value in many firms.[89]

---

[88] Chen, C., M. Kohlbeck and T. Warfield, 2004, Goodwill Valuation Effects of the Initial Adoption of SFAS 142, Working Paper, University of Wisconsin- Madison.

[89] Berger, P., E. Ofek and I. Swary, 1996, Investor Valuation of the Abandonment Option, Journal of Financial Economics, v42, 257-287.

Lang, Stulz and Walkling (1989) use book value as a proxy for the replacement cost of assets when computing Tobin's Q.[90]

The relationship between liquidation and discounted cash flow value is more difficult to discern. It stands to reason that liquidation value should be significantly lower than discounted cash flow value, partly because the latter reflects the value of expected growth potential and the former usually does not. In addition, the urgency associated with the liquidation can have an impact on the proceeds, since the discount on value can be considerable for those sellers who are eager to divest their assets. Kaplan (1989) cited a Merrill Lynch estimate that the speedy sales of the Campeau stake in Federated would bring about 32% less than an orderly sale of the same assets.[91] Holland (1990) estimates the discount to be greater than 50% in the liquidation of the assets of machine tool manufacturer.[92] Williamson (1988) makes the very legitimate point that the extent of the discount is likely to be smaller for assets that are not specialized and can be redeployed elsewhere.[93] Shleifer and Vishny (1992) argue that assets with few potential buyers or buyers who are financially constrained are likely to sell at significant discounts on market value.[94]

In summary, liquidation valuation is likely to yield more realistic estimates of value for firms that are distressed, where the going concern assumption underlying conventional discounted cash flow valuation is clearly violated. For healthy firms with significant growth opportunities, it will provide estimates of value that are far too conservative.

Relative Valuation

In relative valuation, we value an asset based upon how similar assets are priced in the market. A prospective house buyer decides how much to pay for a house by looking at the prices paid for similar houses in the neighborhood. A baseball card

---

90 Lang, L.H.P., R.M. Stulz and R.Walking. 1989. Managerial Performance, Tobin's Q, and The Gains from Successful Tender Offers. Journal of Financial Economics, v29, 137-154.

91 Kaplan, S.N., 1989, Campeau's Acquisition of Federated: Value Destroyed or Value Added? Journal of Financial Economics, v25, 191-212.

92 Holland, M., 1990, When the Machine Stopped, Harvard Business School Press, Cambridge, MA.

93 Williamson, O.E., 1988, Corporate Finance and Corporate Governance, Journal of Finance, v43, 567-592.

94 Shleifer, A., and R. W. Vishny, 1992, Liquidation Values and Debt Capacity: A Market Equilibrium

collector makes a judgment on how much to pay for a Mickey Mantle rookie card by checking transactions prices on other Mickey Mantle rookie cards. In the same vein, a potential investor in a stock tries to estimate its value by looking at the market pricing of "similar" stocks.

Embedded in this description are the three essential steps in relative valuation. The first step is finding comparable assets that are priced by the market, a task that is easier to accomplish with real assets like baseball cards and houses than it is with stocks. All too often, analysts use other companies in the same sector as comparable, comparing a software firm to other software firms or a utility to other utilities, but we will question whether this practice really yields similar companies later in this paper. The second step is scaling the market prices to a common variable to generate standardized prices that are comparable. While this may not be necessary when comparing identical assets (Mickey Mantle rookie cards), it is necessary when comparing assets that vary in size or units. Other things remaining equal, a smaller house or apartment should trade at a lower price than a larger residence. In the context of stocks, this equalization usually requires converting the market value of equity or the firm into multiples of earnings, book value or revenues. The third and last step in the process is adjusting for differences across assets when comparing their standardized values. Again, using the example of a house, a newer house with more updated amenities should be priced higher than a similar sized older house that needs renovation. With stocks, differences in pricing across stocks can be attributed to all of the fundamentals that we talked about in discounted cash flow valuation. Higher growth companies, for instance, should trade at higher multiples than lower growth companies in the same sector. Many analysts adjust for these differences qualitatively, making every relative valuation a story telling experience; analysts with better and more believable stories are given credit for better valuations.

Basis for approach

There is a significant philosophical difference between discounted cash flow and relative valuation. In discounted cash flow valuation, we are attempting to estimate the intrinsic value of an asset based upon its capacity to generate cash flows in the future. In

---

Approach, Journal of Finance, v47, 143-66.

relative valuation, we are making a judgment on how much an asset is worth by looking at what the market is paying for similar assets. If the market is correct, on average, in the way it prices assets, discounted cash flow and relative valuations may converge. If, however, the market is systematically over pricing or under pricing a group of assets or an entire sector, discounted cash flow valuations can deviate from relative valuations.

Harking back to our earlier discussion of discounted cash flow valuation, we argued that discounted cash flow valuation was a search (albeit unfulfilled) for intrinsic value. In relative valuation, we have given up on estimating intrinsic value and essentially put our trust in markets getting it right, at least on average. It can be argued that most valuations are relative valuations. Damodaran (2002) notes that almost 90% of equity research valuations and 50% of acquisition valuations use some combination of multiples and comparable companies and are thus relative valuations.[95]

Standardized Values and Multiples

When comparing identical assets, we can compare the prices of these assets. Thus, the price of a Tiffany lamp or a Mickey Mantle rookie card can be compared to the price at which an identical item was bought or sold in the market. However, comparing assets that are not exactly similar can be a challenge. After all, the price per share of a stock is a function both of the value of the equity in a company and the number of shares outstanding in the firm. Thus, a stock split that doubles the number of units will approximately halve the stock price. To compare the values of "similar" firms in the market, we need to standardize the values in some way by scaling them to a common variable. In general, values can be standardized relative to the earnings firms generate, to the book values or replacement values of the firms themselves, to the revenues that firms generate or to measures that are specific to firms in a sector.

- One of the more intuitive ways to think of the value of any asset is as a multiple of the earnings that asset generates. When buying a stock, it is common to look at the price paid as a multiple of the earnings per share generated by the company. This price/earnings ratio can be estimated using current earnings per share, yielding a current PE, earnings over the last 4 quarters, resulting in a trailing PE, or an expected earnings per share in the next year, providing a forward PE. When buying a business, as opposed to just the equity in the business, it is common to examine the value of the firm as a multiple of the operating income or the earnings before interest, taxes, depreciation and amortization (EBITDA). While, as a buyer of the equity or the firm, a lower multiple is better than a higher one, these multiples will be affected by the growth potential and risk of the business being acquired.

- While financial markets provide one estimate of the value of a business, accountants often provide a very different estimate of value of for the same business. As we noted earlier, investors often look at the relationship between the price they pay for a stock and the book value of equity (or net worth) as a measure of how over- or undervalued a stock is; the price/book value ratio that emerges can vary widely across industries, depending again upon the growth potential and the quality of the investments in each. When valuing businesses, we estimate this ratio using the value of the firm and the book value of all assets or capital (rather than just the equity). For those who believe that book value is not a good measure of the true value of the assets, an alternative is to use the replacement cost of the assets; the ratio of the value of the firm to replacement cost is called Tobin's Q.
- Both earnings and book value are accounting measures and are determined by accounting rules and principles. An alternative approach, which is far less affected by accounting choices, is to use the ratio of the value of a business to the revenues it generates. For equity investors, this ratio is the price/sales ratio (PS), where the market value of equity is divided by the revenues generated by the firm. For firm value, this ratio can be modified as the enterprise value/to sales ratio (VS), where the numerator becomes the market value of the operating assets of the firm. This ratio, again, varies widely across sectors, largely as a function of the profit margins in each. The advantage of using revenue multiples, however, is that it becomes far easier to compare firms in different markets, with different accounting systems at work, than it is to compare earnings or book value multiples.
- While earnings, book value and revenue multiples are multiples that can be computed for firms in any sector and across the entire market, there are some multiples that are specific to a sector. For instance, when internet firms first appeared on the market in the later 1990s, they had negative earnings and negligible revenues and book value.

---

[95] Damodaran, A., 2002, Investment Valuation (Second Edition), John Wiley and Sons, New York.

Analysts looking for a multiple to value these firms divided the market value of each of these firms by the number of hits generated by that firm's web site. Firms with lower market value per customer hit were viewed as under valued. More recently, cable companies have been judged by the market value per cable subscriber, regardless of the longevity and the profitably of having these subscribers. While there are conditions under which sector-specific multiples can be justified, they are dangerous for two reasons. First, since they cannot be computed for other sectors or for the entire market, sector-specific multiples can result in persistent over or under valuations of sectors relative to the rest of the market. Thus, investors who would never consider paying 80 times revenues for a firm might not have the same qualms about paying $2000 for every page hit (on the web site), largely because they have no sense of what high, low or average is on this measure. Second, it is far more difficult to relate sector specific multiples to fundamentals, which is an essential ingredient to using multiples well. For instance, does a visitor to a company's web site translate into higher revenues and profits? The answer will not only vary from company to company, but will also be difficult to estimate looking forward.

There have been relatively few studies that document the usage statistics on these multiples and compare their relative efficacy. Damodaran (2002) notes that the usage of multiples varies widely across sectors, with Enterprise Value/EBITDA multiples dominating valuations of heavy infrastructure businesses (cable, telecomm) and price to book ratios common in financial service company valuations.[96] Fernandez (2001) presents evidence on the relative popularity of different multiples at the research arm of one investment bank – Morgan Stanley Europe – and notes that PE ratios and EV/EBITDA multiples are the most frequently employed.[97] Liu, Nissim and Thomas (2002) compare how well different multiples do in pricing 19,879 firm-year observations between 1982 and 1999 and suggest that multiples of forecasted earnings per share do best in explaining pricing differences, that multiples of sales and operating cash flows do

[96] Damodaran, A, 2002, Investment Valuation, Second Edition, John Wiley and Sons, New York.

[97] Fernandez, P., 2001, Valuation using multiples. How do analysts reach their conclusions?, Working Paper, IESE Business School.

worst and that multiples of book value and EBITDA fall in the middle.[98] Lie and Lie (2002) examine 10 different multiples across 8,621 companies between 1998 and 1999 and arrive at similar conclusions.[99]

Determinants of Multiples

In the introduction to discounted cash flow valuation, we observed that the value of a firm is a function of three variables – it capacity to generate cash flows, the expected growth in these cash flows and the uncertainty associated with these cash flows. Every multiple, whether it is of earnings, revenues or book value, is a function of the same three variables – risk, growth and cash flow generating potential. Intuitively, then, firms with higher growth rates, less risk and greater cash flow generating potential should trade at higher multiples than firms with lower growth, higher risk and less cash flow potential.

The specific measures of growth, risk and cash flow generating potential that are used will vary from multiple to multiple. To look under the hood, so to speak, of equity and firm value multiples, we can go back to fairly simple discounted cash flow models for equity and firm value and use them to derive the multiples. In the simplest discounted cash flow model for equity, which is a stable growth dividend discount model, the value of equity is:

$$\text{Value of Equity} = P_0 = \frac{DPS_1}{k_e - g_n}$$

where DPS_1 is the expected dividend in the next year, k_e is the cost of equity and g_n is the expected stable growth rate. Dividing both sides by the earnings, we obtain the discounted cash flow equation specifying the PE ratio for a stable growth firm.

$$\frac{P_0}{EPS_0} = PE = \frac{\text{Payout Ratio} * (1 + g_n)}{k_e - g_n}$$

The key determinants of the PE ratio are the expected growth rate in earnings per share, the cost of equity and the payout ratio. Other things remaining equal, we would expect higher growth, lower risk and higher payout ratio firms to trade at higher multiples of earnings than firms without these characteristics. In fact, this model can be expanded to

---

[98] Liu, J., D. Nissim, and J. Thomas. 2002. Equity Valuation Using Multiples. *Journal of Accounting Research*, V 40, 135-172.

[99] Lie E., H.J. Lie, 2002, Multiples Used to Estimate Corporate Value. Financial Analysts Journal, v58, 44-54.

allow for high growth in near years and stable growth beyond.[100] Researchers have long recognized that the PE for a stock is a function of both the level and the quality of its growth and its risk. Beaver and Morse (1978) related PE ratios to valuation fundamentals[101], as did earlier work by Edwards and Bell (1961).[102] Peasnell (1982) made a more explicit attempt to connect market values to accounting numbers.[103] Zarowin (1990) looked at the link between PE ratios and analyst forecasts of growth to conclude that PE ratios are indeed positively related to long term expected growth.[104] Leibowitz and Kogelman (1990, 1991, 1992) expanded on the relationship between PE ratios and the excess returns earned on investments, which they titled franchise opportunities, in a series of articles on the topic, noting that for a stock to have a high PE ratio, it needs to generate high growth in conjunction with excess returns on its new investments.[105] Fairfield (1994) provides a generalized version of their model, allowing for changing return on equity over time.[106] While these papers focused primarily on growth and returns, Kane, Marcus and Noe (1996) examine the relationship between PE and risk for the aggregate market and conclude that PE ratios decrease as market volatility increases.[107]

Dividing both sides of the stable growth dividend discount model by the book value of equity, we can estimate the price/book value ratio for a stable growth firm.

$$\frac{P_0}{BV_0} = PBV = \frac{ROE * \text{Payout Ratio} * (1 + g_n)}{k_e - g_n}$$

---

[100] Damodaran, A., 2002, Investment Valuation, John Wiley and Sons, New York. The expanded versions of the models are available in the chapter on PE ratios.

[101] Beaver, W. and D. Morse, 1978, What do P/E ratios mean?, Financial Analysts Journal, v34, 65-76.

[102] Edwards, E. and P. Bell, 1961, The Theory and Measurement of Business Income, University of California Press, Berkeley.

[103] Peasnell, K., 1982, Some Financial Connections between Economic Values and Accounting Numbers, Journal of Business Finance and Accounting, v9, 361-381.

[104] Zarowin, P. 1990. What determines earnings-price ratios: revisited, Journal of Accounting, Auditing, and Finance, v5: 439-57.

[105] Leibowitz, M.L. and S. Kogelman, 1990, Inside the PE Ratio: The Franchise Factor, Financial Analysts Journal, v46, 17-35; Leibowitz, M.L. and S. Kogelman, 1991, The Franchise Factor for Leveraged Firms, Financial Analysts Journal, v47, 29-43.; Leibowitz, M.L. and S. Kogelman, 1992, Franchise Value and the Growth Factor, Financial Analysts Journal, v48, 16-23.

[106] Fairfield, P., 1994, P/E, P/B and the present value of future dividends, Financial Analysts Journal, v50, 23-31.

[107] Kane, A., A.J. Marcus and J. Noh, The P/E Multiple and Market Volatility, Financial Analysts Journal, v52, 16-24.

where ROE is the return on equity and is the only variable in addition to the three that determine PE ratios (growth rate, cost of equity and payout) that affects price to book equity. The strong connection between price to book and return on equity was noted by Wilcox (1984), with his argument that cheap stocks are those that trade at low price to book ratios while maintaining reasonable or even high returns on equity.[108] The papers we referenced in the earlier section on book-value based valuation approaches centered on the Ohlson model can be reframed as a discussion of the determinants of price to book ratios. Penman (1996) draws a distinction between PE ratios and PBV ratios when it comes to the link with return on equity, by noting that while PBV ratios increase with ROE, the relationship between PE ratios and ROE is weaker.[109]

Finally, dividing both sides of the dividend discount model by revenues per share, the price/sales ratio for a stable growth firm can be estimated as a function of its profit margin, payout ratio, risk and expected growth.

$$\frac{P_0}{\text{Sales}_0} = \text{PS} = \frac{\text{Profit Margin} * \text{Payout Ratio} * (1 + g_n)}{k_e - g_n}$$

The net margin is the new variable that is added to the process. While all of these computations are based upon a stable growth dividend discount model, we will show that the conclusions hold even when we look at companies with high growth potential and with other equity valuation models. While less work has been done on revenue multiples than on book value or earnings multiples, Leibowitz (1997) extends his franchise value argument from PE ratios to revenue multiples and notes the importance of what profit margins.[110]

We can do a similar analysis to derive the firm value multiples. The value of a firm in stable growth can be written as:

$$\text{Value of Firm} = V_0 = \frac{FCFF_1}{k_c - g_n}$$

[108] Wilcox, J., 1984, The P/B-ROE Valuation Model, Financial Analysts Journal, 58-66.

[109] Penman, S.H., 1996, The Articulation of Price-Earnings and Market-to-Book Ratios and the Evaluation of Growth, Journal of Accounting Research, v34, 235-259.

[110] Leibowitz, M.L., 1997, Franchise Margins and the Sales-Driven Franchise Value, Financial Analysts Journal, v53, 43-53.

Dividing both sides by the expected free cash flow to the firm yields the Value/FCFF multiple for a stable growth firm.

$$\frac{V_0}{FCFF_1} = \frac{1}{k_c - g_n}$$

The multiple of FCFF that a firm commands will depend upon two variables – its cost of capital and its expected stable growth rate. Since the free cash flow the firm is the after-tax operating income netted against the net capital expenditures and working capital needs of the firm, the multiples of EBIT, after-tax EBIT and EBITDA can also be estimated similarly.

In short, multiples are determined by the same variables and assumptions that underlie discounted cash flow valuation. The difference is that while the assumptions are explicit in the latter, they are often implicit in the use of the former.

Comparable Firms

When multiples are used, they tend to be used in conjunction with comparable firms to determine the value of a firm or its equity. But what is a comparable firm? A comparable firm is one with cash flows, growth potential, and risk similar to the firm being valued. It would be ideal if we could value a firm by looking at how an exactly identical firm - in terms of risk, growth and cash flows - is priced. Nowhere in this definition is there a component that relates to the industry or sector to which a firm belongs. Thus, a telecommunications firm can be compared to a software firm, if the two are identical in terms of cash flows, growth and risk. In most analyses, however, analysts define comparable firms to be other firms in the firm's business or businesses. If there are enough firms in the industry to allow for it, this list is pruned further using other criteria; for instance, only firms of similar size may be considered. The implicit assumption being made here is that firms in the same sector have similar risk, growth, and cash flow profiles and therefore can be compared with much more legitimacy. This approach becomes more difficult to apply when there are relatively few firms in a sector. In most markets outside the United States, the number of publicly traded firms in a particular sector, especially if it is defined narrowly, is small. It is also difficult to define firms in the same sector as comparable firms if differences in risk, growth and cash flow profiles across firms within a sector are large. The tradeoff is therefore a simple one. Defining an industry more broadly increases the number of comparable firms, but it also results in a more diverse group of companies. Boatman and Baskin (1981) compare the precision of PE ratio estimates that emerge from using a random sample from within the same sector and a narrower set of firms with the most similar 10-year average growth rate in earnings and conclude that the latter yields better estimates.[111]

There are alternatives to the conventional practice of defining comparable firms as other firms in the same industry. One is to look for firms that are similar in terms of valuation fundamentals. For instance, to estimate the value of a firm with a beta of 1.2, an expected growth rate in earnings per share of 20% and a return on equity of 40%[112], we would find other firms across the entire market with similar characteristics.[113] Alford (1992) examines the practice of using industry categorizations for comparable firms and compares their effectiveness with using categorizations based upon fundamentals such as risk and growth.[114] Based upon the prediction error from the use of each categorization, he concludes that industry based categorizations match or slightly outperform fundamental based categorization, which he views as evidence that much of the variation in multiples that can be explained by fundamentals can be also explained by industry. In contrast, Cheng and McNamara (2000) and Bhojraj and Lee (2002) argue that picking comparables using a combination of industry categorization and fundamentals such as total assets yields more precise valuations than using just the industry classification.[115]

---

111 Boatman, J.R. and E.F. Baskin, 1981, Asset Valuation in Incomplete Markets, The Accounting Review, 38-53.

112 The return on equity of 40% becomes a proxy for cash flow potential. With a 20% growth rate and a 40% return on equity, this firm will be able to return half of its earnings to its stockholders in the form of dividends or stock buybacks.

113 Finding these firms manually may be tedious when your universe includes 10000 stocks. You could draw on statistical techniques such as cluster analysis to find similar firms.

114 Alford, A.W., 1992, The Effect of the set of Comparable Firms on the Accuracy of the Price Earnings Valuation Method, Journal of Accounting Research, v30, 94-108.

115 Cheng, C. S. A. and R. McNamara, 2000, The valuation accuracy of the price-earnings and price-book benchmark valuation methods, Review of Quantitative Finance and Accounting, v15, 349-370; Bhojraj, S. and C. M. C. Lee (2002): Who is my peer? A valuation-based approach to the selection of comparable firms, Journal of Accounting Research, *v*40, 407-439. Bhojraj S., C. M. C. Lee, Oler D. (2003), What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. Journal of Accounting Research, v41, 745-774.

Controlling for Differences across Firms

No matter how carefully we construct our list of comparable firms, we will end up with firms that are different from the firm we are valuing. The differences may be small on some variables and large on others and we will have to control for these differences in a relative valuation. There are three ways of controlling for these differences.

1. Subjective Adjustments

Relative valuation begins with two choices - the multiple used in the analysis and the group of firms that comprises the comparable firms. In many relative valuations, the multiple is calculated for each of the comparable firms and the average is computed. One issue that does come up with subjective adjustments to industry average multiples is how best to compute that average. Beatty, Riffe and Thompson (1999) examine multiples of earnings, book value and total assets and conclude that the harmonic mean provides better estimates of value than the arithmetic mean.[116] To evaluate an individual firm, the analyst then compare the multiple it trades at to the average computed; if it is significantly different, the analyst can make a subjective judgment about whether the firm's individual characteristics (growth, risk or cash flows) may explain the difference. If, in the judgment of the analyst, the difference on the multiple cannot be explained by the fundamentals, the firm will be viewed as over valued (if its multiple is higher than the average) or undervalued (if its multiple is lower than the average). The weakness in this approach is not that analysts are called upon to make subjective judgments, but that the judgments are often based upon little more than guesswork. All too often, these judgments confirm their biases about companies.

2. Modified Multiples

In this approach, we modify the multiple to take into account the most important variable determining it – the companion variable. To provide an illustration, analysts who compare PE ratios across companies with very different growth rates often divide the PE ratio by the expected growth rate in EPS to determine a growth-adjusted PE ratio or the PEG ratio. This ratio is then compared across companies with different growth rates to find under and over valued companies. There are two implicit assumptions that we make

---

[116] Beatty, R.P., S.M. Riffe, and R. Thompson, 1999, The method of comparables and tax court

when using these modified multiples. The first is that these firms are comparable on all the other measures of value, other than the one being controlled for. In other words, when comparing PEG ratios across companies, we are assuming that they are all of equivalent risk. If some firms are riskier than others, you would expect them to trade at lower PEG ratios. The other assumption generally made is that that the relationship between the multiples and fundamentals is linear. Again, using PEG ratios to illustrate the point, we are assuming that as growth doubles, the PE ratio will double; if this assumption does not hold up and PE ratios do not increase proportional to growth, companies with high growth rates will look cheap on a PEG ratio basis. Easton (2004) notes that one of the weaknesses of the PEG ratio approach is its emphasis on short term growth and provides a way of estimating the expected rate of return for a stock, using the PEG ratio, and concludes that PEG ratios are effective at ranking stocks.[117]

3. Statistical Techniques

Subjective adjustments and modified multiples are difficult to use when the relationship between multiples and the fundamental variables that determine them becomes complex. There are statistical techniques that offer promise, when this happens. In this section, we will consider the advantages of these approaches and potential concerns.

Sector Regressions

In a regression, we attempt to explain a dependent variable by using independent variables that we believe influence the dependent variable. This mirrors what we are attempting to do in relative valuation, where we try to explain differences across firms on a multiple (PE ratio, EV/EBITDA) using fundamental variables (such as risk, growth and cash flows). Regressions offer three advantages over the subjective approach:

a. The output from the regression gives us a measure of how strong the relationship is between the multiple and the variable being used. Thus, if we are contending that higher growth companies have higher PE ratios, the regression should yield clues to both how growth and PE ratios are related (through the coefficient on growth as an

---

valuations of private firms: an empirical investigation, Accounting Horizons 13, 177–199.

[117] Easton, P., 2004, PE Ratios, PEG Ratios and Estimating the Implied Expected Rate of Return on Equity Capital, The Accounting Review, v79, 79-95.

independent variable) and how strong the relationship is (through the t statistics and R squared).

b. If the relationship between a multiple and the fundamental we are using to explain it is non-linear, the regression can be modified to allow for the relationship.

c. Unlike the modified multiple approach, where we were able to control for differences on only one variable, a regression can be extended to allow for more than one variable and even for cross effects across these variables.

In general, regressions seem particularly suited to our task in relative valuation, which is to make sense of voluminous and sometimes contradictory data. There are two key questions that we face when running sector regressions:

- The first relates to how we define the sector. If we define sectors too narrowly, we run the risk of having small sample sizes, which undercut the usefulness of the regression. Defining sectors broadly entails fewer risks. While there may be large differences across firms when we do this, we can control for those differences in the regression.
- The second involves the independent variables that we use in the regression. While the focus in statistics exercises is increasing the explanatory power of the regression (through the R-squared) and including any variables that accomplish this, the focus of regressions in relative valuations is narrower. Since our objective is not to explain away all differences in pricing across firms but only those differences that are explained by fundamentals, we should use only those variables that are related to those fundamentals. The last section where we analyzed multiples using DCF models should yield valuable clues. As an example, consider the PE ratio. Since it is determined by the payout ratio, expected growth and risk, we should include only those variables in the regression. We should not add other variables to this regression, even if doing so increases the explanatory power, if there is no fundamental reason why these variables should be related to PE ratios.

Market Regression

Searching for comparable firms within the sector in which a firm operates is fairly restrictive, especially when there are relatively few firms in the sector or when a firm operates in more than one sector. Since the definition of a comparable firm is not one that is in the same business but one that has the same growth, risk and cash flow characteristics as the firm being analyzed, we need not restrict our choice of comparable firms to those in the same industry. The regression introduced in the previous section controls for differences on those variables that we believe cause multiples to vary across firms. Based upon the variables that determine each multiple, we should be able to regress PE, PBV and PS ratios against the variables that should affect them. As shown in the last section, the fundamentals that determine each multiple are summarized in table 2:

Table 2: Fundamentals Determining Equity Multiples

Multiple	*Fundamental Determinants*
Price Earnings Ratio	Expected Growth, Payout, Risk
Price to Book Equity Ratio	Expected Growth, Payout, Risk, ROE
Price to Sales Ratio	Expected Growth, Payout, Risk, Net Margin
EV to EBITDA	Expected Growth, Reinvestment Rate, Risk, ROC, Tax rate
EV to Capital Ratio	Expected Growth, Reinvestment Rate, Risk, ROC
EV to Sales	Expected Growth, Reinvestment Rate, Risk, Operating Margin

It is, however, possible that the proxies that we use for risk (beta), growth (expected growth rate in earnings per share), and cash flow (payout) are imperfect and that the relationship is not linear. To deal with these limitations, we can add more variables to the regression - e.g., the size of the firm may operate as a good proxy for risk.

The first advantage of this market-wide approach over the “subjective” comparison across firms in the same sector, described in the previous section, is that it does quantify, based upon actual market data, the degree to which higher growth or risk should affect the multiples. It is true that these estimates can contain errors, but those errors are a reflection of the reality that many analysts choose not to face when they make subjective judgments. Second, by looking at all firms in the market, this approach allows us to make more meaningful comparisons of firms that operate in industries with relatively few firms. Third, it allows us to examine whether all firms in an industry are under- or overvalued, by estimating their values relative to other firms in the market.

In one of the earliest regressions of PE ratios against fundamentals across the market, Kisor and Whitbeck(1963) used data from the Bank of New York for 135 stocks to arrive at the following result.[118]

P/E = 8.2 + 1.5 (Growth rate in Earnings) + 6.7 (Payout ratio) - 0.2 (Standard Deviation in EPS changes)

Cragg and Malkiel (1968) followed up by estimating the coefficients for a regression of the price-earnings ratio on the growth rate, the payout ratio and the beta for stocks for the time period from 1961 to 1965.[119]

Year	*Equation*	R^2
1961	$P/E = 4.73 + 3.28\ g + 2.05\ \pi - 0.85\ \beta$	0.70
1962	$P/E = 11.06 + 1.75\ g + 0.78\ \pi - 1.61\ \beta$	0.70
1963	$P/E = 2.94 + 2.55\ g + 7.62\ \pi - 0.27\ \beta$	0.75
1964	$P/E = 6.71 + 2.05\ g + 5.23\ \pi - 0.89\ \beta$	0.75
1965	$P/E = 0.96 + 2.74\ g + 5.01\ \pi - 0.35\ \beta$	0.85

where,

P/E = Price/Earnings Ratio at the start of the year

g = Growth rate in Earnings

π = Earnings payout ratio at the start of the year

β = Beta of the stock

They concluded that while such models were useful in explaining PE ratios, they were of little use in predicting performance. In both of these studies, the three variables used – payout, risk and growth – represent the three variables that were identified as the determinants of PE ratios in an earlier section.

The regressions were updated in Damodaran (1996, 2002) using a much broader sample of stocks and for a much wider range of multiples.[120] The results for PE ratios from 1987 to 1991 are summarized below.

---

[118] Kisor, M., Jr., and V.S. Whitbeck, 1963, A New Tool in Investment Decision-Making, Financial Analysts Journal, v19, 55-62.

[119] Cragg, J.G., and B.G. Malkiel, 1968, The Consensus and Accuracy of Predictions of the Growth of Corporate Earnings, Journal of Finance, v23, 67-84.

[120] Damodaran, A., 1996 & 2004, Investment Valuation, John Wiley and Sons (first and second editions). These regressions look at all stocks listed on the COMPUSTAT database and similar regressions are run using price to book, price to sales and enterprise value multiples. The updated versions of these regressions

Year	*Regression*	*R squared*
1987	PE = 7.1839 + 13.05 PAYOUT - 0.6259 BETA + 6.5659 EGR	0.9287
1988	PE = 2.5848 + 29.91 PAYOUT - 4.5157 BETA + 19.9143 EGR	0.9465
1989	PE = 4.6122 + 59.74 PAYOUT - 0.7546 BETA + 9.0072 EGR	0.5613
1990	PE = 3.5955 + 10.88 PAYOUT - 0.2801 BETA + 5.4573 EGR	0.3497
1991	PE = 2.7711 + 22.89 PAYOUT - 0.1326 BETA + 13.8653 EGR	0.3217

Note the volatility in the R-squared over time and the changes in the coefficients on the independent variables. For instance, the R squared in the regressions reported above declines from 0.93 in 1987 to 0.32 in 1991 and the coefficients change dramatically over time. Part of the reason for these shifts is that earnings are volatile and the price-earnings ratios reflect this volatility. The low R-squared for the 1991 regression can be ascribed to the recession's effects on earnings in that year. These regressions are clearly not stable, and the predicted values are likely to be noisy. In addition, the regressions for book value and revenue multiples consistently have higher explanatory power than the regressions for price earnings ratios.

Limitations of Statistical Techniques

Statistical techniques are not a panacea for research or for qualitative analysis. They are tools that every analyst should have access to, but they should remain tools. In particular, when applying regression techniques to multiples, we need to be aware of both the distributional properties of multiples that we talked about earlier in the paper and the relationship among and with the independent variables used in the regression.

- The distribution of multiple values across the population is not normal for a very simple reason; most multiples are restricted from taking on values below zero but can be very large positive values.[121] This can pose problems when using standard regression techniques, and these problems are accentuated with small samples, where the asymmetry in the distribution can be magnified by the existence of a few large outliers.

---

are online at http://www.damodaran.com. The growth rate over the previous 5 years was used as the expected growth rate and the betas were estimated from the CRSP tape.

[121] Damodaran, A., 2006, Damodaran on Valuation (Second Edition), John Wiley and Sons, New York. The distributional characteristics of multiples are described in chapter 7.

- In a multiple regression, the independent variables are themselves supposed to be independent of each other. Consider, however, the independent variables that we have used to explain valuation multiples – cash flow potential or payout ratio, expected growth and risk. Across a sector and over the market, it is quite clear that high growth companies will tend to be risky and have low payout. This correlation across independent variables creates "multicollinearity" which can undercut the explanatory power of the regression.
- The distributions for multiples change over time, making comparisons of PE ratios or EV/EBITDA multiples across time problematic. By the same token, a multiple regression where we explain differences in a multiple across companies at a point in time will itself lose predictive power as it ages. A regression of PE ratios against growth rates in early 2005 may therefore not be very useful in valuing stocks in early 2006.
- As a final note of caution, the R-squared on relative valuation regressions will almost never be higher than 70% and it is common to see them drop to 30 or 35%. Rather than ask the question of how high an R-squared has to be to be meaningful, we would focus on the predictive power of the regression. When the R-squared decreases, the ranges on the forecasts from the regression will increase.

Reconciling Relative and Discounted Cash Flow Valuations

The two approaches to valuation – discounted cash flow valuation and relative valuation – will generally yield different estimates of value for the same firm at the same point in time. It is even possible for one approach to generate the result that the stock is under valued while the other concludes that it is over valued. Furthermore, even within relative valuation, we can arrive at different estimates of value depending upon which multiple we use and what firms we based the relative valuation on.

The differences in value between discounted cash flow valuation and relative valuation come from different views of market efficiency, or put more precisely, market inefficiency. In discounted cash flow valuation, we assume that markets make mistakes, that they correct these mistakes over time, and that these mistakes can often occur across entire sectors or even the entire market. In relative valuation, we assume that while markets make mistakes on individual stocks, they are correct on average. In other words, when we value a new software company relative to other small software companies, we are assuming that the market has priced these companies correctly, on average, even though it might have made mistakes in the pricing of each of them individually. Thus, a stock may be over valued on a discounted cash flow basis but under valued on a relative basis, if the firms used for comparison in the relative valuation are all overpriced by the market. The reverse would occur, if an entire sector or market were underpriced.

Kaplan and Ruback (1995) examine the transactions prices paid for 51 companies in leveraged buyout deals and conclude that discounted cash flow valuations yield values very similar to relative valuations, at least for the firms in their sample.[122] They used the compressed APV approach, described in an earlier section, to estimate discounted cash flow values and multiples of EBIT and EBITDA to estimate relative values. Berkman, Bradbury and Ferguson (2000) use the PE ratio and discounted cash flow valuation models to value 45 newly listed companies on the New Zealand Stock Exchange and conclude that both approaches explain about 70% of the price variation and have similar accuracy.[123] In contrast to these findings, Kim and Ritter (1998) value a group of IPOs using PE and price to book ratios and conclude that multiples have only modest predictive ability.[124] Lee, Myers and Swaminathan (1999) compare valuations obtained for the Dow 30 stocks using both multiples and a discounted cash flow model, based upon residual income, and conclude that prices are more likely to converge on the latter in the long term. While the evidence seems contradictory, it can be explained by the fact the studies that find relative valuation works well look at cross sectional differences across stocks, whereas studies that look at pricing differences that correct over time conclude that intrinsic valuations are more useful.[125]

Directions for future research

As we survey the research done on valuation in the last few decades, there are three key trends that emerge from the research. First, the focus has shifted from valuing

---

[122] Kaplan, S.N. and R.S. Ruback, 1995, The Valuation of Cash Flow Forecasts: An Empirical Analysis, Journal of Finance, v50, 1059-1093.
[123] Berkman, H., M.E. Bradbury and J. Ferguson, 2000, The Accuracy of Price-Earnings and Discounted Cash Flow Methods of IPO Equity Valuation, Journal of International Financial Management and Accounting, v11, 71-83.
[124] Kim, M. and J. R. Ritter (1999): Valuing IPOs, Journal of Financial Economics, v53, 409-437.
[125] Lee, C.M.C., J. Myers and B.Swaminathan, 1999, What is the intrinsic value of the Dow?, Journal of Finance, v54, 1693-1741.

stocks through models such as the dividend discount model to valuing businesses, representing the increased use of valuation models in acquisitions and corporate restructuring (where the financing mix is set by the acquirer) and the possibility that financial leverage can change quickly over time. Second, the connections between corporate finance and valuation have become clearer as value is linked to a firm's actions. In particular, the excess return models link value directly to the quality of investment decisions, whereas adjusted present value models make value a function of financing choices. Third, the comforting conclusion is that all models lead to equivalent values, with consistent assumptions, which should lead us to be suspicious of new models that claim to be more sophisticated and yield more precise values than prior iterations.

The challenges for valuation research in the future lie in the types of companies that we are called upon to value. First, the shift of investments from developed markets to emerging markets in Asia and Latin America has forced us to re-examine the assumptions we make about value. In particular, the interrelationship between corporate governance and value, and the question of how best to deal with the political and economic risk endemic to emerging markets have emerged as key topics. Second, the entry of young companies into public markets, often well before they have established revenue and profit streams, requires us to turn our attention to estimation questions: How best do we estimate the revenues and margins for a firm that has an interesting product idea but no commercial products? How do we forecast the reinvestment needs and estimate discount rates for such a firm? Third, with both emerging market and young companies, we need to reassess our dependence on current financial statement values as the basis for valuation. For firms in transition, in markets that are themselves changing, we need to be able to allow for significant changes in fundamentals, be they risk parameters, debt ratios and growth rats, over time. In short, we need dynamic valuation models rather than the static ones that we offer as the default currently. Fourth, as the emphasis has shifted from growth to excess returns as the driver of value, the importance of tying corporate strategy to value has also increased. After all, corporate strategy is all about creating new barriers to entry and augmenting or preserving existing ones, and much work needs to be done at the intersection of strategy and valuation. Understanding why a company earns excess returns in the first place and why those excess returns may come under assault is a pre-requisit for good valuation. Finally, while the increase in computing power and easy access to statistical tools has opened the door to more sophisticated variations in valuation, it has also increased the potential for misuse of these tools. Research on how best to incorporate statistical tools into the conventional task of valuing a business is needed. In particular, is there a place for simulations in valuation and if so, what is it? How about scenario analysis or neural networks? The good news is that there is a great deal of interesting work left to be done in valuation. The bad news is that it will require a mix of interdisciplinary skills including accounting, corporate strategy, statistics and corporate finance for this research to have a significant impact.

Conclusion

Since valuation is key to so much of what we do in finance, it is not surprising that there are a myriad of valuation approaches in use. In this paper, we examined three different approaches to valuation, with numerous sub-approaches within each. The first is discounted cash flow valuation, where the value of a business or asset is determined by its cash flows and can be estimated in one of four ways: (a) expected cash flows can be discounted back at a risk-adjusted discount rate (b) uncertain cash flows can be converted into certainty equivalents and discounted back at a riskfree rate (c) expected cash flows can be broken down into normal (representing a fair return on capital invested) and excess return cash flows and valued separately and (d) the value of the asset or business is first estimated on an all-equity funded basis and the effects of debt on value are computed separately. Not surprisingly, given their common roots, these valuation approaches can be shown to yield the same value for an asset, if we make consistent assumptions. In practice, though, proponents of these approaches continue to argue for their superiority and arrive at very different asset values, often because of difference in the implicit assumptions that they make within each approach.

The second approach has its roots in accounting, and builds on the notion that there is significant information in the book value of a firm's assets and equity. While there are few who would claim that the book value is a good measure of the true value, there are approaches that build on the book value and accrual earnings to arrive at consistent estimates of value. In recent years, there has also been a push towards fair value accounting with the ultimate objective of making balance sheets more informative and value relevant.

The third approach to valuation is relative valuation, where we value an asset based upon how similar assets are priced. It is built on the assumption that the market, while it may be wrong in how it prices individual assets, gets it right on average and is clearly the dominant valuation approach in practice. Relative valuation is built on standardized prices, where we scale the market value to some common measure such as earnings, book value or revenues, but the determinants of these multiples are the same ones that underlie discounted cash flow valuation.

Equity Risk Premiums (ERP): Determinants, Estimation and Implications – The 2015 Edition

Updated: March 2015

Aswath Damodaran

Stern School of Business

adamodar@stern.nyu.edu

Electronic copy available at: http://ssrn.com/abstract=2581517

Equity Risk Premiums (ERP): Determinants, Estimation and Implications – The 2015 Edition

Equity risk premiums are a central component of every risk and return model in finance and are a key input in estimating costs of equity and capital in both corporate finance and valuation. Given their importance, it is surprising how haphazard the estimation of equity risk premiums remains in practice. We begin this paper by looking at the economic determinants of equity risk premiums, including investor risk aversion, information uncertainty and perceptions of macroeconomic risk. In the standard approach to estimating the equity risk premium, historical returns are used, with the difference in annual returns on stocks versus bonds over a long time period comprising the expected risk premium. We note the limitations of this approach, even in markets like the United States, which have long periods of historical data available, and its complete failure in emerging markets, where the historical data tends to be limited and volatile. We look at two other approaches to estimating equity risk premiums – the survey approach, where investors and managers are asked to assess the risk premium and the implied approach, where a forward-looking estimate of the premium is estimated using either current equity prices or risk premiums in non-equity markets. In the next section, we look at the relationship between the equity risk premium and risk premiums in the bond market (default spreads) and in real estate (cap rates) and how that relationship can be mined to generated expected equity risk premiums. We close the paper by examining why different approaches yield different values for the equity risk premium, and how to choose the "right" number to use in analysis.

(This is the eighth update of this piece. The first update was in the midst of the financial crisis in 2008 and there have been annual updates at the start of each year from 2009 through 2014.)

Electronic copy available at: http://ssrn.com/abstract=2581517

The notion that risk matters, and that riskier investments should have higher expected returns than safer investments, to be considered good investments, is intuitive and central to risk and return models in finance. Thus, the expected return on any investment can be written as the sum of the riskfree rate and a risk premium to compensate for the risk. The disagreement, in both theoretical and practical terms, remains on how to measure the risk in an investment, and how to convert the risk measure into an expected return that compensates for risk. A central number in this debate is the premium that investors demand for investing in the 'average risk' equity investment (or for investing in equities as a class), i.e., the equity risk premium.

In this paper, we begin by examining competing risk and return models in finance and the role played by equity risk premiums in each of them. We argue that equity risk premiums are central components in every one of these models and consider what the determinants of these premiums might be. We follow up by looking at three approaches for estimating the equity risk premium in practice. The first is to survey investors or managers with the intent of finding out what they require as a premium for investing in equity as a class, relative to the riskfree rate. The second is to look at the premiums earned historically by investing in stocks, as opposed to riskfree investments. The third is to back out an equity risk premium from market prices today. We consider the pluses and minuses of each approach and how to choose between the very different numbers that may emerge from these approaches.

Equity Risk Premiums: Importance and Determinants

Since the equity risk premium is a key component of every valuation, we should begin by looking at not only why it matters in the first place but also the factors that influence its level at any point in time and why that level changes over time. In this section, we look at the role played by equity risk premiums in corporate financial analysis, valuation and portfolio management, and then consider the determinants of equity risk premiums.

Why does the equity risk premium matter?

The equity risk premium reflects fundamental judgments we make about how much risk we see in an economy/market and what price we attach to that risk. In the process, it affects the expected return on every risky investment and the value that we estimate for that investment. Consequently, it makes a difference in both how we allocate wealth across different asset classes and which specific assets or securities we invest in within each asset class.

A Price for Risk

To illustrate why the equity risk premium is the price attached to risk, consider an alternate (though unrealistic) world where investors are risk neutral. In this world, the value of an asset would be the present value of expected cash flows, discounted back at a risk free rate. The expected cash flows would capture the cash flows under all possible scenarios (good and bad) and there would be no risk adjustment needed. In the real world, investors are risk averse and will pay a lower price for risky cash flows than for riskless cash flows, with the same expected value. How much lower? That is where equity risk premiums come into play. In effect, the equity risk premium is the premium that investors demand for the average risk investment, and by extension, the discount that they apply to expected cash flows with average risk. When equity risk premiums rise, investors are charging a higher price for risk and will therefore pay lower prices for the same set of risky expected cash flows.

Expected Returns and Discount Rates

Building on the theme that the equity risk premium is the price for taking risk, it is a key component into the expected return that we demand for a risky investment. This expected return, is a determinant of both the cost of equity and the cost of capital, essential inputs into corporate financial analysis and valuation.

While there are several competing risk and return models in finance, they all share some common assumptions about risk. First, they all define risk in terms of variance in actual returns around an expected return; thus, an investment is riskless when actual returns are always equal to the expected return. Second, they argue that risk has to be measured from the perspective of the marginal investor in an asset, and that this marginal investor is well diversified. Therefore, the argument goes, it is only the risk that an investment adds on to a diversified portfolio that should be measured and compensated. In fact, it is this view of risk that leads us to break the risk in any investment into two components. There is a firm-specific component that measures risk that relates only to that investment or to a few investments like it, and a market component that contains risk that affects a large subset or all investments. It is the latter risk that is not diversifiable and should be rewarded.

All risk and return models agree on this fairly crucial distinction, but they part ways when it comes to how to measure this market risk. In the capital asset pricing model (CAPM), the market risk is measured with a beta, which when multiplied by the equity risk premium yields the total risk premium for a risky asset. In the competing models, such as the arbitrage pricing and multi-factor models, betas are estimated against individual market risk factors, and each factor has it own price (risk premium). Table 1 summarizes four models, and the role that equity risk premiums play in each one:

Table 1: Equity Risk Premiums in Risk and Return Models

	Model	*Equity Risk Premium*
The CAPM	Expected Return = Riskfree Rate + $\text{Beta}_{\text{Asset}}$ (Equity Risk Premium)	Risk Premium for investing in the market portfolio, which includes all risky assets, relative to the riskless rate.
Arbitrage pricing model (APM)	$\text{Expected Return} = \text{Riskfree Rate} + \sum_{j=1}^{j=k} \beta_j (\text{Risk Premium}_j)$	Risk Premiums for individual (unspecified) market risk factors.
Multi-Factor Model	$\text{Expected Return} = \text{Riskfree Rate} + \sum_{j=1}^{j=k} \beta_j (\text{Risk Premium}_j)$	Risk Premiums for individual (specified) market risk factors
Proxy Models	Expected Return = a + b (Proxy 1) + c (Proxy 2) (where the proxies are firm characteristics such as market capitalization, price to book ratios or return momentum)	No explicit risk premium computation, but coefficients on proxies reflect risk preferences.

All of the models other than proxy models require three inputs. The first is the riskfree rate, simple to estimate in currencies where a default free entity exists, but more complicated in markets where there are no default free entities. The second is the beta (in the CAPM) or betas (in the APM or multi-factor models) of the investment being analyzed, and the third is the appropriate risk premium for the portfolio of all risky assets (in the CAPM) and the factor risk premiums for the market risk factors in the APM and multi-factor models. While I examine the issues of riskfree rate and beta estimation in companion pieces, I will concentrate on the measurement of the risk premium in this paper.

Note that the equity risk premium in all of these models is a market-wide number, in the sense that it is not company specific or asset specific but affects expected returns on all risky investments. Using a larger equity risk premium will increase the expected returns for all risky investments, and by extension, reduce their value. Consequently, the choice of an equity risk premium may have much larger consequences for value than firm-specific inputs such as cash flows, growth and even firm-specific risk measures (such as betas).

Investment and Policy Implications

It may be tempting for those not in the midst of valuation or corporate finance analysis to pay little heed to the debate about equity risk premium, but it would be a mistake to do so, since its effects are far reaching.

- The amounts set aside by both corporations and governments to meet future pension fund and health care obligations are determined by their expectations of returns from investing in equity markets, i.e., their views on the equity risk premium. Assuming that the equity risk premium is 6% will lead to far less being set aside each year to cover future obligations than assuming a premium of 4%. If the actual premium delivered by equity markets is only 2%, the fund's assets will be insufficient to meet its liabilities, leading to fund shortfalls which have to be met by raising taxes (for governments) or reducing profits (for corporations) In some cases, the pension benefits can be put at risk, if plan administrators use unrealistically high equity risk premiums, and set aside too little each year.
- Business investments in new assets and capacity is determined by whether the businesses think they can generate higher returns on those investments than the cost that they attach to the capital in that investment. If equity risk premiums increase, the cost of equity and capital will have to increase with them, leading to less overall investment in the economy and lower economic growth.
- Regulated monopolies, such as utility companies, are often restricted in terms of the prices that they charge for their products and services. The regulatory commissions that determine "reasonable" prices base them on the assumption that these companies have to earn a fair rate of return for their equity investors. To come up with this fair rate of return, they need estimates of equity risk premiums; using higher equity risk premiums will translate into higher prices for the customers in these companies.[1]
- Judgments about how much you should save for your retirement or health care and where you should invest your savings are clearly affected by how much return you think you can make on your investments. Being over optimistic about equity risk premiums will lead you to save too little to meet future needs and to over investment in risky asset classes.

Thus, the debate about equity risk premiums has implications for almost every aspect of our lives.

Market Timing and Risk Premiums

Any one who invests has a view on equity risk premiums, though few investors are explicit about their views. In particular, if you believe that markets are efficient, you

---

[1] The Society of Utility and Regulatory Financial Analysts (SURFA) has annual meetings of analysts involved primarily in this debate. Not surprisingly, they spend a good chunk of their time discussing equity risk premiums, with analysts working for the utility firms arguing for higher equity risk premiums and analysts working for the state or regulatory authorities wanting to use lower risk premiums.

are arguing that the equity risk premiums built into market prices today are correct. If you believe that stock markets are over valued or in a bubble, you are asserting that the equity risk premiums built into prices today are too low, relative to what they should be (based on the risk in equities and investor risk aversion). Conversely, investors who believe that stocks are collectively underpriced or cheap are also making a case that the equity risk premium in the market today is much higher than what you should be making (again based on the risk in equities and investor risk aversion). Thus, every debate about the overall equity market can be translated into a debate about equity risk premiums.

Put differently, asset allocation decisions that investors make are explicitly or implicitly affected by investor views on risk premiums and how they vary across asset classes and geographically. Thus, if you believe that equity risk premiums are low, relative to the risk premiums in corporate bond markets (which take the form or default spreads on bonds), you will allocated more of your overall portfolio to bonds. Your allocation of equities across geographical markets are driven by your perceptions of equity risk premiums in those markets, with more of your portfolio going into markets where the equity risk premium is higher than it should be (given the risk of those markets). Finally, if you determine that the risk premiums in financial assets (stocks and bonds) are too low, relative to what you can earn in real estate or other real assets, you will redirect more of your portfolio into the latter.

By making risk premiums the focus of asset allocation decisions, you give focus to those decisions. While it is very difficult to compare PE ratios for stocks to interest rates on bonds and housing price indicators, you can compare equity risk premiums to default spreads to real estate capitalization rates to make judgments about where you get the best trade off on risk and return. In fact, we will make these comparisons later in this paper.

What are the determinants of equity risk premiums?

Before we consider different approaches for estimating equity risk premiums, we should examine the factors that determine equity risk premiums. After all, equity risk premiums should reflect not only the risk that investors see in equity investments but also the price they attach to that risk.

Risk Aversion and Consumption Preferences

The first and most critical factor, obviously, is the risk aversion of investors in the markets. As investors become more risk averse, equity risk premiums will climb, and as risk aversion declines, equity risk premiums will fall. While risk aversion will vary across investors, it is the collective risk aversion of investors that determines equity risk premium, and changes in that collective risk aversion will manifest themselves as changes in the equity risk premium. While there are numerous variables that influence risk aversion, we will focus on the variables most likely to change over time.

a. Investor Age: There is substantial evidence that individuals become more risk averse as they get older. The logical follow up to this proposition is that markets with older investors, in the aggregate, should have higher risk premiums than markets with younger investors, for any given level of risk. Bakshi and Chen (1994), for instance, examined risk premiums in the United States and noted an increase in risk premiums as investors aged.[2] Liu and Spiegel computed the ratio of the middle-age cohort (40-49 years) to the old-age cohort (60-69) and found that PE ratios are closely and positively related to the MO ratio for the US equity market from 1954 to 2010; since the equity risk premium is inversely related to the PE, this would suggest that investor age does play a role in determining equity risk premiums.[3]
b. Preference for current consumption: We would expect the equity risk premium to increase as investor preferences for current over future consumption increase. Put another way, equity risk premiums should be lower, other things remaining equal, in markets where individuals are net savers than in markets where individuals are net consumers. Consequently, equity risk premiums should increase as savings rates decrease in an economy. Rieger, Wang and Hens (2012) compare equity risk premiums and time discount factors across 27 countries and find that premiums are higher in countries where investors are more short term.[4]

Relating risk aversion to expected equity risk premiums is not straightforward. While the direction of the relationship is simple to establish – higher risk aversion should translate into higher equity risk premiums- getting beyond that requires us to be more precise in our judgments about investor utility functions, specifying how investor utility relates to wealth (and variance in that wealth). As we will see later in this paper, there has been a significant angst among financial economics that most conventional utility models do not do a good job of explaining observed equity risk premiums.

---

[2] Bakshi, G. S., and Z. Chen, 1994, *Baby Boom, Population Aging, and Capital Markets*, The Journal of Business, LXVII, 165-202.

[3] Liu, Z. and M.M. Siegel, 2011, *Boomer Retirement: Headwinds for US Equity Markets?* FRBSF Economic Letters, v26.

[4] Rieger, M.O., M. Wang and T. Hens, 2012, International Evidence on the Equity Risk Premium Puzzle and Time Discounting, SSRN Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2120442

Economic Risk

The risk in equities as a class comes from more general concerns about the health and predictability of the overall economy. Put in more intuitive terms, the equity risk premium should be lower in an economy with predictable inflation, interest rates and economic growth than in one where these variables are volatile. Lettau, Ludwigson and Wachter (2008) link the changing equity risk premiums in the United States to shifting volatility in the real economy.[5] In particular, they attribute that that the lower equity risk premiums of the 1990s (and higher equity values) to reduced volatility in real economic variables including employment, consumption and GDP growth. One of the graphs that they use to illustrate the correlation looks at the relationship between the volatility in GDP growth and the dividend/ price ratio (which is the loose estimate that they use for equity risk premiums), and it is reproduced in figure 1.

Figure 1: Volatility in GDP growth and Equity Risk Premiums (US)

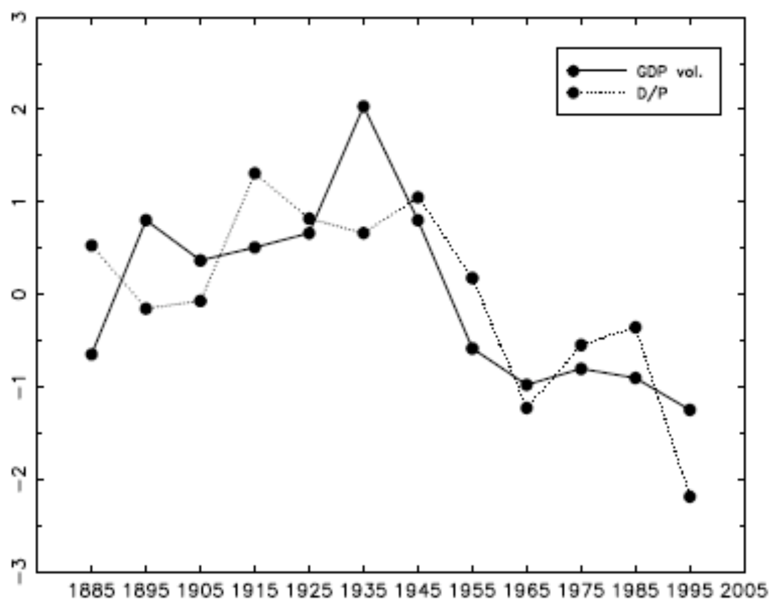


Figure 3
GDP volatility and the D/P ratio—Prewar evidence
This figure plots the standard deviations of GDP growth and the mean D/P ratio by decade starting in 1880 until 2000. Both series are demeaned and divided by their standard deviation. The GDP data are from Ray Fair's website (http://fairmodel.econ.yale.edu/RAYFAIR/PDF/2002DTBL.HTM) based on Balke and Gordon (1989). The dividend yield data is from Robert Shiller's website (http://aida.econ.yale.edu/~shiller/data/ie_data.htm).

Note how closely the dividend yield has tracked the volatility in the real economy over this very long time period.

---

[5] Lettau, M., S.C. Ludvigson and J.A. Wachter, 2008. *The Declining Equity Risk Premium: What role does macroeconomic risk play?* Review of Financial Studies, v21, 1653-1687.

Gollier (2001) noted that the linear absolute risk tolerance often assumed in standard models breaks down when there is income inequality and the resulting concave absolute risk tolerance should lead to higher equity risk premiums.[6] Hatchondo (2008) attempted to quantify the impact on income inequality on equity risk premiums. In his model, which is narrowly structured, the equity risk premium is higher in an economy with unequal income than in an egalitarian setting, but only by a modest amount (less than 0.50%).[7]

A related strand of research examines the relationship between equity risk premium and inflation, with mixed results. Studies that look at the relationship between the level of inflation and equity risk premiums find little or no correlation. In contrast, Brandt and Wang (2003) argue that news about inflation dominates news about real economic growth and consumption in determining risk aversion and risk premiums.[8] They present evidence that equity risk premiums tend to increase if inflation is higher than anticipated and decrease when it is lower than expected. Another strand of research on the Fisher equation, which decomposes the riskfree rate into expected inflation and a real interest rate, argues that when inflation is stochastic, there should be a third component in the risk free rate: an inflation risk premium, reflecting uncertainty about future inflation.[9] Reconciling the findings, it seems reasonable to conclude that it is not so much the level of inflation that determines equity risk premiums but uncertainty about that level, and that some of the inflation uncertainty premium may be captured in the risk free rate, rather than in the equity risk premiums.

Since the 2008 crisis, with its aftermath oflow government bond rates and a simmering economic crisis, equity risk premiums in the United States have behaved differently than they have historically. Connolly and Dubofsky (2015) find that equity risk premiums have increased (decreased) as US treasury bond rates decrease (increase), and have moved inversely with inflation (with higher inflation leading to lower equity risk premiums), both behaviors at odds with the relationship in the pre-2008 time period, suggesting a structural break in 2008.[10]

---

[6] Gollier, C., 2001. *Wealth Inequality and Asset Pricing*, Review of Economic Studies, v68, 181–203.

[7] Hatchondo, J.C., 2008, *A Quantitative Study of the Role of Income Inequality on Asset Prices*, Economic Quarterly, v94, 73–96.

[8] Brandt, M.W. and K.Q. Wang. 2003. *Time-varying risk aversion and unexpected inflation*, Journal of Monetary Economics, v50, pp. 1457-1498.

[9] Benninga, S., and A. Protopapadakis, 1983, *Real and Nominal Interest Rates under Uncertainty: The Fisher Problem and the Term Structure*, Journal of Political Economy, vol. 91, pp. 856–67.

[10] Connolly, R. and D. Dubofsky, 2015, *Risk Perceptions, Inflation and Financial Asset Returns: A Tale of Two Connections*, Working Paper, SSRN #2527213.

Information

When you invest in equities, the risk in the underlying economy is manifested in volatility in the earnings and cash flows reported by individual firms in that economy. Information about these changes is transmitted to markets in multiple ways, and it is clear that there have been significant changes in both the quantity and quality of information available to investors over the last two decades. During the market boom in the late 1990s, there were some who argued that the lower equity risk premiums that we observed in that period were reflective of the fact that investors had access to more information about their investments, leading to higher confidence and lower risk premiums in 2000. After the accounting scandals that followed the market collapse, there were others who attributed the increase in the equity risk premium to deterioration in the quality of information as well as information overload. In effect, they were arguing that easy access to large amounts of information of varying reliability was making investors less certain about the future.

As these contrary arguments suggest, the relationship between information and equity risk premiums is complex. More precise information should lead to lower equity risk premiums, other things remaining equal. However, precision here has to be defined in terms of what the information tells us about future earnings and cash flows. Consequently, it is possible that providing more information about last period's earnings may create more uncertainty about future earnings, especially since investors often disagree about how best to interpret these numbers. Yee (2006) defines earnings quality in terms of volatility of future earnings and argues that equity risk premiums should increase (decrease) as earnings quality decreases (increases).[11]

Empirically, is there a relationship between earnings quality and observed equity risk premiums? The evidence is mostly anecdotal, but there are several studies that point to the deteriorating quality of earnings in the United States, with the blame distributed widely. First, the growth of technology and service firms has exposed inconsistencies in accounting definitions of earnings and capital expenditures – the treatment of R&D as an operating expense is a prime example. Second, audit firms have been accused of conflicts of interest leading to the abandonment of their oversight responsibility. Finally, the earnings game, where analysts forecast what firms will earn and firms then try to beat these forecasts has led to the stretching (and breaking) of accounting rules and standards. If earnings have become less informative in the aggregate, it stands to reason that equity

---

[11] Yee, K. K.,, 2006, *Earnings Quality and the Equity Risk Premium: A Benchmark Model*, Contemporary Accounting Research, 23: 833–877.

investors will demand large equity risk premiums to compensate for the added uncertainty.

Information differences may be one reason why investors demand larger risk premiums in some emerging markets than in others. After all, markets vary widely in terms of transparency and information disclosure requirements. Markets like Russia, where firms provide little (and often flawed) information about operations and corporate governance, should have higher risk premiums than markets like India, where information on firms is not only more reliable but also much more easily accessible to investors. Lau, Ng and Zhang (2011) look at time series variation in risk premiums in 41 countries and conclude that countries with more information disclosure, measured using a variety of proxies, have less volatile risk premiums and that the importance of information is heightened during crises (illustrated using the 1997 Asian financial crisis and the 2008 Global banking crisis).[12]

Liquidity and Fund Flows

In addition to the risk from the underlying real economy and imprecise information from firms, equity investors also have to consider the additional risk created by illiquidity. If investors have to accept large discounts on estimated value or pay high transactions costs to liquidate equity positions, they will be pay less for equities today (and thus demand a large risk premium).

The notion that market for publicly traded stocks is wide and deep has led to the argument that the net effect of illiquidity on aggregate equity risk premiums should be small. However, there are two reasons to be skeptical about this argument. The first is that not all stocks are widely traded and illiquidity can vary widely across stocks; the cost of trading a widely held, large market cap stock is very small but the cost of trading an over-the-counter stock will be much higher. The second is that the cost of illiquidity in the aggregate can vary over time, and even small variations can have significant effects on equity risk premiums. In particular, the cost of illiquidity seems to increase when economies slow down and during periods of crisis, thus exaggerating the effects of both phenomena on the equity risk premium.

While much of the empirical work on liquidity has been done on cross sectional variation across stocks (and the implications for expected returns), there have been attempts to extend the research to look at overall market risk premiums. Gibson and Mougeot (2004) look at U.S. stock returns from 1973 to 1997 and conclude that liquidity

[12] Lau. S.T., L. Ng and B. Zhang, 2011, *Information Environment and Equity Risk Premium Volatility around the World*, Management Science, Forthcoming.

accounts for a significant component of the overall equity risk premium, and that its effect varies over time.[13] Baekart, Harvey and Lundblad (2006) present evidence that the differences in equity returns (and risk premiums) across emerging markets can be partially explained by differences in liquidity across the markets.[14]

Another way of framing the liquidity issue is in terms of funds flows, where the equity risk premium is determined by funds flows into and out of equities. Thus, if more funds are flowing into an equity market, either from other asset classes or other geographies, other things remaining equal, the equity risk premium should decrease, whereas funds flowing out of an equity market will lead to higher equity risk premiums.

Catastrophic Risk

When investing in equities, there is always the potential for catastrophic risk, i.e. events that occur infrequently but can cause dramatic drops in wealth. Examples in equity markets would include the great depression from 1929-30 in the United States and the collapse of Japanese equities in the last 1980s. In cases like these, many investors exposed to the market declines saw the values of their investments drop so much that it was unlikely that they would be made whole again in their lifetimes.[15] While the possibility of catastrophic events occurring may be low, they cannot be ruled out and the equity risk premium has to reflect that risk.

Rietz (1988) uses the possibility of catastrophic events to justify higher equity risk premiums and Barro (2006) extends this argument. In the latter's paper, the catastrophic risk is modeled as both a drop in economic output (an economic depression) and partial default by the government on its borrowing.[16] Gabaix (2009) extends the Barro-Rietz model to allow for time varying losses in disasters.[17] Barro, Nakamura, Steinsson and Ursua (2009) use panel data on 24 countries over more than 100 years to examine the empirical effects of disasters.[18] They find that the average length of a disaster is six years

[13] Gibson R., Mougeot N., 2004, *The Pricing of Systematic Liquidity Risk: Empirical Evidence from the US Stock Market*. Journal of Banking and Finance, v28: 157–78.

[14] Bekaert G., Harvey C. R., Lundblad C., 2006, *Liquidity and Expected Returns: Lessons from Emerging Markets*, The Review of Financial Studies.

[15] An investor in the US equity markets who invested just prior to the crash of 1929 would not have seen index levels return to pre-crash levels until the 1940s. An investor in the Nikkei in 1987, when the index was at 40000, would still be facing a deficit of 50% (even after counting dividends) in 2008,

[16] Rietz, T. A., 1988, *The equity premium~: A solution*, Journal of Monetary Economics, v22, 117-131; Barro R J., 2006, *Rare Disasters and Asset Markets in the Twentieth Century*, Quarterly Journal of Economics, August, 823-866.

[17]Gabaix, Xavier, 2012, *Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance*, The Quarterly Journal of Economics, v127, 645-700.

[18] Barro, R., E. Nakamura, J. Steinsson and J. Ursua, 2009, *Crises and Recoveries in an Empirical Model of Consumption Disasters*, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1594554.

and that half of the short run impact is reversed in the long term. Investigating the asset pricing implications, they conclude that the consequences for equity risk premiums will depend upon investor utility functions, with some utility functions (power utility, for instance) yielding low premiums and others generating much higher equity risk premiums. Barro and Ursua (2008) look back to 1870 and identify 87 crises through 2007, with an average impact on stock prices of about 22%, and estimate that investors would need to generate an equity risk premium of 7% to compensate for risk taken.[19] Wachter (2012) builds a consumption model, where consumption follows a normal distribution with low volatility most of the time, with a time-varying probability of disasters that explains high equity risk premiums.[20]

There have been attempts to measure the likelihood of catastrophic risk and incorporate them into models that predict equity risk premiums. In a series of papers with different co-authors, Bollerslev uses the variance risk premium, i.e., the difference between the implied variance in stock market options and realized variance, as a proxy for expectations of catastrophic risk, and documents a positive correlation with equity risk premiums.[21] Kelly (2012) looks at extreme stock market movements as a measure of expected future jump (catastrophic) risk and finds a positive link between jump risk and equity risk premiums.[22] Guo, Liu, Wang, Zhou and Zuo (2014) refine this analysis by decomposing jumps into bad (negative) and good (positive) ones and find that it is the risk of downside jumps that determines equity risk premiums..[23] Maheu, McCurdy and Zhao (2013) used a time-varying jump-arrival process and a two-component GARCH model on US stock market data from 1926 to 2011, and estimated that each additional jump per year increased the equity risk premium by 0.1062% and that there were, on average, 34 jumps a year, leading to a jump equity risk premium of 3.61%.[24]

The banking and financial crisis of 2008, where financial and real estate markets plunged in the last quarter of the year, has provided added ammunition to this school. As

---

19 Barro, R. and J. Ursua, 2008, Macroeconomic Crises since 1870, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1124864.

20 Wachter, J.A., 2013, *Can time-varying risk of rare disasters explain aggregate stock market volatility?* Journal of Finance, v68, 987-1035.

21 Bollerslev, T. M., T. H. Law, and G. Tauchen, 2008, *Risk, Jumps, and Diversification*, Journal of Econometrics, 144, 234-256; Bollerslev, T. M., G. Tauchen, and H. Zhou, 2009, *Expected Stock Returns and Variance Risk Premia*, Review of Financial Studies, 101-3, 552-573; Bollerselv, T.M., and V. Todorov, 2011, *Tails, Fears, and Risk Premia*, Journal of Finance, 66-6, 2165-2211.

22 Kelly, B., 2012, *Tail Risk and Asset Prices*, Working Paper, University of Chicago.

23 Guo, H., Z. Liu, K. Wang, H. Zhou and H. Zuo, 2014, *Good Jumps, Bad Jumps and Conditional Equity Risk Premium*, Working Paper, SSRN #2516074.

24 Maheu, J.M., T.H. McCurdy and X. Wang, 2013, *Do Jumps Contribute to the Dynamics of the Equity Premium,* Journal of Financial Economics, v110, 457-477.

we will see later in the paper, risk premiums in all markets (equity, bond and real estate) climbed sharply during the weeks of the market crisis. In fact, the series of macro crises in the last four years that have affected markets all over the world has led some to hypothesize that the globalization may have increased the frequency and probability of disasters and by extension, equity risk premiums, in all markets.

Government Policy

The prevailing wisdom, at least until 2008, was that while government policy affected equity risk premiums in emerging markets, it was not a major factor in determining equity risk premiums in developed markets. The banking crisis of 2008 and the government responses to it have changed some minds, as both the US government and European governments have made policy changes that at times have calmed markets and at other times roiled them, potentially affecting equity risk premiums.

Pastor and Veronesi (2012) argue that uncertainty about government policy can translate into higher equity risk premiums.[25] The model they develop has several testable implications. First, government policy changes will be more likely just after economic downturns, thus adding policy uncertainty to general economic uncertainty and pushing equity risk premiums upwards. Second, you should expect to see stock prices fall, on average, across all policy changes, with the magnitude of the negative returns increasing for policy changes create more uncertainty. Third, policy changes will increase stock market volatility and the correlation across stocks.

Lam and Zhang (2014) try to capture the potential policy shocks from either an unstable government (government stability) or an incompetent bureaucracy (bureaucracy quality) in 49 countries from 1995 to 2006, using two measures of policy uncertainty drawn from the international country risk guide (ICG). They do find that equity risk premiums are higher in countries with more policy risk from either factor, with more bureaucratic risk increasing the premium by approximately 8%.[26]

The behavioral/ irrational component

Investors do not always behave rationally, and there are some who argue that equity risk premiums are determined, at least partially, by quirks in human behavior. While there are several strands to this analysis, we will focus on three:

---

[25] Pástor, L. and P. Veronesi, 2012. *Uncertainty about Government policy and Stock Prices*. Journal of Finance 67: 1219-1264.

[26] Lam, S.S. and W. Zhang, 2014, *Does Policy Uncertainty matter for International Equity Markets?* Working Paper, SSRN #2297133.

a. <u>The Money Illusion</u>: As equity prices declined significantly and inflation rates increased in the late 1970s, Modigliani and Cohn (1979) argued that low equity values of that period were the consequence of investors being inconsistent about their dealings with inflation. They argued that investors were guilty of using historical growth rates in earnings, which reflected past inflation, to forecast future earnings, but current interest rates, which reflected expectations of future inflation, to estimate discount rates.[27] When inflation increases, this will lead to a mismatch, with high discount rates and low cash flows resulting in asset valuations that are too low (and risk premiums that are too high). In the Modigliani-Cohn model, equity risk premiums will rise in periods when inflation is higher than expected and drop in periods when inflation in lower than expected. Campbell and Voulteenaho (2004) update the Modigliani-Cohn results by relating changes in the dividend to price ratio to changes in the inflation rate over time and find strong support for the hypothesis.[28]

b. <u>Narrow Framing:</u> In conventional portfolio theory, we assume that investors assess the risk of an investment in the context of the risk it adds to their overall portfolio, and demand a premium for this risk. Behavioral economists argue that investors offered new gambles often evaluate those gambles in isolation, separately from other risks that they face in their portfolio, leading them to over estimate the risk of the gamble. In the context of the equity risk premium, Benartzi and Thaler (1995) use this "narrow framing" argument to argue that investors over estimate the risk in equity, and Barberis, Huang and Santos (2001) build on this theme.[29]

The Equity Risk Premium Puzzle

While many researchers have focused on individual determinants of equity risk premiums, there is a related question that has drawn almost as much attention. Are the equity risk premiums that we have observed in practice compatible with the theory? Mehra and Prescott (1985) fired the opening shot in this debate by arguing that the observed historical risk premiums (which they estimated at about 6% at the time of their analysis) were too high, and that investors would need implausibly high risk-aversion

---

[27] Modigliani, Franco and Cohn, Richard. 1979, *Inflation, Rational Valuation, and the Market*, Financial Analysts Journal, v37(3), pp. 24-44.

[28] Campbell, J.Y. and T. Vuolteenaho, 2004, *Inflation Illusion and Stock Prices,* American Economic Review, v94, 19-23.

[29] Benartzi, S. and R. Thaler, 1995, *Myopic Loss Aversion and the Equity Premium Puzzle*, Quarterly Journal of Economics; Barberis, N., M. Huang, and T. Santos, 2001, *Prospect Theory and Asset Prices*, Quarterly Journal of Economics, v 116(1), 1-53.

coefficients to demand these premiums.[30] In the years since, there have been many attempts to provide explanations for this puzzle:

1. Statistical artifact: The historical risk premium obtained by looking at U.S. data is biased upwards because of a survivor bias (induced by picking one of the most successful equity markets of the twentieth century). The true premium, it is argued, is much lower. This view is backed up by a study of large equity markets over the twentieth century, which concluded that the historical risk premium is closer to 4% than the 6% cited by Mehra and Prescott.[31] However, even the lower risk premium would still be too high, if we assumed reasonable risk aversion coefficients.
2. Disaster Insurance: A variation on the statistical artifact theme, albeit with a theoretical twist, is that the observed volatility in an equity market does not fully capture the potential volatility, which could include rare but disastrous events that reduce consumption and wealth substantially. Reitz, referenced earlier, argues that investments that have dividends that are proportional to consumption (as stocks do) should earn much higher returns than riskless investments to compensate for the possibility of a disastrous drop in consumption. Prescott and Mehra (1988) counter than the required drops in consumption would have to be of such a large magnitude to explain observed premiums that this solution is not viable. [32] Berkman, Jacobsen and Lee (2011) use data from 447 international political crises between 1918 and 2006 to create a crisis index and note that increases in the index increase equity risk premiums, with disproportionately large impacts on the industries most exposed to the crisis.[33]
3. Taxes: One possible explanation for the high equity returns in the period after the Second World War is the declining marginal tax rate during that period. McGrattan and Prescott (2001), for instance, provide a hypothetical illustration where a drop in the tax rate on dividends from 50% to 0% over 40 years would cause equity prices to rise about 1.8% more than the growth rate in GDP; adding the dividend yield to this expected price appreciation generates returns similar to

---

[30] Mehra, Rajnish, and Edward C.Prescott, 1985, *The Equity Premium: A Puzzle*, Journal of Monetary Economics, v15, 145–61. Using a constant relative risk aversion utility function and plausible risk aversion coefficients, they demonstrate the equity risk premiums should be much lower (less than 1%).

[31] Dimson, E., P. March and M. Staunton, 2002, *Triumph of the Optimists*, Princeton University Press.

[32] Mehra, R. and E.C. Prescott, 1988, *The Equity Risk Premium: A Solution?* Journal of Monetary Economics, v22, 133-136.

[33] Berkman, H., B. Jacobsen and J. Lee, 2011, *Time-varying Disaster Risk and Stock Returns*, Journal of Financial Economics, v101, 313-332

the observed equity risk premium.[34] In reality, though, the drop in marginal tax rates was much smaller and cannot explain the surge in equity risk premiums.

4. Alternative Preference Structures: There are some who argue that the equity risk premium puzzle stems from its dependence upon conventional expected utility theory to derive premiums. In particular, the constant relative risk aversion (CRRA) function used by Mehra and Prescott in their paper implies that if an investor is risk averse to variation in consumption across different states of nature at a point in time, he or she will also be equally risk averse to consumption variation across time. Epstein and Zin consider a class of utility functions that separate risk aversion (to consumption variation at a point in time) from risk aversion to consumption variation across time. They argue that individuals are much more risk averse when it comes to the latter and claim that this phenomenon explain the larger equity risk premiums.[35] Put in more intuitive terms, individuals will choose a lower and more stable level of wealth and consumption that they can sustain over the long term over a higher level of wealth and consumption that varies widely from period to period. Constantinides (1990) adds to this argument by noting that individuals become used to maintaining past consumption levels and that even small changes in consumption can cause big changes in marginal utility. The returns on stocks are correlated with consumption, decreasing in periods when people have fewer goods to consume (recessions, for instance); the additional risk explains the higher observed equity risk premiums.[36]
5. Myopic Loss Aversion: Myopic loss aversion refers to the finding in behavioral finance that the loss aversion already embedded in individuals becomes more pronounced as the frequency of their monitoring increases. Thus, investors who receive constant updates on equity values actually perceive more risk in equities, leading to higher risk premiums. The paper that we cited earlier by Benartzi and Thaler yields estimates of the risk premium very close to historical levels using a one-year time horizon for investors with plausible loss aversion characteristics (of about 2, which is backed up by the experimental research).

In conclusion, it is not quite clear what to make of the equity risk premium puzzle. It is true that historical risk premiums are higher than could be justified using conventional

---

[34] McGrattan, E.R., and E.C. Prescott. 2001, *Taxes, Regulations, and Asset Prices*, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=292522.

[35] Epstein, L.G., and S.E. Zin. 1991. *Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis*, Journal of Political Economy, v99, 263–286.

[36] Constantinides, G.M. 1990. *Habit Formation: A Resolution of the Equity Premium Puzzle*, Journal of Political Economy, v98, no. 3 (June):519–543.

utility models for wealth. However, that may tell us more about the dangers of using historical data and the failures of classic utility models than they do about equity risk premiums. In fact, the last decade of poor stock returns in the US and declining equity risk premiums may have made the equity risk premium puzzle less of a puzzle, since explaining a historical premium of 4% (the premium in 2011) is far easier than explaining a historical premium of 6% (the premium in 1999).

Estimation Approaches

There are three broad approaches used to estimate equity risk premiums. One is to survey subsets of investors and managers to get a sense of their expectations about equity returns in the future. The second is to assess the returns earned in the past on equities relative to riskless investments and use this historical premium as the expectation. The third is to attempt to estimate a forward-looking premium based on the market rates or prices on traded assets today; we will categorize these as implied premiums.

Survey Premiums

If the equity risk premium is what investors demand for investing in risky assets today, the most logical way to estimate it is to ask these investors what they require as expected returns. Since investors in equity markets number in the millions, the challenge is often finding a subset of investors that best reflects the aggregate market. In practice, se see surveys of investors, managers and even academics, with the intent of estimating an equity risk premium.

Investors

When surveying investors, we can take one of two tacks. The first is to focus on individual investors and get a sense of what they expect returns on equity markets to be in the future. The second is to direct the question of what equities will deliver as a premium at portfolio managers and investment professionals, with the rationale that their expectations should matter more in the aggregate, since they have the most money to invest.

a. Individual Investors: The oldest continuous index of investor sentiment about equities was developed by Robert Shiller in the aftermath of the crash of 1987 and has been updated since.[37] UBS/Gallup has also polled individual investors since 1996 about their optimism about future stock prices and reported a measure of investor

[37] The data is available at http://bit.ly/NcgTW7.

sentiment.[38] While neither survey provides a direct measure of the equity risk premium, they both yield broad measure of where investors expect stock prices to go in the near future. The Securities Industry Association (SIA) surveyed investors from 1999 to 2004 on the expected return on stocks and yields numbers that can be used to extract equity risk premiums. In the 2004 survey, for instance, they found that the median expected return across the 1500 U.S. investors they questioned was 12.8%, yielding a risk premium of roughly 8.3% over the treasury bond rate at that time.[39]

b. Institutional Investors/ Investment Professionals: Investors Intelligence, an investment service, tracks more than a hundred newsletters and categorizes them as bullish, bearish or neutral, resulting in a consolidated advisor sentiment index about the future direction of equities. Like the Shiller and UBS surveys, it is a directional survey that does not yield an equity risk premium. Merrill Lynch, in its monthly survey of institutional investors globally, explicitly poses the question about equity risk premiums to these investors. In its February 2007 report, for instance, Merrill reported an average equity risk premium of 3.5% from the survey, but that number jumped to 4.1% by March, after a market downturn.[40] As markets settled down in 2009, the survey premium has also settled back to 3.76% in January 2010. Through much of 2010, the survey premium stayed in a tight range (3.85% - 3.90%) but the premium climbed to 4.08% in the January 2012 update. In February 2014, the survey yielded a risk premium of 4.6%, though it may not be directly comparable to the earlier numbers because of changes in the survey.[41]

While survey premiums have become more accessible, very few practitioners seem to be inclined to use the numbers from these surveys in computations and there are several reasons for this reluctance:

1. Survey risk premiums are responsive to recent stock prices movements, with survey numbers generally increasing after bullish periods and decreasing after market decline. Thus, the peaks in the SIA survey premium of individual investors occurred in the bull market of 1999, and the more moderate premiums of 2003 and 2004 occurred after the market collapse in 2000 and 2001.
2. Survey premiums are sensitive not only to whom the question is directed at but how the question is asked. For instance, individual investors seem to have higher

[38] The data is available at http://www.ubs.com/us/en/wealth/misc/investor-watch.html

[39] See http://www.sifma.org/research/surveys.aspx. The 2004 survey seems to be the last survey done by SIA. The survey yielded expected stock returns of 10% in 2003, 13% in 2002, 19% in 2001, 33% in 2000 and 30% in 1999.

[40] See http://www.ml.com/index.asp?id=7695_8137_47928.

[41] Global Fund Manager Survey, Bank of America Merrill Lynch, February 2014.

(and more volatile) expected returns on equity than institutional investors and the survey numbers vary depending upon the framing of the question.[42]

3. In keeping with other surveys that show differences across sub-groups, the premium seems to vary depending on who gets surveyed. Kaustia, Lehtoranta and Puttonen (2011) surveyed 1,465 Finnish investment advisors and note that not only are male advisors more likely to provide an estimate but that their estimated premiums are roughly 2% lower than those obtained from female advisors, after controlling for experience, education and other factors.[43]
4. Studies that have looked at the efficacy of survey premiums indicate that if they have any predictive power, it is in the wrong direction. Fisher and Statman (2000) document the negative relationship between investor sentiment (individual and institutional) and stock returns.[44] In other words, investors becoming more optimistic (and demanding a larger premium) is more likely to be a precursor to poor (rather than good) market returns.

As technology aids the process, the number and sophistication of surveys of both individual and institutional investors will also increase. However, it is also likely that these survey premiums will be more reflections of the recent past rather than good forecasts of the future.

Managers

As noted in the first section, equity risk premiums are a key input not only in investing but also in corporate finance. The hurdle rates used by companies – costs of equity and capital – are affected by the equity risk premiums that they use and have significant consequences for investment, financing and dividend decisions. Graham and Harvey have been conducting annual surveys of Chief Financial Officers (CFOs) or companies for roughly the last decade with the intent of estimating what these CFOs think is a reasonable equity risk premium (for the next 10 years over the ten-year bond rate). In their March 2014 survey, they report an average equity risk premium of 3.73% across survey respondents, down slightly from the average premium of 4.27% a year earlier. The median premium in the March 2014 survey was 3.3%.[45]

---

[42] Asking the question "What do you think stocks will do next year?" generates different numbers than asking "What should the risk premium be for investing in stocks?"

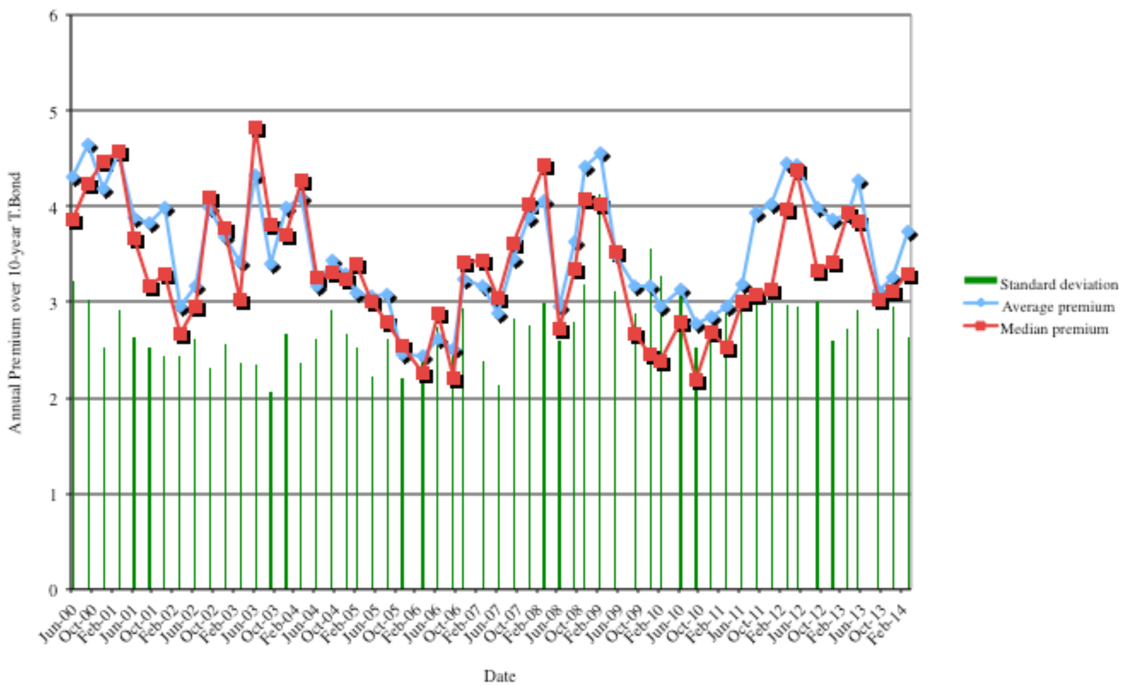
[43] Kaustia, M., A. Lehtoranta and V. Puttonen, 2011, *Sophistication and Gender Effects in Financial Advisers Expectations*, Working Paper, Aalto University.

[44] Fisher, K.L., and M. Statman, 2000, *Investor Sentiment and Stock Returns*, Financial Analysts Journal, v56, 16-23.

[45] Graham, J.R. and C.R. Harvey, 2014, *The Equity Risk Premium in 2014*, Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2422008 . See also Graham, J.R. and C.R. Harvey,

To get a sense of how these assessed equity risk premiums have behaved over time, we have graphed the average and median values of the premium and the cross sectional standard deviation in the estimates in each CFO survey, from 2001 to 2014, in Figure 2.

Figure 2: CFO Survey Premiums



Note the survey premium peak was in February 2009, right after the crisis, at 4.56% and had its lowest recording (2.5%) in September 2006. The average across all 14 years of surveys (more than 10,000 responses) was 3.54%, but the standard deviation in the survey responses did increase after the 2008 crisis.

Academics

Most academics are neither big players in equity markets, nor do they make many major corporate finance decisions. Notwithstanding this lack of real world impact, what they think about equity risk premiums may matter for two reasons. The first is that many of the portfolio managers and CFOs that were surveyed in the last two sub-sections received their first exposure to the equity risk premium debate in the classroom and may have been influenced by what was presented as the right risk premium in that setting. The

---

2009, *The Equity Risk Premium amid a Global Financial Crisis*, Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1405459.

second is that practitioners often offer academic work (textbooks and papers) as backing for the numbers that they use.

Welch (2000) surveyed 226 financial economists on the magnitude of the equity risk premium and reported interesting results. On average, economists forecast an average annual risk premium (arithmetic) of about 7% for a ten-year time horizon and 6-7% for one to five-year time horizons. As with the other survey estimates, there is a wide range on the estimates, with the premiums ranging from 2% at the pessimistic end to 13% at the optimistic end. Interestingly, the survey also indicates that economists believe that their estimates are higher than the consensus belief and try to adjust the premiums down to reflect that view.[46]

Fernandez (2010) examined widely used textbooks in corporate finance and valuation and noted that equity risk premiums varied widely across the books and that the moving average premium has declined from 8.4% in 1990 to 5.7% in 2010.[47] In a more recent survey, Fernandez, Aguirreamalloa and L. Corres (2011) compared both the level and standard deviation of equity risk premium estimates for analysts, companies and academics in the United States:[48]

Group	*Average Equity Risk Premium*	*Standard deviation in Equity Risk Premium estimates*
Academics	5.6%	1.6%
Analysts	5.0%	1.1%
Companies	5.5%	1.6%

The range on equity risk premiums in use is also substantial, with a low of 1.5% and a high of 15%, often citing the same sources. The same authors also report survey responses from the same groups (academics, analysts and companies) in 88 countries in 2014 and note that those in emerging markets use higher risk premiums (not surprisingly) than those in developed markets.[49]

---

[46] Welch, I., 2000, *Views of Financial Economists on the Equity Premium and on Professional Controversies*, Journal of Business, v73, 501-537.

[47] Fernandez, P., 2010, *The Equity Premium in 150 Textbooks*, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1473225. He notes that the risk premium actually varies within the book in as many as a third of the textbooks surveyed.

[48] Fernandez, P., J. Aguirreamalloa and L. Corres, 2011, Equity Premium used in 2011 for the USA by Analysts, Companies and Professors: A Survey, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1805852&rec=1&srcabs=1822182.

[49] Fernandez, P., P. Linares and I.F. Acin, 2014, Market Risk Premium used in 88 countries in 2014, A Survey with 8228 Answers, http://ssrn.com/abstract=2450452.

Historical Premiums

While our task is to estimate equity risk premiums in the future, much of the data we use to make these estimates is in the past. Most investors and managers, when asked to estimate risk premiums, look at historical data. In fact, the most widely used approach to estimating equity risk premiums is the historical premium approach, where the actual returns earned on stocks over a long time period is estimated, and compared to the actual returns earned on a default-free (usually government security). The difference, on an annual basis, between the two returns is computed and represents the historical risk premium. In this section, we will take a closer look at the approach.

Estimation Questions and Consequences

While users of risk and return models may have developed a consensus that historical premium is, in fact, the best estimate of the risk premium looking forward, there are surprisingly large differences in the actual premiums we observe being used in practice, with the numbers ranging from 3% at the lower end to 12% at the upper end. Given that we are almost all looking at the same historical data, these differences may seem surprising. There are, however, three reasons for the divergence in risk premiums: different time periods for estimation, differences in riskfree rates and market indices and differences in the way in which returns are averaged over time.

1. Time Period

Even if we agree that historical risk premiums are the best estimates of future equity risk premiums, we can still disagree about how far back in time we should go to estimate this premium. For decades, Ibbotson Associates was the most widely used estimation service, reporting stock return data and risk free rates going back to 1926,[50] and Duff and Phelps now provides the same service[51]. There are other less widely used databases that go further back in time to 1871 or even to 1792.[52]

While there are many analysts who use all the data going back to the inception date, there are almost as many analysts using data over shorter time periods, such as fifty, twenty or even ten years to come up with historical risk premiums. The rationale

[50] Ibbotson Stocks, Bonds, Bills and Inflation Yearbook (SBBI), 2011 Edition, Morningstar.

[51] Duff and Phelps, 2014 Valuation Handbook, Industry Cost of Capital.

[52] Siegel, in his book, Stocks for the Long Run, estimates the equity risk premium from 1802-1870 to be 2.2% and from 1871 to 1925 to be 2.9%. (Siegel, Jeremy J., Stocks for the Long Run, Second Edition, McGraw Hill, 1998). Goetzmann and Ibbotson estimate the premium from 1792 to 1925 to be 3.76% on an arithmetic average basis and 2.83% on a geometric average basis. Goetzmann. W.N. and R. G. Ibbotson, 2005, History and the Equity Risk Premium, Working Paper, Yale University. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=702341.

presented by those who use shorter periods is that the risk aversion of the average investor is likely to change over time, and that using a shorter and more recent time period provides a more updated estimate. This has to be offset against a cost associated with using shorter time periods, which is the greater noise in the risk premium estimate. In fact, given the annual standard deviation in stock returns[53] between 1928 and 2014 of 19.90% (approximated to 20%), the standard error associated with the risk premium estimate can be estimated in table 2 follows for different estimation periods:[54]

Table 2: Standard Errors in Historical Risk Premiums

Estimation Period	*Standard Error of Risk Premium Estimate*
5 years	$20\%/\sqrt{5} = 8.94\%$
10 years	$20\%/\sqrt{10} = 6.32\%$
25 years	$20\% / \sqrt{25} = 4.00\%$
50 years	$20\% / \sqrt{50} = 2.83\%$
80 years	$20\% / \sqrt{80} = 2.23\%$

Even using all of the entire data (about 85 years) yields a substantial standard error of 2.2%. Note that that the standard errors from ten-year and twenty-year estimates are likely to be almost as large or larger than the actual risk premium estimated. This cost of using shorter time periods seems, in our view, to overwhelm any advantages associated with getting a more updated premium.

What are the costs of going back even further in time (to 1871 or before)? First, the data is much less reliable from earlier time periods, when trading was lighter and record keeping more haphazard. Second, and more important, the market itself has changed over time, resulting in risk premiums that may not be appropriate for today. The U.S. equity market in 1871 more closely resembled an emerging market, in terms of volatility and risk, than a mature market. Consequently, using the earlier data may yield premiums that have little relevance for today's markets.

There are two other solutions offered by some researchers. The first is to break the annual data down into shorter return intervals – quarters or even months – with the intent of increasing the data points over any given time period. While this will increase

[53] For the historical data on stock returns, bond returns and bill returns check under "updated data" in http://www.damodaran.com.

[54] The standard deviation in annual stock returns between 1928 and 2014 is 19.90%; the standard deviation in the risk premium (stock return – bond return) is a little higher at 21.59%. These estimates of the standard error are probably understated, because they are based upon the assumption that annual returns are uncorrelated over time. There is substantial empirical evidence that returns are correlated over time, which would make this standard error estimate much larger. The raw data on returns is provided in Appendix 1.

the sample size, the effect on the standard error will be minimal.[55] The second is to use the entire data but to give a higher weight to more recent data, thus getting more updated premiums while preserving the data. While this option seems attractive, weighting more recent data will increase the standard error of the estimate. After all, using only the last ten years of data is an extreme form of time weighting, with the data during that period being weighted at one and the data prior to the period being weighted at zero.

2. Riskfree Security and Market Index

The second estimation question we face relates to the riskfree rate. We can compare the expected return on stocks to either short-term government securities (treasury bills) or long term government securities (treasury bonds) and the risk premium for stocks can be estimated relative to either. Given that the yield curve in the United States has been upward sloping for most of the last eight decades, the risk premium is larger when estimated relative to short term government securities (such as treasury bills) than when estimated against treasury bonds.

Some practitioners and a surprising number of academics (and textbooks) use the treasury bill rate as the riskfree rate, with the alluring logic that there is no price risk in a treasury bill, whereas the price of a treasury bond can be affected by changes in interest rates over time. That argument does make sense, but only if we are interested in a single period equity risk premium (say, for next year). If your time horizon is longer (say 5 or 10 years), it is the treasury bond that provides the more predictable returns.[56] Investing in a 6-month treasury bill may yield a guaranteed return for the next six months, but rolling over this investment for the next five years will create reinvestment risk. In contrast, investing in a ten-year treasury bond, or better still, a ten-year zero coupon bond will generate a guaranteed return for the next ten years.[57]

The riskfree rate chosen in computing the premium has to be consistent with the riskfree rate used to compute expected returns. Thus, if the treasury bill rate is used as the riskfree rate, the premium has to be the premium earned by stocks over that rate. If the treasury bond rate is used as the riskfree rate, the premium has to be estimated relative to that rate. For the most part, in corporate finance and valuation, the riskfree rate will be a

---

[55] If returns are uncorrelated over time, the variance in quarterly (monthly) risk premiums will be approximately one-quarter (one twelfth) the variance in annual risk premiums.

[56] For more on risk free rates, see Damodaran, A., 2008, *What is the riskfree rate?* Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1317436.

[57] There is a third choice that is sometimes employed, where the short term government security (treasury bills) is used as the riskfree rate and a "term structure spread" is added to this to get a normalized long term rate.

long-term default-free (government) bond rate and not a short-term rate. Thus, the risk premium used should be the premium earned by stocks over treasury bonds.

The historical risk premium will also be affected by how stock returns are estimated. Using an index with a long history, such as the Dow 30, seems like an obvious solution, but returns on the Dow may not be a good reflection of overall returns on stocks. In theory, at least, we would like to use the broadest index of stocks to compute returns, with two caveats. The first is that the index has to be market-weighted, since the overall returns on equities will be tilted towards larger market cap stocks. The second is that the returns should be free of survivor bias; estimating returns only on stocks that have survived that last 80 years will yield returns that are too high. Stock returns should incorporate those equity investments from earlier years that did not make it through the estimation period, either because the companies in question went bankrupt or were acquired.

Finally, there is some debate about whether the equity risk premiums should be computed using nominal returns or real returns. While the choice clearly makes a difference, if we estimate the return on stocks or the government security return standing alone, it is less of an issue, when computing equity risk premiums, where we look at the difference between the two values.

3. Averaging Approach

The final sticking point when it comes to estimating historical premiums relates to how the average returns on stocks, treasury bonds and bills are computed. The arithmetic average return measures the simple mean of the series of annual returns, whereas the geometric average looks at the compounded return[58]. Many estimation services and academics argue for the arithmetic average as the best estimate of the equity risk premium. In fact, if annual returns are uncorrelated over time, and our objective was to estimate the risk premium for the next year, the arithmetic average is the best and most unbiased estimate of the premium. There are, however, strong arguments that can be made for the use of geometric averages. First, empirical studies seem to indicate that returns on stocks are negatively correlated[59] over time. Consequently, the arithmetic

---

[58] The compounded return is computed by taking the value of the investment at the start of the period ($Value_0$) and the value at the end ($Value_N$), and then computing the following:

$$\text{Geometric Average} = \left(\frac{\text{Value}_N}{\text{Value}_0}\right)^{1/N} - 1$$

[59] In other words, good years are more likely to be followed by poor years, and vice versa. The evidence on negative serial correlation in stock returns over time is extensive, and can be found in Fama and French (1988). While they find that the one-year correlations are low, the five-year serial correlations are strongly

average return is likely to over state the premium. Second, while asset pricing models may be single period models, the use of these models to get expected returns over long periods (such as five or ten years) suggests that the estimation period may be much longer than a year. In this context, the argument for geometric average premiums becomes stronger. Indro and Lee (1997) compare arithmetic and geometric premiums, find them both wanting, and argue for a weighted average, with the weight on the geometric premium increasing with the time horizon.[60]

In closing, the averaging approach used clearly matters. Arithmetic averages will be yield higher risk premiums than geometric averages, but using these arithmetic average premiums to obtain discount rates, which are then compounded over time, seems internally inconsistent. In corporate finance and valuation, at least, the argument for using geometric average premiums as estimates is strong.

Estimates for the United States

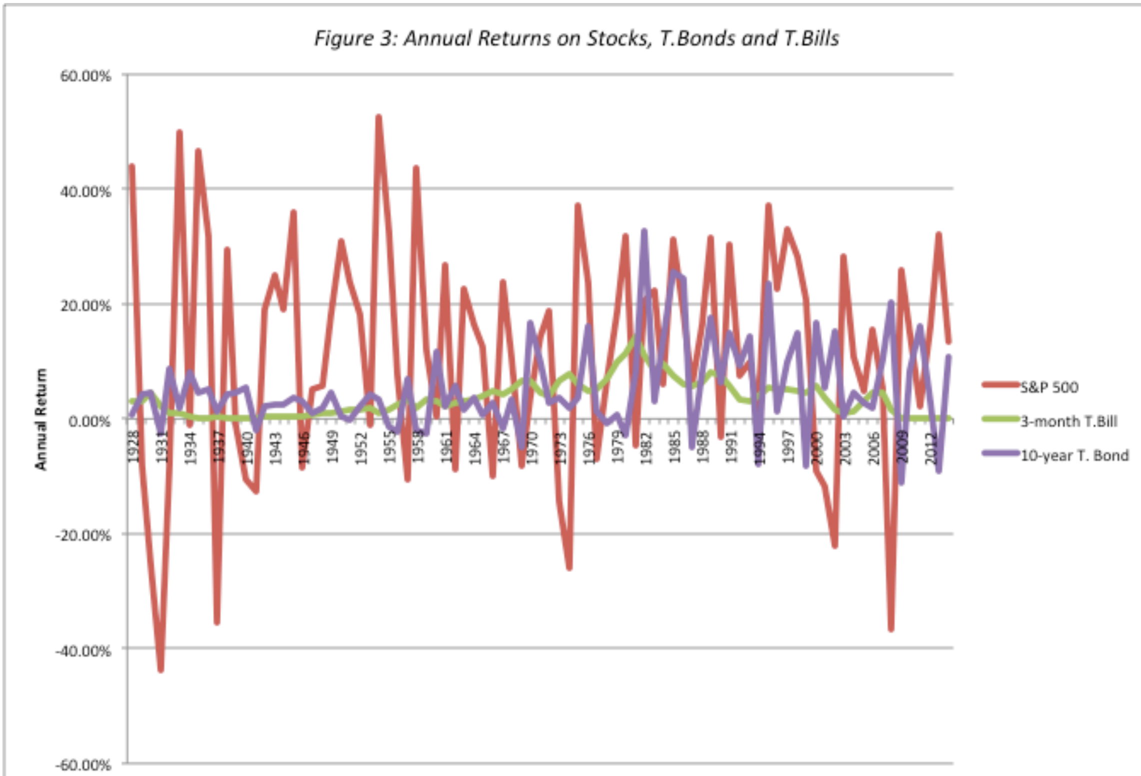
The questions of how far back in time to go, what riskfree rate to use and how to average returns (arithmetic or geometric) may seem trivial until you see the effect that the choices you make have on your equity risk premium. Rather than rely on the summary values that are provided by data services, we will use raw return data on stocks, treasury bills and treasury bonds from 1928 to 2014 to make this assessment.[61] In figure 3, we begin with a chart of the annual returns on stock, treasury bills and bonds for each year:

---

negative for all size classes. Fama, E.F. and K.R. French, 1992, *The Cross-Section of Expected Returns*, Journal of Finance, Vol 47, 427-466.

[60] Indro, D.C. and W. Y. Lee, 1997, *Biases in Arithmetic and Geometric Averages as Estimates of Long-run Expected Returns and Risk Premium*, Financial Management, v26, 81-90.

[61] The raw data for treasury rates is obtained from the Federal Reserve data archive (http://research.stlouisfed.org/fred2/) at the Fed site in St. Louis, with the 3-month treasury bill rate used for treasury bill returns and the 10-year treasury bond rate used to compute the returns on a constant maturity 10-year treasury bond. The stock returns represent the returns on the S&P 500. Appendix 1 provides the returns by year on stocks, bonds and bills, by year, from 1928 through the current year.



It is difficult to make much of this data other than to state the obvious, which is that stock returns are volatile, which is at the core of the demand for an equity risk premium in the first place. In table 3, we present summary statistics for stock, 3-month Treasury bill and ten-year Treasury bond returns from 1928 to 2014:

Table 3: Summary Statistics- U.S. Stocks, T.Bills and T. Bonds- 1928-2014

	Stocks	T. Bills	T. Bonds
Mean	11.53%	3.53%	5.28%
Standard Error	2.13%	0.33%	0.84%
Median	14.22%	3.11%	3.61%
Standard Deviation	19.90%	3.06%	7.83%
Kurtosis	2.98	3.82	4.39
Skewness	-0.41	0.96	0.94
Minimum	-43.84%	0.03%	-11.12%
Maximum	52.56%	14.30%	32.81%
25th percentile	-1.19%	1.01%	2.20%
75th percentile	26.11%	5.32%	8.93%

While U.S. equities have delivered much higher returns than treasuries over this period, they have also been more volatile, as evidenced both by the higher standard deviation in returns and by the extremes in the distribution. Using this table, we can take a first shot at estimating a risk premium by taking the difference between the average returns on stocks and the average return on treasuries, yielding a risk premium of 8.00% for stocks over T.Bills (11.53%-3.53%) and 6.25% for stocks over T.Bonds (11.53%-5.28%). Note, though, that these represent arithmetic average, long-term premiums for stocks over treasuries.

How much will the premium change if we make different choices on historical time periods, riskfree rates and averaging approaches? To answer this question, we estimated the arithmetic and geometric risk premiums for stocks over both treasury bills and bonds over different time periods in table 4, with standard errors reported in brackets below the arithmetic averages:

Table 4: Historical Equity Risk Premiums (ERP) –Estimation Period, Riskfree Rate and Averaging Approach

	Arithmetic Average		*Geometric Average*	
	Stocks - T. Bills	Stocks - T. Bonds	Stocks - T. Bills	Stocks - T. Bonds
1928-2014	8.00%	6.25%	6.11%	4.60%
	(2.17%)	(2.32%)		
1965-2014	6.19%	4.12%	4.84%	3.14%
	(2.42%	(2.74%)		
2005-2014	7.94%	4.06%	6.18%	2.73%
	(6.05%)	(8.65%)		

Note that even with only three slices of history considered, the premiums range from 2.73% to 8.00%, depending upon the choices made. If we take the earlier discussion about the "right choices" to heart, and use a long-term geometric average premium over the long-term rate as the risk premium to use in valuation and corporate finance, the equity risk premium that we would use would be 4.60%. The caveats that we would offer, though, are that this estimate comes with significant standard error and is reflective of time periods (such as 1920s and 1930s) when the U.S. equity market (and investors in it) had very different characteristics.

There have been attempts to extend the historical time period to include years prior to 1926 (the start of the Ibbotson database). Goetzmann and Jorion (1999) estimate the returns on stocks and bonds between 1792 and 1925 and report an arithmetic average premium, for stocks over bonds, of 2.76% and a geometric average premium of 2.83%.[62] The caveats about data reliability and changing market characteristics that we raised in an earlier section apply to these estimates.

---

[62] Jorion, Philippe and William N. Goetzmann, 1999, *Global Stock Markets in the Twentieth Century*, Journal of Finance, 54(3), 953-980.

There is one more troublesome (or at least counter intuitive) characteristic of historical risk premiums. The geometric average equity risk premium through the end of 2007 was 4.79%, higher than the 3.88% estimated though the end of 2008; in fact, every single equity risk premium number in this table would have been much higher, if we had stopped with 2007 as the last year. Adding the data for 2008, an abysmal year for stocks and a good year for bonds, lowers the historical premium dramatically, even when computed using a long period of history. In effect, the historical risk premium approach would lead investors to conclude, after one of worst stock market crisis in several decades, that stocks were less risky than they were before the crisis and that investors should therefore demand lower premiums. In contrast, adding the data for 2009, a good year for stocks (+25.94%) and a bad year for bonds (-11.12%) would have increased the equity risk premium from 3.88% to 4.29%. As a general rule, historical risk premiums will tend to rise when markets are buoyant and investors are less risk averse and will fall as markets collapse and investor fears rise.

Global Estimates

If it is difficult to estimate a reliable historical premium for the US market, it becomes doubly so when looking at markets with short, volatile and transitional histories. This is clearly true for emerging markets, where equity markets have often been in existence for only short time periods (Eastern Europe, China) or have seen substantial changes over the last few years (Latin America, India). It also true for many West European equity markets. While the economies of Germany, Italy and France can be categorized as mature, their equity markets did not share the same characteristics until recently. They tended to be dominated by a few large companies, many businesses remained private, and trading was thin except on a few stocks.

Notwithstanding these issues, services have tried to estimate historical risk premiums for non-US markets with the data that they have available. To capture some of the danger in this practice, Table 5 summarizes historical arithmetic average equity risk premiums for major non-US markets below for 1976 to 2001, and reports the standard error in each estimate:[63]

Table 5: Risk Premiums for non-US Markets: 1976- 2001

Country	*Weekly average*	*Weekly standard deviation*	*Equity Risk Premium*	*Standard error*
Canada	0.14%	5.73%	1.69%	3.89%

[63] Salomons, R. and H. Grootveld, 2003, *The equity risk premium: Emerging vs Developed Markets,* Emerging Markets Review, v4, 121-144.

France	0.40%	6.59%	4.91%	4.48%
Germany	0.28%	6.01%	3.41%	4.08%
Italy	0.32%	7.64%	3.91%	5.19%
Japan	0.32%	6.69%	3.91%	4.54%
UK	0.36%	5.78%	4.41%	3.93%
India	0.34%	8.11%	4.16%	5.51%
Korea	0.51%	11.24%	6.29%	7.64%
Chile	1.19%	10.23%	15.25%	6.95%
Mexico	0.99%	12.19%	12.55%	8.28%
Brazil	0.73%	15.73%	9.12%	10.69%

Before we attempt to come up with rationale for why the equity risk premiums vary across countries, it is worth noting the magnitude of the standard errors on the estimates, largely because the estimation period includes only 25 years. Based on these standard errors, we cannot even reject the hypothesis that the equity risk premium in each of these countries is zero, let alone attach a value to that premium.

If the standard errors on these estimates make them close to useless, consider how much more noise there is in estimates of historical risk premiums for some emerging market equity markets, which often have a reliable history of ten years or less, and very large standard deviations in annual stock returns. Historical risk premiums for emerging markets may provide for interesting anecdotes, but they clearly should not be used in risk and return models.

The survivor bias

Given how widely the historical risk premium approach is used, it is surprising that the flaws in the approach have not drawn more attention. Consider first the underlying assumption that investors' risk premiums have not changed over time and that the average risk investment (in the market portfolio) has remained stable over the period examined. We would be hard pressed to find anyone who would be willing to sustain this argument with fervor. The obvious fix for this problem, which is to use a more recent time period, runs directly into a second problem, which is the large noise associated with historical risk premium estimates. While these standard errors may be tolerable for very long time periods, they clearly are unacceptably high when shorter periods are used.

Even if there is a sufficiently long time period of history available, and investors' risk aversion has not changed in a systematic way over that period, there is a final problem. Markets such as the United States, which have long periods of equity market history, represent "survivor markets". In other words, assume that one had invested in the largest equity markets in the world in 1926, of which the United States was one.[64] In the period extending from 1926 to 2000, investments in many of the other equity markets would have earned much smaller premiums than the US equity market, and some of them would have resulted in investors earning little or even negative returns over the period. Thus, the survivor bias will result in historical premiums that are larger than expected premiums for markets like the United States, even assuming that investors are rational and factor risk into prices.

How can we mitigate the survivor bias? One solution is to look at historical risk premiums across multiple equity markets across very long time periods. In the most comprehensive attempt of this analysis, Dimson, Marsh and Staunton (2002, 2008) estimated equity returns for 17 markets and obtained both local and a global equity risk premium.[65] In their most recent update in 2015, they provide the risk premiums from 1900 to 2014 for 20 markets, with standard errors on each estimate (reported in table 6):[66]

Table 6: Historical Risk Premiums across Equity Markets – 1900 – 2014 (in %)

	Stocks minus Short term Governments				*Stocks minus Long term Governments*			
Country	*Geometric Mean*	*Arithmetic Mean*	*Standard Error*	*Standard Deviation*	*Geometric Mean*	*Arithmetic Mean*	*Standard Error*	*Standard Deviation*
Australia	6.6%	8.1%	1.6%	17.5%	5.6%	7.5%	1.9%	20.0%
Austria	5.5%	10.4%	3.5%	37.4%	2.5%	21.5%	14.4%	153.5%
Belgium	3.0%	5.4%	2.2%	23.9%	2.3%	4.4%	2.0%	21.1%
Canada	4.2%	5.6%	1.6%	16.9%	3.5%	5.1%	1.7%	18.2%
Denmark	3.1%	5.0%	1.9%	20.5%	2.0%	3.6%	1.7%	17.9%
Finland	5.9%	9.5%	2.8%	29.9%	5.1%	8.7%	2.8%	30.1%
France	6.1%	8.7%	2.3%	24.2%	3.0%	5.3%	2.1%	22.8%
Germany	6.0%	9.9%	3.0%	31.5%	5.0%	8.4%	2.7%	28.6%
Ireland	3.5%	5.8%	2.0%	21.3%	2.6%	4.5%	1.8%	19.6%
Italy	5.7%	9.5%	2.9%	31.6%	3.1%	6.5%	2.7%	29.5%

[64] Jorion, Philippe and William N. Goetzmann, 1999, *Global Stock Markets in the Twentieth Century*, Journal of Finance, 54(3), 953-980. They looked at 39 different equity markets and concluded that the US was the best performing market from 1921 to the end of the century. They estimated a geometric average premium of 3.84% across all of the equity markets that they looked at, rather than just the US and estimated that the survivor bias added 1.5% to the US equity risk premium (with arithmetic averages) and 0.9% with geometric averages.

[65] Dimson, E.,, P Marsh and M Staunton, 2002, *Triumph of the Optimists: 101 Years of Global Investment Returns,* Princeton University Press, NJ; Dimson, E.,, P Marsh and M Staunton, 2008, The Worldwide Equity Risk Premium: a smaller puzzle, Chapter 11 in the Handbook of the Equity Risk Premium, edited by R. Mehra, Elsevier.

[66] *Credit Suisse Global Investment Returns Sourcebook,* 2015, Credit Suisse/ London Business School. Summary data is accessible at the Credit Suisse website.

Japan	6.1%	9.3%	2.6%	27.7%	5.1%	9.1%	3.0%	32.6%
Netherlands	4.4%	6.5%	2.1%	22.5%	3.2%	5.6%	2.1%	22.3%
New Zealand	4.4%	5.9%	1.7%	18.1%	3.9%	5.5%	1.7%	17.9%
Norway	3.1%	5.9%	2.4%	26.1%	2.3%	5.3%	2.6%	27.7%
South Africa	6.3%	8.4%	2.0%	21.7%	5.4%	7.1%	1.8%	19.6%
Spain	3.4%	5.5%	2.0%	21.6%	1.9%	3.9%	1.9%	20.7%
Sweden	3.9%	5.9%	1.9%	20.5%	3.0%	5.3%	2.0%	21.5%
Switzerland	3.7%	5.3%	1.7%	18.7%	2.1%	3.6%	1.6%	17.5%
U.K.	4.3%	6.1%	1.8%	19.7%	3.7%	5.0%	1.6%	17.3%
U.S.	5.6%	7.5%	1.8%	19.6%	4.4%	6.5%	1.9%	20.7%
Europe	3.4%	5.2%	1.8%	19.3%	3.1%	4.4%	1.5%	16.1%
World-ex U.S.	3.6%	5.2%	1.7%	18.6%	2.8%	3.9%	1.4%	14.7%
World	4.3%	5.7%	1.6%	17.0%	3.2%	4.5%	1.5%	15.5%

In making comparisons of the numbers in this table to prior years, note that this database was modified in two ways: the world estimates are now weighted by market capitalization and the issue of survivorship bias has been dealt with frontally by incorporating the return histories of three markets (Austria, China and Russia) where equity investors would have lost their entire investment during the century. Note that the risk premiums, averaged across the markets, are lower than risk premiums in the United States. For instance, the geometric average risk premium for stocks over long-term government bonds, across the non-US markets, is only 2.8%, lower than the 4.4% for the US markets. The results are similar for the arithmetic average premium, with the average premium of 3.9% across non-US markets being lower than the 6.5% for the United States. In effect, the difference in returns captures the survivorship bias, implying that using historical risk premiums based only on US data will results in numbers that are too high for the future. Note that the "noise" problem persists, even with averaging across 20 markets and over 112 years. The standard error in the global equity risk premium estimate is 1.5%, suggesting that the range for the historical premium remains a large one.

Decomposing the historical equity risk premium

As the data to compute historical risk premiums has become richer, those who compute historical risk premiums have also become more creative, breaking down the historical risk premiums into its component parts, partly to understand the drivers of the premiums and partly to get better predictors for the future. Ibbotson and Chen (2013) started this process by breaking down the historical risk premium into four components:[67]

1. The income return is the return earned by stockholders from dividends and stock buybacks.
2. The second is the inflation rate during the estimation time period
3. The third is the growth rate in real earnings (earnings cleansed of inflation) during the estimation period
4. The change in PE ratio over the period, since an increase (decrease) in the PE ratio will raise (lower) the realized return on stocks during an estimation period.

Using the argument that the first three are sustainable and generated by "the productivity of corporations in the economy" and the fourth is not, they sum up the first three components to arrive at what they term a "supply-side" equity risk premium.

Following the same playbook, Dimson, Marsh and Staunton decompose the realized equity risk premium in each market into three components: the level of dividends, the growth in those dividends and the effects on stock price of a changing multiple for dividend (price to dividend ratio). For the United States, they attribute 1.67% of the overall premium of 5.59% (for stocks over treasury bills) to growth in real dividends and 0.57% to expansion in the price to dividend ratio. Of the global premium of 4.32%, 0.57% can be attributed to growth in dividends and 0.53% to increases in the price to dividend ratio.

While there is some value in breaking down a historical risk premium, notice that none of these decompositions remove the basic problems with historical risk premiums, which is that they are backward looking and noisy. Thus, a supply side premium has to come with all of the caveats that a conventional historical premium with the added noise created by the decomposition, i.e, measuring inflation and real earnings.

Historical Premium Plus

If we accept the proposition that historical risk premiums are the best way to estimate future risk premiums and also come to terms with the statistical reality that we need long time periods of history to get reliable estimates, we are trapped when it comes to estimating risk premiums in most emerging markets, where historical data is either non-existent or unreliable. Furthermore, the equity risk premium that we estimate becomes the risk premium that we use for all stocks within a market, no matter what their

---

[67] Ibbotson, R. and P. Chen, 2003, *Long-Run Stock Returns: Participating in the Real Economy*, Financial Analysts Journal, pp.88-98.

differences are on market capitalization and growth potential; in effect, we assume that the betas we use will capture differences in risk across companies.

In this section, we consider one way out of this box, where we begin with the US historical risk premium (4.60%) or the global premium from the DMS data (3.20%) as the base premium for a mature equity market and then build additional premiums for riskier markets or classes of stock. For the first part of this section, we stay within the US equity market and consider the practice of adjusting risk premiums for company-specific characteristics, with market capitalization being the most common example. In the second part, we extend the analysis to look at emerging markets in Asia, Latin American and Eastern Europe, and take a look at the practice of estimating country risk premiums that augment the US equity risk premium. Since many of these markets have significant exposures to political and economic risk, we consider two fundamental questions in this section. The first relates to whether there should be an additional risk premium when valuing equities in these markets, because of the country risk. As we will see, the answer will depend upon whether we think country risk is diversifiable or non-diversifiable, view markets to be open or segmented and whether we believe in a one-factor or a multi-factor model. The second question relates to estimating equity risk premiums for emerging markets. Depending upon our answer to the first question, we will consider several solutions.

Small cap and other risk premiums

In computing an equity risk premium to apply to all investments in the capital asset pricing model, we are essentially assuming that betas carry the weight of measuring the risk in individual firms or assets, with riskier investments having higher betas than safer investments. Studies of the efficacy of the capital asset pricing model over the last three decades have cast some doubt on whether this is a reasonable assumption, finding that the model understates the expected returns of stocks with specific characteristics; small market cap companies and companies low price to book ratios, in particular, seem to earn much higher returns than predicted by the CAPM. It is to counter this finding that many practitioners add an additional premium to the required returns (and costs of equity) of smaller market cap companies.

The CAPM and Market Capitalization

In one of very first studies to highlight the failure of the traditional capital asset pricing model to explain returns at small market cap companies, Banz (1981) looked returns on stocks from 1936-1977 and concluded that investing in the smallest companies (the bottom 20% of NYSE firms in terms of capitalization) would have generated about 6% more, after adjusting for beta risk, than larger cap companies.[68] In the years since, there has been substantial research on both the origins and durability of the small cap premium, with mixed conclusions. First, there is evidence of a small firm premium in markets outside the United States as well. Studies find small cap premiums of about 7% from 1955 to 1984 in the United Kingdom,[69] 8.8% in France and 3% in Germany,[70] and a premium of 5.1% for Japanese stocks between 1971 and 1988.[71] Dimson, March and Staunton (2015), in their updated assessment of equity risk premiums in global markets, also compute small cap premiums in 23 markets over long time periods (which range from 113 years for some markets to less for others). Of the 23 markets, small cap stocks have not outperformed the rest of the market in only Norway, Finland and the Netherlands; the small cap premium, over the long term, has been higher in developed markets than in emerging markets. Second, while the small cap premium has been persistent in US equity markets, it has also been volatile, with large cap stocks outperforming small cap stocks for extended periods. In figure 4, we look at the difference in returns between small cap (defined as bottom 10% of firms in terms of market capitalization) and all US stocks between 1927 and 2014; note that the premium was pronounced in the 1970s and disappeared for much of the 1980s.[72]

---

[68] Banz, R., 1981, *The Relationship between Return and Market Value of Common Stocks*, Journal of Financial Economics, v9.

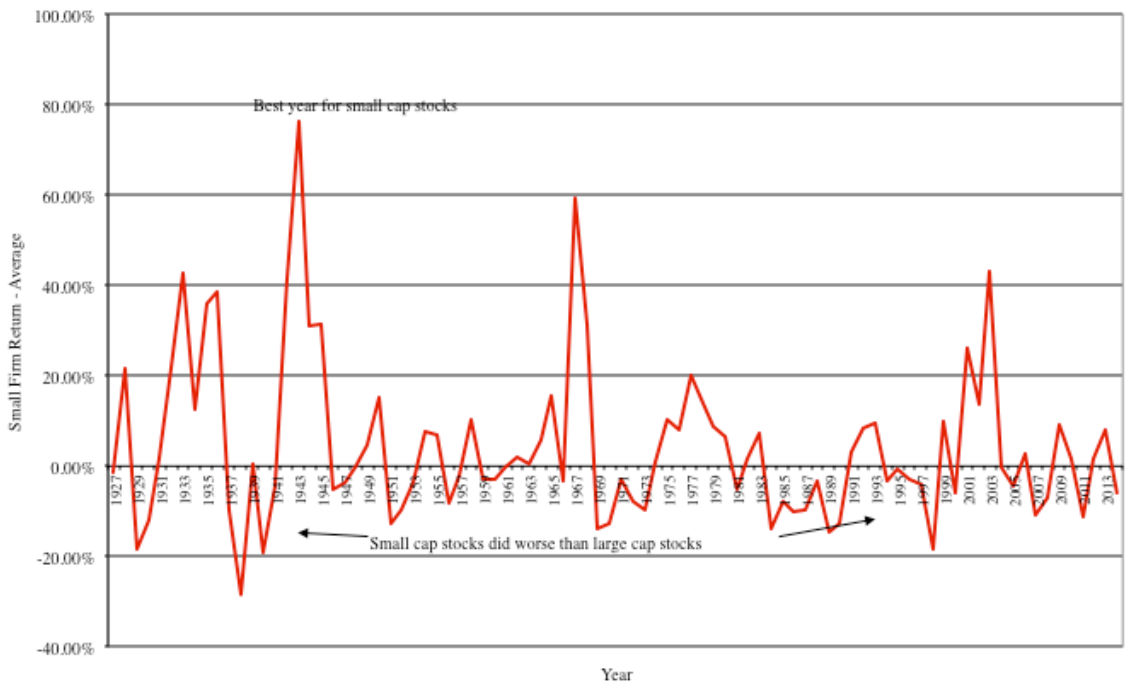
[69] Dimson, E. and P.R. Marsh, 1986, *Event Studies and the Size Effect: The Case of UK Press Recommendations*, Journal of Financial Economics, v17, 113-142.

[70] Bergstrom,G.L., R.D. Frashure and J.R. Chisholm, 1991, *The Gains from international small-company diversification* in Global Portfolios: Quantiative Strategies for Maximum Performance, Edited By R.Z. Aliber and B.R. Bruce, Business One Irwin, Homewood.

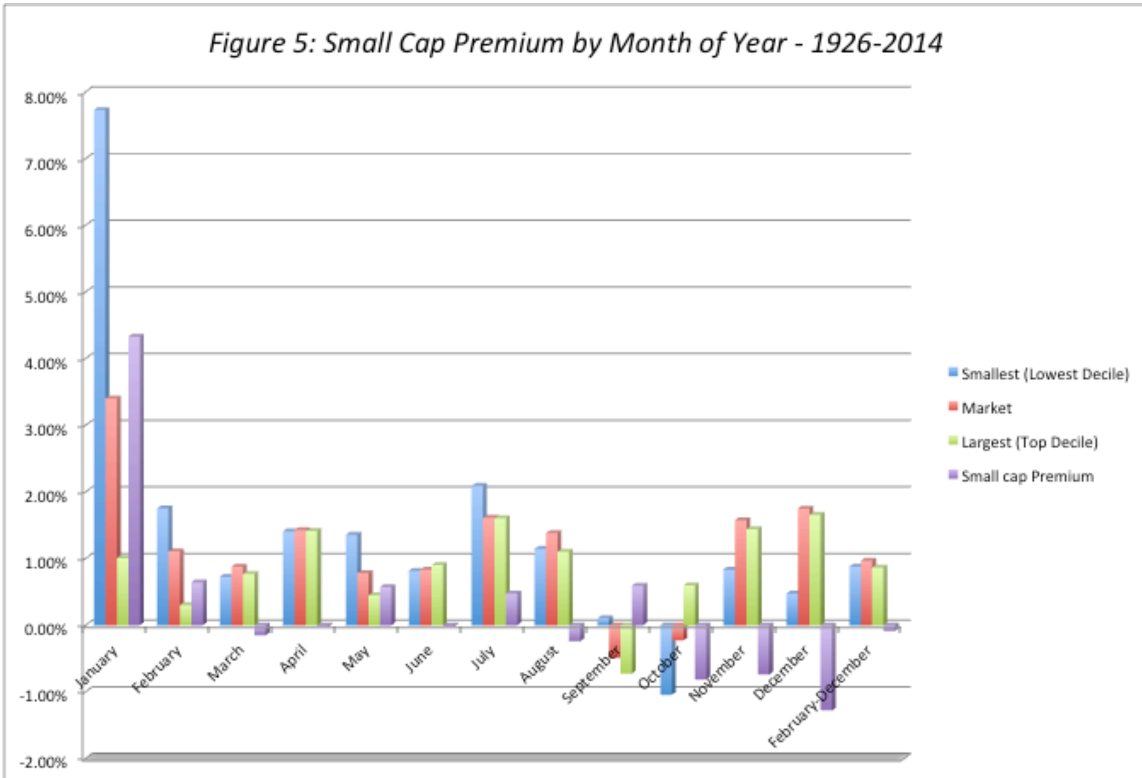
[71] Chan, L.K., Y. Hamao, and J. Lakonishok, 1991, *Fundamentals and Stock Returns in Japan*, Journal of Finance. v46. 1739-1789.

[72] The raw data for this table is obtained from Professor Ken French's website at Dartmouth. These premiums are based on value weighted portfolios. If equally weighted portfolios are used, the small cap premium is larger (almost 10.71%).

Figure 4: Small Firm Premium over time- 1927 -2014



The average premium for stocks in the smallest companies, in terms of market capitalization, between 1926 and 2013 was 4.33%, but the standard error in that estimate is 1.96%. Third, much of the premium is generated in one month of the year: January. As Figure 5 shows, eliminating that month from our calculations would essentially dissipate the entire small stock premium. That would suggest that size itself is not the source of risk, since small firms in January remain small firms in the rest of the year, but that the small firm premium, if it exists, comes from some other risk that is more pronounced or prevalent in January than in the rest of the year.



Source: Raw data from Ken French

Finally, a series of studies have argued that market capitalization, by itself, is not the reason for excess returns but that it is a proxy for other ignored risks such as illiquidity and poor information.

In summary, while the empirical evidence supports the notion that small cap stocks have earned higher returns after adjusting for beta risk than large cap stocks, it is not as conclusive, nor as clean as it was initially thought to be. The argument that there is, in fact, no small cap premium and that we have observed over time is just an artifact of history cannot be rejected out of hand.

The Small Cap Premium

If we accept the notion that there is a small cap premium, there are two ways in which we can respond to the empirical evidence that small market cap stocks seem to earn higher returns than predicted by the traditional capital asset pricing model. One is to view this as a market inefficiency that can be exploited for profit: this, in effect, would require us to load up our portfolios with small market cap stocks that would then proceed to deliver higher than expected returns over long periods. The other is to take the excess returns as evidence that betas are inadequate measures of risk and view the additional returns are compensation for the missed risk. The fact that the small cap premium has endured for as long as it has suggests that the latter is the more reasonable path to take.

If CAPM betas understate the true risk of small cap stocks, what are the solutions? The first is to try and augment the model to reflect the missing risk, but this would require being explicit about this risk. For instance, there are models that include additional factors for illiquidity and imperfect information that claim to do better than the CAPM in predicting future returns. The second and simpler solution that is adopted by many practitioners is to add a premium to the expected return (from the CAPM) of small cap stocks. To arrive at this premium, analysts look at historical data on the returns on small cap stocks and the market, adjust for beta risk, and attribute the excess return to the small cap effect. As we noted earlier, using the data from 1926-2014, we would estimate a small cap premium of 4.33%. Duff and Phelps present a richer set of estimates, where the premiums are computed for stocks in 25 different size classes (with size measured on eight different dimensions including market capitalization, book value and net income). Using the Fama/French data, we present excess returns for firms broken down by ten market value classes in Table 7, with the standard error for each estimate.

Table 7: Excess Returns by Market Value Class: US Stocks from 1927 – 2014

Excess Return = Return on Portfolio – Return on Market

Decile	*Average*	*Standard Error*	*Maximum*	*Minimum*
Smallest	4.33%	1.96%	76.28%	-28.42%
2	1.63%	1.14%	41.25%	-17.96%
3	1.47%	0.77%	41.98%	-13.54%
4	0.64%	0.55%	15.56%	-7.33%
5	0.05%	0.53%	11.63%	-16.05%
6	-0.01%	0.51%	15.21%	-14.01%
7	-0.51%	0.55%	7.48%	-19.50%
8	-1.50%	0.81%	11.20%	-29.42%
9	-2.13%	1.02%	21.96%	-36.09%
Largest	-3.98%	1.56%	31.29%	-65.57%

Note that the market capitalization effect shows up at both extremes – the smallest firms earn higher returns than expected whereas the largest firms earn lower returns than expected. The small firm premium is statistically significant only for the lowest and three highest size deciles.

Perils of the approach

While the small cap premium may seem like a reasonable way of dealing with the failure of the CAPM to capture the risk in smaller companies, there are significant costs to using the approach.

a. Standard Error on estimates: One of the dangers we noted with using historical risk premiums is the high standard error in our estimates. This danger is magnified when we look at sub-sets of stocks, based on market capitalization or any other characteristic, and extrapolate past returns. The standard errors on the small cap premiums that are estimated are likely to be significant, as is evidenced in table 7.
b. Small versus Large Cap: At least in its simplest form, the small cap premium adjustment requires us to divide companies into small market companies and the rest of the market, with stocks falling on one side of the line having much higher required returns (and costs of equity) than stocks falling on the other side.
c. Understanding Risk: Even in its more refined format, where the required returns are calibrated to market cap, using small cap premiums allows analysts to evade basic questions about what it is that makes smaller cap companies riskier, and whether these factors may vary across companies.
d. Small cap companies become large cap companies over time: When valuing companies, we attach high growth rates to revenues, earnings and value over time. Consequently, companies that are small market cap companies now grow to become large market cap companies over time. Consistency demands that we adjust the small cap premium as we go further into a forecast period.
e. Other risk premiums: Using a small cap premium opens the door to other premiums being used to augment expected returns. Thus, we could adjust expected returns upwards for stocks with price momentum and low price to book ratios, reflecting the excess returns that these characteristics seem to deliver, at least on paper. Doing so will deliver values that are closer to market prices, across assets, but undercuts the rationale for intrinsic valuation, i.e., finding market mistakes.

There is another reason why we are wary about adjusting costs of equity for a small cap effect. If, as is the practice now, we add a small cap premium of between 4% to 5% to the cost of equity of small companies, without attributing this premium to any specific risk factor, we are exposed to the risk of double counting risk. For instance, assume that the small cap premium that we have observed over the last few decades is attributable to the lower liquidity (and higher transactions costs) of trading small cap stocks. Adding that premium on to the discount rate will reduce the estimated values of small cap and private businesses. If we attach an illiquidity discount to this value, we are double counting the effect of illiquidity.

The small cap premium is firmly entrenched in practice, with analysts generally adding on 4% to 5% to the conventional cost of equity for small companies, with the definition of small shifting from analyst to analyst. Even if you believe that small cap companies are more exposed to market risk than large cap ones, this is an extremely sloppy and lazy way of dealing with that risk, since risk ultimately has to come from something fundamental (and size is not a fundamental factor). Thus, if you believe that small cap stocks are more prone to failure or distress, it behooves you to measure that risk directly and incorporate it into the cost of equity. If it is illiquidity that is at the heart of the small cap premium, then you should be measuring liquidity risk and incorporating it into the cost of equity and you certainly should not be double counting the risk by first incorporating a small cap premium into the discount rate and then applying an illiquidity discount to value.

The question of whether there is a small cap premium ultimately is not a theoretical one but a practical one. While those who incorporate a small cap premium justify the practice with the historical data, we will present a more forward-looking approach, where we use market pricing of small capitalization stocks to see if the market builds in a small cap premium, later in this paper.

Country Risk Premiums

As both companies and investors get used to the reality of a global economy, they have also been forced to confront the consequences of globalization for equity risk premiums and hurdle rates. Should an investor putting his money in Indian stocks demand a higher risk premium for investing in equities that one investing in German stocks? Should a US consumer product company investing in Brazil demand the same hurdle rates for its Brazilian investments as it does for its US investments? In effect, should we demand one global equity risk premium that we use for investments all over the world or should we use higher equity risk premiums in some markets than in others?

The arguments for no country risk premium

Is there more risk in investing in a Malaysian or Brazilian stock than there is in investing in the United States? The answer, to most, seems to be obviously affirmative, with the solution being that we should use higher equity risk premiums when investing in riskier emerging markets. There are, however, three distinct and different arguments offered against this practice.

1. Country risk is diversifiable

In the risk and return models that have developed from conventional portfolio theory, and in particular, the capital asset pricing model, the only risk that is relevant for purposes of estimating a cost of equity is the market risk or risk that cannot be diversified away. The key question in relation to country risk then becomes whether the additional risk in an emerging market is diversifiable or non-diversifiable risk. If, in fact, the additional risk of investing in Malaysia or Brazil can be diversified away, then there should be no additional risk premium charged. If it cannot, then it makes sense to think about estimating a country risk premium.

But diversified away by whom? Equity in a publicly traded Brazilian, or Malaysian, firm can be held by hundreds or even thousands of investors, some of whom may hold only domestic stocks in their portfolio, whereas others may have more global exposure. For purposes of analyzing country risk, we look at the marginal investor – the investor most likely to be trading on the equity. If that marginal investor is globally diversified, there is at least the potential for global diversification. If the marginal investor does not have a global portfolio, the likelihood of diversifying away country risk declines substantially. Stulz (1999) made a similar point using different terminology.[73] He differentiated between segmented markets, where risk premiums can be different in each market, because investors cannot or will not invest outside their domestic markets, and open markets, where investors can invest across markets. In a segmented market, the marginal investor will be diversified only across investments in that market, whereas in an open market, the marginal investor has the opportunity (even if he or she does not take it) to invest across markets. It is unquestionable that investors today in most markets have more opportunities to diversify globally than they did three decades ago, with international mutual funds and exchange traded funds, and that many more of them take advantage of these opportunities. It is also true still that a significant home bias exists in most investors' portfolios, with most investors over investing in their home markets.

Even if the marginal investor is globally diversified, there is a second test that has to be met for country risk to be diversifiable. All or much of country risk should be country specific. In other words, there should be low correlation across markets. Only then will the risk be diversifiable in a globally diversified portfolio. If, on the other hand, the returns across countries have significant positive correlation, country risk has a market risk component, is not diversifiable and can command a premium. Whether

---

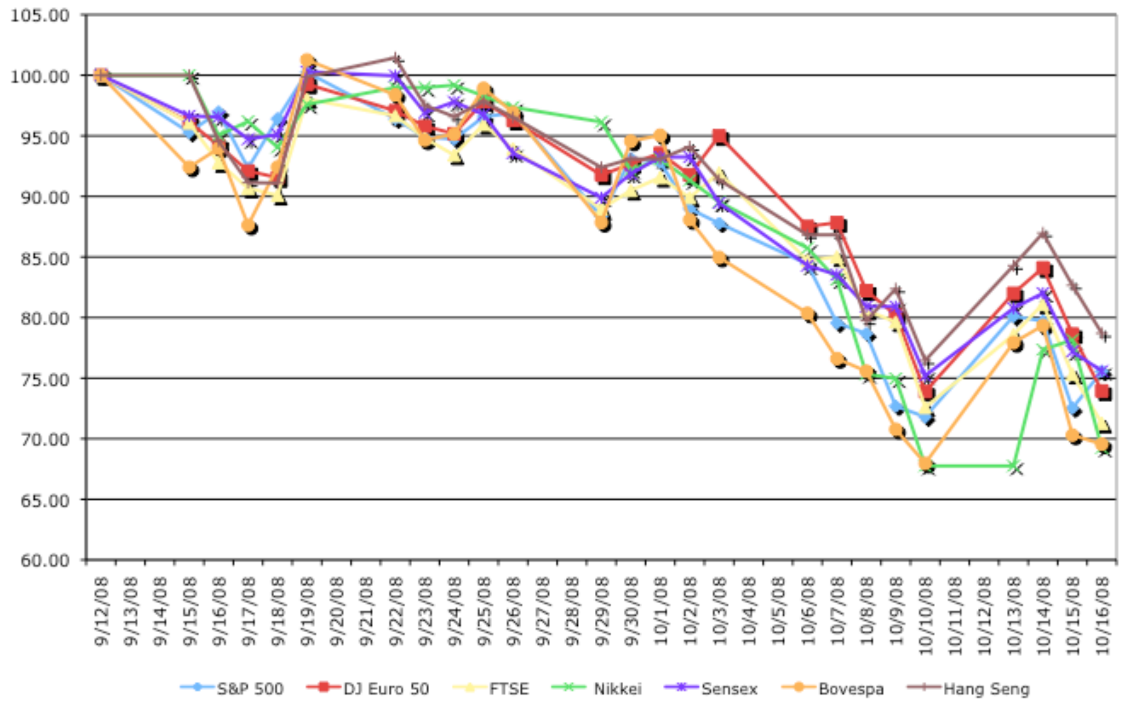
[73] Stulz, R.M., *Globalization, Corporate finance, and the Cost of Capital,* Journal of Applied Corporate Finance, v12. 8-25.

returns across countries are positively correlated is an empirical question. Studies from the 1970s and 1980s suggested that the correlation was low, and this was an impetus for global diversification.[74] Partly because of the success of that sales pitch and partly because economies around the world have become increasingly intertwined over the last decade, more recent studies indicate that the correlation across markets has risen. The correlation across equity markets has been studied extensively over the last two decades and while there are differences, the overall conclusions are as follows:

1. The correlation across markets has increased over time, as both investors and firms have globalized. Yang, Tapon and Sun (2006) report correlations across eight, mostly developed markets between 1988 and 2002 and note that the correlation in the 1998-2002 time period was higher than the correlation between 1988 and 1992 in every single market; to illustrate, the correlation between the Hong Kong and US markets increased from 0.48 to 0.65 and the correlation between the UK and the US markets increased from 0.63 to 0.82.[75] In the global returns sourcebook, from Credit Suisse, referenced earlier for historical risk premiums for different markets, the authors estimate the correlation between developed and emerging markets between 1980 and 2013, and note that it has increased from 0.57 in 1980 to 0.88 in 2013.
2. The correlation across equity markets increases during periods of extreme stress or high volatility.[76] This is borne out by the speed with which troubles in one market, say Russia, can spread to a market with little or no obvious relationship to it, say Brazil. The contagion effect, where troubles in one market spread into others is one reason to be skeptical with arguments that companies that are in multiple emerging markets are protected because of their diversification benefits. In fact, the market crisis in the last quarter of 2008 illustrated how closely bound markets have become, as can be seen in figure 6:

[74] Levy, H. and M. Sarnat, 1970, *International Diversification of Investment Portfolios*, American Economic Review 60(4), 668-75.
[75] Yang, Li , Tapon, Francis and Sun, Yiguo, 2006, *International correlations across stock markets and industries: trends and patterns 1988-2002*, Applied Financial Economics, v16: 16, 1171-1183
[76] Ball, C. and W. Torous, 2000, *Stochastic correlation across international stock markets*, Journal of Empirical Finance. v7, 373-388.

Figure 6: The globalization of risk



Between September 12, 2008 and October 16, 2008, markets across the globe moved up and down together, with emerging markets showing slightly more volatility.

3. The downside correlation increases more than upside correlation: In a twist on the last point, Longin and Solnik (2001) report that it is not high volatility per se that increases correlation, but downside volatility. Put differently, the correlation between global equity markets is higher in bear markets than in bull markets.[77]
4. Globalization increases exposure to global political uncertainty, while reducing exposure to domestic political uncertainty: In the most direct test of whether we should be attaching different equity risk premiums to different countries due to systematic risk exposure, Brogaard, Dai, Ngo and Zhang (2014) looked at 36 countries from 1991-2010 and measured the exposure of companies in these countries to global political uncertainty and domestic political uncertainty.[78] They find that the costs of capital of companies in integrated markets are more highly

[77] Longin, F. and B. Solnik, 2001, *Extreme Correlation of International Equity Markets*, Journal of Finance, v56 , pg 649-675.
[78] Brogaard, J., L. Dai, P.T.H. Ngo, B. Zhuang, 2014, *The World Price of Political Uncertainty*, SSRN #2488820.

influenced by global uncertainty (increasing as uncertainty increases) and those in segmented markets are more highly influenced by domestic uncertainty.[79]

2. A Global Capital Asset Pricing Model

The other argument against adjusting for country risk comes from theorists and practitioners who believe that the traditional capital asset pricing model can be adapted fairly easily to a global market. In their view, all assets, no matter where they are traded, should face the same global equity risk premium, with differences in risk captured by differences in betas. In effect, they are arguing that if Malaysian stocks are riskier than US stocks, they should have higher betas and expected returns.

While the argument is reasonable, it flounders in practice, partly because betas do not seem capable of carry the weight of measuring country risk.

1. If betas are estimated against local indices, as is usually the case, the average beta within each market (Brazil, Malaysia, US or Germany) has to be one. Thus, it would be mathematically impossible for betas to capture country risk.
2. If betas are estimated against a global equity index, such as the Morgan Stanley Capital Index (MSCI), there is a possibility that betas could capture country risk but there is little evidence that they do in practice. Since the global equity indices are market weighted, it is the companies that are in developed markets that have higher betas, whereas the companies in small, very risky emerging markets report low betas. Table 8 reports the average beta estimated for the ten largest market cap companies in Brazil, India, the United States and Japan against the MSCI.

Table 8: Betas against MSCI – Large Market Cap Companies

Country	*Average Beta (against local index)*	*Average Beta (against MSCI)*
India	0.97	0.83
Brazil	0.98	0.81
United States	0.96	1.05
Japan	0.94	1.03

[a] The betas were estimated using two years of weekly returns from January 2006 to December 2007 against the most widely used local index (Sensex in India, Bovespa in Brazil, S&P 500 in the US and the Nikkei in Japan) and the MSCI using two years of weekly returns.

The emerging market companies consistently have lower betas, when estimated against global equity indices, than developed market companies. Using these betas with a global equity risk premium will lead to lower costs of equity for emerging market companies than developed market companies. While there are creative fixes

[79] The implied costs of capital for companies in the 36 countries were computed and related to global political uncertainty, measured using the US economic policy uncertainty index, and to domestic political uncertainty, measured using domestic national elections.

that practitioners have used to get around this problem, they seem to be based on little more than the desire to end up with higher expected returns for emerging market companies.[80]

3. *Country risk is better reflected in the cash flows*

The essence of this argument is that country risk and its consequences are better reflected in the cash flows than in the discount rate. Proponents of this point of view argue that bringing in the likelihood of negative events (political chaos, nationalization and economic meltdowns) into the expected cash flows effectively risk adjusts the cashflows, thus eliminating the need for adjusting the discount rate.

This argument is alluring but it is wrong. The expected cash flows, computed by taking into account the possibility of poor outcomes, is not risk adjusted. In fact, this is exactly how we should be calculating expected cash flows in any discounted cash flow analysis. Risk adjustment requires us to adjust the expected cash flow further for its risk, i.e. compute certainty equivalent cash flows in capital budgeting terms. To illustrate why, consider a simple example where a company is considering making the same type of investment in two countries. For simplicity, let us assume that the investment is expected to deliver $ 90, with certainty, in country 1 (a mature market); it is expected to generate $ 100 with 90% probability in country 2 (an emerging market) but there is a 10% chance that disaster will strike (and the cash flow will be $0). The expected cash flow is $90 on both investments, but only a risk neutral investor would be indifferent between the two. A risk averse investor would prefer the investment in the mature market over the emerging market investment, and would demand a premium for investing in the emerging market.

In effect, a full risk adjustment to the cash flows will require us to go through the same process that we have to use to adjust discount rates for risk. We will have to estimate a country risk premium, and use that risk premium to compute certainty equivalent cash flows.[81]

The arguments for a country risk premium

There are elements in each of the arguments in the previous section that are persuasive but none of them is persuasive enough.

---

[80] There are some practitioners who multiply the local market betas for individual companies by a beta for that market against the US. Thus, if the beta for an Indian chemical company is 0.9 and the beta for the Indian market against the US is 1.5, the global beta for the Indian company will be 1.35 (0.9*1.5). The beta for the Indian market is obtained by regressing returns, in US dollars, for the Indian market against returns on a US index (say, the S&P 500).

[81] In the simple example above, this is how it would work. Assume that we compute a country risk premium of 3% for the emerging market to reflect the risk of disaster. The certainty equivalent cash flow on the investment in that country would be $90/1.03 = $87.38.

- Investors have become more globally diversified over the last three decades and portions of country risk can therefore be diversified away in their portfolios. However, the significant home bias that remains in investor portfolios exposes investors disproportionately to home country risk, and the increase in correlation across markets has made a portion of country risk into non-diversifiable or market risk.
- As stocks are traded in multiple markets and in many currencies, it is becoming more feasible to estimate meaningful global betas, but it also is still true that these betas cannot carry the burden of capturing country risk in addition to all other macro risk exposures.
- Finally, there are certain types of country risk that are better embedded in the cash flows than in the risk premium or discount rates. In particular, risks that are discrete and isolated to individual countries should be incorporated into probabilities and expected cash flows; good examples would be risks associated with nationalization or related to acts of God (hurricanes, earthquakes etc.).

After you have diversified away the portion of country risk that you can, estimated a meaningful global beta and incorporated discrete risks into the expected cash flows, you will still be faced with residual country risk that has only one place to go: the equity risk premium.

There is evidence to support the proposition that you should incorporate additional country risk into equity risk premium estimates in riskier markets:

1. Historical equity risk premiums: Donadelli and Prosperi (2011) look at historical risk premiums in 32 different countries (13 developed and 19 emerging markets) and conclude that emerging market companies had both higher average returns and more volatility in these returns between 1988 and 2010 (see table 9).

Table 9: Historical Equity Risk Premiums (Monthly) by Region

Region	*Monthly ERP*	*Standard deviation*
Developed Markets	0.62%	4.91%
Asia	0.97%	7.56%
Latin America	2.07%	8.18%
Eastern Europe	2.40%	15.66%
Africa	1.41%	6.03%

While we remain cautious about using historical risk premiums over short time periods (and 22 years is short in terms of stock market history), the evidence is consistent with the argument that country risk should be incorporated into a larger equity risk premium.[82]

2. Survey premiums: Earlier in the paper, we referenced a paper by Fernandez et al (2014) that surveyed academics, analysts and companies in 82 countries on equity risk premiums. The reported average premiums vary widely across markets and are higher for riskier emerging markets, as can be seen in table 10.

Table 10: Survey Estimates of Equity Risk Premium: By Region

Region	Number	Average	Median
Africa	11	10.14%	9.85%
Developed Markets	20	5.44%	5.29%
Eastern Europe	15	8.29%	8.25%
Emerging Asia	12	8.33%	8.08%
EU Troubled	7	8.36%	8.31%
Latin America	15	9.45%	9.39%
Middle East	8	7.14%	6.79%
Grand Total	**88**	**7.98%**	**7.82%**

Again, while this does not conclusively prove that country risk commands a premium, it does indicate that those who do valuations in emerging market countries seem to act like it does. Ultimately, the question of whether country risk matters and should affect the equity risk premium is an empirical one, not a theoretical one, and for the moment, at least, the evidence seems to suggest that you should incorporate country risk into your discount rates. This could change as we continue to move towards a global economy, with globally diversified investors and a global equity market, but we are not there yet.

Estimating a Country Risk Premium

If country risk is not diversifiable, either because the marginal investor is not globally diversified or because the risk is correlated across markets, we are then left with the task of measuring country risk and considering the consequences for equity risk premiums. In this section, we will consider three approaches that can be used to estimate country risk premiums, all of which build off the historical risk premiums estimated in the last section. To approach this estimation question, let us start with the basic proposition that the risk premium in any equity market can be written as:

---

[82] Donadelli, M. and L. Prosperi, 2011, *The Equity Risk Premium: Empirical Evidence from Emerging Markets*, Working Paper, http://ssrn.com/abstract=1893378.

Equity Risk Premium = Base Premium for Mature Equity Market + Country Risk Premium

The country premium could reflect the extra risk in a specific market. This boils down our estimation to estimating two numbers – an equity risk premium for a mature equity market and the additional risk premium, if any, for country risk. To estimate a mature market equity risk premium, we can look at one of two numbers. The first is the historical risk premium that we estimated for the United States, which yielded 4.60% as the geometric average premium for stocks over treasury bonds from 1928 to 2014. If we do this, we are arguing that the US equity market is a mature market, and that there is sufficient historical data in the United States to make a reasonable estimate of the risk premium. The other is the average historical risk premium across 20 equity markets, approximately 3.3%, that was estimated by Dimson et al (see earlier reference), as a counter to the survivor bias that they saw in using the US risk premium. Consistency would then require us to use this as the equity risk premium, in every other equity market that we deem mature; the equity risk premium in January 2015 would be 4.60% in Germany, France and the UK, for instance. For markets that are not mature, however, we need to measure country risk and convert the measure into a country risk premium, which will augment the mature market premium.

Measuring Country Risk

There are at least three measures of country risk that we can use. The first is the sovereign rating attached to a country by ratings agencies. The second is to subscribe to services that come up with broader measures of country risk that explicitly factor in the economic, political and legal risks in individual countries. The third is go with a market-based measure such as the volatility in the country's currency or markets.

i. Sovereign Ratings

One of the simplest and most accessible measures of country risk is the rating assigned to a country's debt by a ratings agency (S&P, Moody's and Fitch, among others, all provide country ratings). These ratings measure default risk (rather than equity risk) but they are affected by many of the factors that drive equity risk – the stability of a country's currency, its budget and trade balances and political uncertainty, among other variables[83].

[83] The process by which country ratings are obtained in explained on the S&P web site at http://www.ratings.standardpoor.com/criteria/index.htm.

To get a measure of country ratings, consider six countries – Germany, Brazil, China, India, Russia and Greece. In January 2015, the Moody's ratings for the countries are summarized in table 11:

Table 11: Sovereign Ratings in January 2015 – Moody's

Country	*Foreign Currency Rating*	*Local Currency Rating*
Brazil	Baa2	Baa2
China	Aa3	Aa3
Germany	Aaa	Aaa
Greece	Caa1	Caa1
India	Baa3	Baa3
Russia	Baa2	Baa2

What do these ratings tell us? First, the local currency and foreign currency ratings are identical for all of the countries on the list. There are a few countries (not on this list) where the two ratings diverge, and when they do, the local currency ratings tend to be higher (or at worst equal to) the foreign currency ratings for most countries, because a country should be in a better position to pay off debt in the local currency than in a foreign currency. Second, at least based on Moody's assessments in 2015, Germany is the safest company in this group, followed by China, Russia, Brazil, India and Greece, in that order. Third, ratings do change over time. In fact, Brazil's rating has risen from B1 in 2001 to its current rating of Baa2, reflecting both strong economic growth and a more robust political system. Appendix 2 contains the current ratings – local currency and foreign currency – for the countries that are tracked by Moody's in January 2015.[84]

While ratings provide a convenient measure of country risk, there are costs associated with using them as the only measure. First, ratings agencies often lag markets when it comes to responding to changes in the underlying default risk. The ratings for India, according to Moody's, were unchanged from 2004 to 2007, though the Indian economy grew at double-digit rates over that period. Similarly, Greece's ratings did not plummet until the middle of 2011, though their financial problems were visible well before that time. Second, the ratings agency focus on default risk may obscure other risks that could still affect equity markets. For instance, rising commodity (and especially oil) prices pushed up the ratings for commodity supplying countries (like Russia), even

[84] In a disquieting reaction to the turmoil of the market crisis in the last quarter of 2008, Moody's promoted the notion that Aaa countries were not all created equal and slotted these countries into three groups – resistant Aaa (the stongest), resilient Aaa (weaker but will probably survive intact) and vulnerable Aaa (likely to face additional default risk.

though there was little improvement in the rest of the economy. Finally, not all countries have ratings; much of sub-Saharan Africa, for instance, is unrated.

ii. Country Risk Scores

Rather than focus on just default risk, as rating agencies do, some services have developed numerical country risk scores that take a more comprehensive view of risk. These risk scores are often estimated from the bottom-up by looking at economic fundamentals in each country. This, of course, requires significantly more information and, as a consequence, most of these scores are available only to commercial subscribers.

The Political Risk Services (PRS) group, for instance, considers political, financial and economic risk indicators to come up with a composite measure of risk (ICRG) for each country that ranks from 0 to 100, with 0 being highest risk and 100 being the lowest risk.[85] Appendix 3 classifies countries based on composite country risk measures from the PRS Group in January 2014.[86] Harvey (2005) examined the efficacy of these scores and found that they were correlated with costs of capital, but only for emerging market companies.

The Economist, the business newsmagazine, also operates a country risk assessment unit that measures risk from 0 to 100, with 0 being the least risk and 100 being the most risk. In September 2008, Table 12 the following countries were ranked as least and most risky by their measure:

---

[85] The PRS group considers three types of risk – political risk, which accounts for 50% of the index, financial risk, which accounts for 25%, and economic risk, which accounts for the balance. While this table is dated, updated numbers are available for a hefty price. We have used the latest information in the public domain. Some university libraries have access to the updated data. While we have not updated the numbers, out of concerns about publishing proprietary data, you can get the latest PRS numbers by paying $99 on their website (http://www.prsgroup.com).

[86] Harvey, C.R., *Country Risk Components, the Cost of Capital, and Returns in Emerging Markets*, Working paper, Duke University. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=620710.

Table 12: Country Risk Scores – The Economist

Economist.com rankings

Country risk

Selected countries and territories, September 2008 (except where noted)

Least risky

Rank		Score*
1	Switzerland †	12
2	Finland **	14
	Norway **	14
	Sweden ††	14
5	Canada **	17
	Denmark †	17
	Netherlands §	17
8	Germany ††	18
9	Austria **	19
	France ††	19
11	Belgium ††	20
12	Singapore	21
13	Japan **	23
14	Ireland #	24
	Britain	24
	United States †	24

Most risky

Rank		Score
120	Zimbabwe	86
119	Iraq	80
118	Sudan	76
117	Myanmar	75
116	Nicaragua	69
115	Jamaica	68
114	Kenya	66
113	Cuba	64
112	Cambodia	62
111	Côte d'Ivoire	61
	Ecuador	61
	Pakistan	61
	Venezuela	61
	Vietnam	61
106	Syria	60

*Out of 100, with higher numbers indicating more risk. Scores are based on indicators from three categories: currency risk, sovereign debt risk and banking risk.

† May 2008; ** July 2008; †† June 2008; § August 2008; # February 2008

In fact, comparing the PRS and Economist measures of country risk provides some insight into the problems with using their risk measures. The first is that the measures may be internally consistent but are not easily comparable across different services. The Economist, for instance, assigns its lowest scores to the safest countries whereas PRS assigns the highest scores to these countries. The second is that, by their very nature, a significant component of these measures have to be black boxes to prevent others from replicating them at no cost. Third, the measures are not linear and the services do not claim that they are; a country with a risk score of 60 in the Economist measure is not twice as risky as a country with a risk score of 30.

iii. Market-based Measures

To those analysts who feel that ratings agencies are either slow to respond to changes in country risk or take too narrow a view of risk, there is always the alternative of using market based measures.

- Bond default spread: We can compute a default spread for a country if it has bonds that are denominated in currencies such as the US dollar, Euro or Yen, where there is a riskfree rate to compare it to. In January 2015, for instance, a 10-year US dollar denominated bond issued by the Brazilian government had a yield to maturity of 3.87%, giving it a default spread of 1.70% over the 10-year US treasury bond rate (2.17%), as of the same time.
- Credit Default Swap Spreads: In the last few years, credit default swaps (CDS) markets have developed, allowing us to obtain updated market measures of default risk in different entities. In particular, there are CDS spreads for countries (governments) that yield measures of default risk that are more updated and precise, at least in some cases, than bond default spreads.[87] Table 13 summarizes the CDS spreads for all countries where a CDS spread was available, in January 2015:

Table 13: Credit Default Swap Spreads (in basis points)– January 2015

Country	Moody's rating	CDS Spread	CDS Spread adj for US	Country	Moody's rating	CDS Spread	CDS Spread adj for US	Country	Moody's rating	CDS Spread	CI Spr adj U
Abu Dhabi	Aa2	1.43%	1.12%	Hungary	Ba1	2.64%	2.33%	Poland	A2	1.46%	1.1
Argentina	Caa1	83.48%	83.17%	Iceland	Baa3	2.27%	1.96%	Portugal	Ba1	3.09%	2.7
Australia	Aaa	0.97%	0.66%	India	Baa3	2.64%	2.33%	Qatar	Aa2	1.57%	1.2
Austria	Aaa	0.81%	0.50%	Indonesia	Baa3	2.82%	2.51%	Romania	Baa3	2.23%	1.9
Bahrain	Baa2	3.18%	2.87%	Ireland	Baa1	1.26%	0.95%	Russia	Baa2	5.63%	5.3
Belgium	Aa3	1.20%	0.89%	Israel	A1	0.42%	0.11%	Saudi Arabia	Aa3	1.39%	1.0
Brazil	Baa2	3.17%	2.86%	Italy	Baa2	2.34%	2.03%	Slovakia	A2	1.32%	1.0
Bulgaria	Baa2	2.99%	2.68%	Japan	A1	1.55%	1.24%	Slovenia	Ba1	2.14%	1.8
Chile	Aa3	1.77%	1.46%	Kazakhstan	Baa2	4.16%	3.85%	South Africa	Baa2	2.96%	2.6
China	Aa3	1.78%	1.47%	Korea	Aa3	1.17%	0.86%	Spain	Baa2	1.79%	1.4
Colombia	Baa2	2.57%	2.26%	Latvia	Baa1	1.92%	1.61%	Sweden	Aaa	0.65%	0.3
Costa Rica	Ba1	3.58%	3.27%	Lebanon	B2	4.69%	4.38%	Switzerland	Aaa	0.72%	0.4
Croatia	Ba1	3.65%	3.34%	Lithuania	Baa1	1.88%	1.57%	Thailand	Baa1	1.91%	1.6
Cyprus	B3	6.35%	6.04%	Malaysia	A3	2.15%	1.84%	Tunisia	Ba3	3.38%	3.0
Czech Republic	A1	1.25%	0.94%	Mexico	A3	2.05%	1.74%	Turkey	Baa3	2.77%	2.4
Denmark	Aaa	0.79%	0.48%	Morocco	Ba1	2.55%	2.24%	Uganda	B1	0.31%	0.0
Egypt	Caa1	3.56%	3.25%	Netherlands	Aaa	0.78%	0.47%	Ukraine	Caa3	15.74%	15.
Estonia	A1	1.20%	0.89%	New Zealand	Aaa	1.01%	0.70%	UAE	Aa2	1.54%	1.2
Finland	Aaa	0.81%	0.50%	Norway	Aaa	0.61%	0.30%	United Kingdom	Aa1	0.77%	0.4
France	Aa1	1.22%	0.91%	Pakistan	Caa1	10.41%	10.10%	United States	Aaa	0.31%	0.0
Germany	Aaa	0.74%	0.43%	Panama	Baa2	2.09%	1.78%	Venezuela	Caa1	18.06%	17.
Greece	Caa1	10.76%	10.45%	Peru	A3	2.23%	1.92%	Vietnam	B1	3.15%	2.8

[87] The spreads are usually stated in US dollar or Euro terms.

Hong Kong	Aa1	1.12%	0.81%	Philippines	Baa2	1.98%	1.67%

Source: Bloomberg
Spreads are for 10-year US $ CDS.

In January 2015, for instance, the CDS market yielded a spread of 3.17% for the Brazilian Government, higher than the 1.70% that we obtained from the 10-year dollar denominated Brazilian bond. However, the CDS market does have some counterparty risk exposure and other risk exposures that are incorporated into the spreads. In fact, there is no country with a zero CDS spread, indicating either that there is no entity with default risk or that the CDS spread is not a pure default spread. To counter that problem, we netted the US CDS spread of 0.31% from each country's CDS to get a modified measure of country default risk.[88] Using this approach for Brazil, for instance, yields a netted CDS spread of 2.86% for the country.

- Market volatility: In portfolio theory, the standard deviation in returns is generally used as the proxy for risk. Extending that measure to emerging markets, there are some analysts who argue that the best measure of country risk is the volatility in local stock prices. Stock prices in emerging markets will be more volatile that stock prices in developed markets, and the volatility measure should be a good indicator of country risk. While the argument makes intuitive sense, the practical problem with using market volatility as a measure of risk is that it is as much a function of the underlying risk as it is a function of liquidity. Markets that are risky and illiquid often have low volatility, since you need trading to move stock prices. Consequently, using volatility measures will understate the risk of emerging markets that are illiquid and overstate the risk of liquid markets.

Market-based numbers have the benefit of constant updating and reflect the points of view of investors at any point in time. However, they also are also afflicted with all of the problems that people associate with markets – volatility, mood shifts and at times, irrationality. They tend to move far more than the other two measures – sovereign ratings and country risk scores – sometimes for good reasons and sometimes for no reason at all.

b. Estimating Country Risk Premium (for Equities)

How do we link a country risk measure to a country risk premium? In this section, we will look at three approaches. The first uses default spreads, based upon

---

[88] If we assume that there is default risk in the US, we would subtract the default spread associated with this risk from the 0.67% first, before netting the value against other CDS spreads. Thus, if the default spread for the US is 0.15%, we would subtract out only 0.52% (0.67% - 0.15%) from each country's CDS spread to get to a corrected default spread for that country.

country bonds or ratings, whereas the latter two use equity market volatility as an input in estimating country risk premiums.

1. Default Spreads

The simplest and most widely used proxy for the country risk premium is the default spread that investors charge for buying bonds issued by the country. This default spread can be estimated in one of three ways.

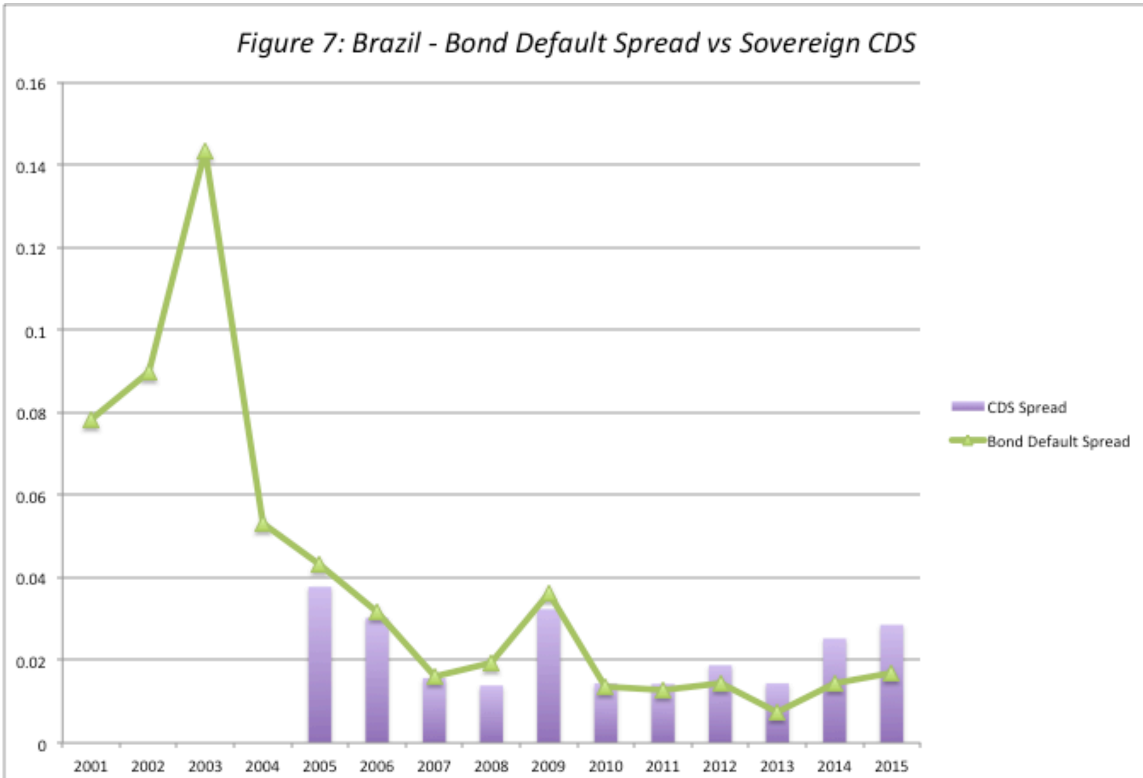
a. Current Default Spread on Sovereign Bond or CDS market: As we noted in the last section, the default spread comes from either looking at the yields on bonds issued by the country in a currency where there is a default free bond yield to which it can be compared or spreads in the CDS market.[89] With the 10-year US dollar denominated Brazilian bond that we cited as an example in the last section, the default spread would have amounted to 1.70% in January 2015: the difference between the interest rate on the Brazilian bond and a treasury bond of the same maturity. The netted CDS market spread on the same day for the default spread was 2.86%. Bekaert, Harvey, Lundblad and Siegel (2014) break down the sovereign bond default spread into four components, including global economic conditions, country-specific economic factors, sovereign bond liquidity and policial risk, and find that it is the political risk component that best explain money flows into and out of the country equity markets.[90]

b. Average (Normalized) spread on bond: While we can make the argument that the default spread in the dollar denominated is a reasonable measure of the default risk in Brazil, it is also a volatile measure. In figure 7, we have graphed the yields on the dollar denominated ten-year Brazilian Bond and the U.S. ten-year treasury bond and highlighted the default spread (as the difference between the two yields) from January 2000 to January 2015. In the same figure, we also show the 10-year CDS spreads from 2005 to 2015,[91] the spreads have also changed over time but move with the bond default spreads.

[89] You cannot compare interest rates across bonds in different currencies. The interest rate on a peso bond cannot be compared to the interest rate on a dollar denominated bond.

[90] Bekaert, G., C.R. Harvey, C.T. Lundblad and S. Siegel, 2014, *Political Risk Spreads*, Journal of International Business Studies, v45, 471-493.

[91] Data for the sovereign CDS market is available only from the last part of 2004.



Note that the bond default spread widened dramatically during 2002, mostly as a result of uncertainty in neighboring Argentina and concerns about the Brazilian presidential elections.[92] After the elections, the spreads decreased just as quickly and continued on a downward trend through the middle of last year. Since 2004, they have stabilized, with a downward trend; they spiked during the market crisis in the last quarter of 2008 but have settled back into pre-crisis levels. Given this volatility, a reasonable argument can be made that we should consider the average spread over a period of time rather than the default spread at the moment. If we accept this argument, the normalized default spread, using the average spreads over the last 5 years of data would be 1.65% (bond default spread) or 1.99% (CDS spread). Using this approach makes sense only if the economic fundamentals of the country have not changed significantly (for the better or worse) during the period but will yield misleading values, if there have been structural shifts in the economy. In 2008, for instance, it would have made sense to use averages over time for a country like Nigeria, where oil price movements created volatility in spreads over time, but not for countries like China and India, which saw their economies expand and mature dramatically over the period or Venezuela, where government capriciousness

[92] The polls throughout 2002 suggested that Lula Da Silva who was perceived by the market to be a leftist would beat the establishment candidate. Concerns about how he would govern roiled markets and any poll that showed him gaining would be followed by an increase in the default spread.

made operating private businesses a hazardous activity (with a concurrent tripling in default spreads). In fact, the last year has seen a spike in the Brazilian default spread, partly the result of another election and partly because of worries about political corruption and worse in large Brazilian companies.

<u>c. Imputed or Synthetic Spread</u>: The two approaches outlined above for estimating the default spread can be used only if the country being analyzed has bonds denominated in US dollars, Euros or another currency that has a default free rate that is easily accessible. Most emerging market countries, though, do not have government bonds denominated in another currency and some do not have a sovereign rating. For the first group (that have sovereign rating but no foreign currency government bonds), there are two solutions. If we assume that countries with the similar default risk should have the same sovereign rating, we can use the typical default spread for other countries that have the same rating as the country we are analyzing and dollar denominated or Euro denominated bonds outstanding. Thus, Bulgaria, with a Baa2 rating, would be assigned the same default spread as Brazil, which also has Baa2 rating, and dollar denominated bonds and CDS prices from which we can extract a default spread. For the second group, we are on even more tenuous grounds. Assuming that there is a country risk score from the Economist or PRS for the country, we could look for other countries that are rated and have similar scores and assign the default spreads that these countries face. For instance, we could assume that Cuba and Cameroon, which fall within the same score grouping from PRS, have similar country risk; this would lead us to attach Cuba's rating of Caa1 to Cameroon (which is not rated) and to use the same default spread (based on this rating) for both countries.

In table 14, we have estimated the typical default spreads for bonds in different sovereign ratings classes in January 2015. One problem that we had in obtaining the numbers for this table is that relatively few emerging markets have dollar or Euro denominated bonds outstanding. Consequently, there were some ratings classes where there was only one country with data and several ratings classes where there were none. To mitigate this problem, we used spreads from the CDS market, referenced in the earlier section. We were able to get default spreads for 65 countries, categorized by rating class, and we averaged the spreads across multiple countries in the same ratings class.[93] An alternative approach to estimating default spread is to assume that sovereign ratings are

[93] There were thirteen Baa2 rated countries, with ten-year CDS spreads, in January 2015. The average spread across the these countries is 2.68%. We noticed wide variations across countries in the same ratings class, and no discernible trend when compared to the January 2014 averages. Consequently, we decided to use the same default spreads that we used last year.

comparable to corporate ratings, i.e., a Ba1 rated country bond and a Ba1 rated corporate bond have equal default risk. In this case, we can use the default spreads on corporate bonds for different ratings classes. Table 14 summarizes the typical default spreads by sovereign rating class in January 2015, and compares it to the default spreads for similar corporate ratings.

Table 14: Default Spreads by Ratings Class – Sovereign vs. Corporate in January 2015

Moody's rating	*Sovereign Bonds/CDS*	Corporate Bonds
Aaa/AAA	0.00%	0.42%
Aa1/AA+	0.40%	0.60%
Aa2/AA	0.50%	0.78%
Aa3/AA-	0.60%	0.87%
A1/A+	0.70%	0.96%
A2/A	0.85%	0.97%
A3/A-	1.20%	1.10%
Baa1/BBB+	1.60%	1.36%
Baa2/BBB	1.90%	1.67%
Baa3/BBB-	2.20%	2.22%
Ba1/BB+	2.50%	2.61%
Ba2/BB	3.00%	2.97%
Ba3/BB-	3.60%	3.33%
B1/B+	4.50%	3.74%
B2/B	5.50%	4.10%
B3/B-	6.50%	4.45%
Caa1/ CCC+	7.50%	4.86%
Caa2/CCC	9.00%	7.50%
Caa3/ CCC-	10.00%	10.00%

Note that the corporate bond spreads, at least in January 2015, were slightly larger than the sovereign spreads for the higher ratings classes, converge for the intermediate ratings and widen again at the lowest ratings. Using this approach to estimate default spreads for Brazil, with its rating of Baa2 would result in a spread of 1.90% (1.67%), if we use sovereign spreads (corporate spreads). These spreads are down from post-crisis levels at the end of 2008 but are still larger than the actual spreads on Brazilian sovereign bonds in early 2014.

Figure 8 depicts the alternative approaches to estimating default spreads for four countries, Brazil, China, India and Poland, in early 2015:

Figure 8: Approaches for estimating Sovereign Default Spreads

Estimating a default spread for a country or sovereign entity

Market Based estimates

Sovereign Bond spread
1. Find a bond issued by the country, denominated in US$ or Euros.
2. Compute the default spread by comparing to US treasury bond (if US $) or German Euro bond (if Euros).

CDS Market
1. Find a 10-year CDS for the country (if one exists)
2. Net out US CDS
2. This is your default spread.

Country	Sovereign Bond Yield	Currency	Risk free rate	Default Spread	CDS Spread (net of US)
Brazil	3.87%	US $	2.17%	1.70%	2.86%
China	NA	NA	NA	NA	1.47%
India	NA	NA	NA	NA	2.33%
Poland	1.70%	Euro	0.50%	1.20%	1.15%

Rating/Risk score based estimates
Step 1: Find a sovereign rating (local currency) for the country (on Moody's or S&P)
Step 2: Look up the default spread for that rating in the lookup table below:

Moody's rating	*Sovereign Bonds/CDS*
Aaa/AAA	0.00%
Aa1/AA+	0.40%
Aa2/AA	0.50%
Aa3/AA-	0.60%
A1/A+	0.70%
A2/A	0.85%
A3/A-	1.20%
Baa1/BBB+	1.60%
Baa2/BBB	1.90%
Baa3/BBB-	2.20%
Ba1/BB+	2.50%
Ba2/BB	3.00%
Ba3/BB-	3.60%
B1/B+	4.50%
B2/B	5.50%
B3/B-	6.50%
Caa1/ CCC+	7.50%
Caa2/CCC	9.00%
Caa3/ CCC-	10.00%

Country	Moody's Rating	Default Spread
Brazil	Baa2	1.90%
China	Aa3	0.60%
India	Baa3	2.20%
Poland	A2	0.85%

With some countries, without US-dollar (or Euro) denominated sovereign bonds or CDS spreads, you don't have a choice since the only estimate of the default spread comes from the sovereign rating. With other countries, such as Brazil, you have multiple estimates of the default spreads: 1.70% from the dollar denominated bond, 3.17% from the CDS spread, 2.86% from the netted CDS spread and 1.90% from the sovereign rating look up table (table 14). You could choose one of these approaches and stay consistent over time or average across them.

Analysts who use default spreads as measures of country risk typically add them on to both the cost of equity and debt of every company traded in that country. Thus, the cost of equity for an Indian company, estimated in U.S. dollars, will be 2.2% higher than the cost of equity of an otherwise similar U.S. company, using the January 2015 measure of the default spread, based upon the rating. In some cases, analysts add the default spread to the U.S. risk premium and multiply it by the beta. This increases the cost of equity for high beta companies and lowers them for low beta firms.[94]

94 In a companion paper, I argue for a separate measure of company exposure to country risk called lambda that is scaled around one (just like beta) that is multiplied by the country risk premium to estimate the cost

While many analysts use default spreads as proxies for country risk, the evidence for its use is still thin. Abuaf (2011) examines ADRs from ten emerging markets and relates the returns on these ADRs to returns on the S&P 500 (which yields a conventional beta) and to the CDS spreads for the countries of incorporation. He finds that ADR returns as well as multiples (such as PE ratios) are correlated with movement in the CDS spreads over time and argues for the addition of the CDS spread (or some multiple of it) to the costs of equity and capital to incorporate country risk.[95]

2. Relative Equity Market Standard Deviations

There are some analysts who believe that the equity risk premiums of markets should reflect the differences in equity risk, as measured by the volatilities of these markets. A conventional measure of equity risk is the standard deviation in stock prices; higher standard deviations are generally associated with more risk. If you scale the standard deviation of one market against another, you obtain a measure of relative risk. For instance, the relative standard deviation for country X (against the US) would be computed as follows:

$$\text{Relative Standard Deviation}_{\text{Country X}} = \frac{\text{Standard Deviation}_{\text{Country X}}}{\text{Standard Deviation}_{\text{US}}}$$

If we assume a linear relationship between equity risk premiums and equity market standard deviations, and we assume that the risk premium for the US can be computed (using historical data, for instance) the equity risk premium for country X follows:

$$\text{Equity risk premium}_{\text{Country X}} = \text{Risk Premum}_{\text{US}} * \text{Relative Standard Deviation}_{\text{Country X}}$$

Assume, for the moment, that you are using an equity risk premium for the United States of 5.75%. The annualized standard deviation in the S&P 500 in two years preceding January 2014, using weekly returns, was 10.85%, whereas the standard deviation in the Bovespa (the Brazilian equity index) over the same period was 22.25%.[96] Using these values, the estimate of a total risk premium for Brazil would be as follows.

$$\text{Equity Risk Premium}_{Brazil} = 5.75\% * \frac{22.25\%}{10.85\%} = 11.77\%$$

The country risk premium for Brazil can be isolated as follows:

$$\text{Country Risk Premium}_{Brazil} = 11.77\% - 5.75\% = 6.02\%$$

---

of equity. See Damodaran, A., 2007, Measuring Company Risk Exposure to Country Risk, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=889388.

95 Abuaf, N., 2011, *Valuing Emerging Market Equities – The Empirical Evidence*, Journal of Applied Finance, v21, 123-138.

96 If the dependence on historical volatility is troubling, the options market can be used to get implied volatilities for both the US market (14.16%) and for the Bovespa (24.03%).

Table 15 lists country volatility numbers for some of the Latin American markets and the resulting total and country risk premiums for these markets, based on the assumption that the equity risk premium for the United States is 5.75%. Appendix 4 contains a more complete list of emerging markets, with equity risk premiums and country risk premiums estimated for each.

Table 15: Equity Market Volatilities and Risk Premiums (Weekly returns: Feb 13-Feb 15): Latin American Countries

Country	*Standard deviation in Equities (weekly)*	*Relative Volatility (to US)*	*Total Equity Risk Premium*	*Country risk premium*
Argentina	35.50%	3.27	18.78%	13.03%
Brazil	22.25%	2.05	11.77%	6.02%
Chile	13.91%	1.28	7.36%	1.61%
Colombia	16.00%	1.47	8.46%	2.71%
Costa Rica	8.78%	0.81	4.64%	-1.11%
Mexico	14.81%	1.36	7.83%	2.08%
Panama	6.18%	0.57	3.27%	-2.48%
Peru	16.15%	1.49	8.54%	2.79%
US	**10.87%**	**1.00**	5.75%	0.00%
Venezuela	40.03%	3.68	21.18%	15.43%

While this approach has intuitive appeal, there are problems with using standard deviations computed in markets with widely different market structures and liquidity. Since equity market volatility is affected by liquidity, with more liquid markets often showing higher volatility, this approach will understate premiums for illiquid markets and overstate the premiums for liquid markets. For instance, the standard deviations for Panama and Costa Rica are lower than the standard deviation in the S&P 500, leading to equity risk premiums for those countries that are lower than the US. The second problem is related to currencies since the standard deviations are usually measured in local currency terms; the standard deviation in the U.S. market is a dollar standard deviation, whereas the standard deviation in the Brazilian market is based on nominal Brazilian Real returns. This is a relatively simple problem to fix, though, since the standard deviations can be measured in the same currency – you could estimate the standard deviation in dollar returns for the Brazilian market.

3. Default Spreads + Relative Standard Deviations

In the first approach to computing equity risk premiums, we assumed that the default spreads (actual or implied) for the country were good measures of the additional risk we face when investing in equity in that country. In the second approach, we argued that the information in equity market volatility can be used to compute the country risk premium. In the third approach, we will meld the first two, and try to use the information in both the country default spread and the equity market volatility.

The country default spreads provide an important first step in measuring country equity risk, but still only measure the premium for default risk. Intuitively, we would expect the country equity risk premium to be larger than the country default risk spread. To address the issue of how much higher, we look at the volatility of the equity market in a country relative to the volatility of the bond market used to estimate the spread. This yields the following estimate for the country equity risk premium.

$$\text{Country Risk Premium=Country Default Spread*}\left(\frac{\sigma_{\text{Equity}}}{\sigma_{\text{Country Bond}}}\right)$$

To illustrate, consider again the case of Brazil. As noted earlier, the default spread for Brazil in January 2015, based upon its sovereign rating, was 1.90%. We computed annualized standard deviations, using two years of weekly returns, in both the equity market and the government bond, in early March 2015. The annualized standard deviation in the Brazilian dollar denominated ten-year bond was 11.97%, well below the standard deviation in the Brazilian equity index of 22.25%. The resulting country equity risk premium for Brazil is as follows:

$$\text{Brazil Country Risk Premium} = 1.90\% * \frac{22.25\%}{11.97\%} = 3.53\%$$

Unlike the equity standard deviation approach, this premium is in addition to a mature market equity risk premium. Thus, assuming a 5.75% mature market premium, we would compute a total equity risk premium for Brazil of 8.22%:

Brazil's Total Equity Risk Premium = 5.75% + 3.53% = 9.28%

Note that this country risk premium will increase if the country rating drops or if the relative volatility of the equity market increases.

Why should equity risk premiums have any relationship to country bond spreads? A simple explanation is that an investor who can make 1.90% risk premium on a dollar-denominated Brazilian government bond would not settle for a risk premium of 1.90% (in dollar terms) on Brazilian equity. Playing devil's advocate, however, a critic could argue that the interest rate on a country bond, from which default spreads are extracted, is not really an expected return since it is based upon the promised cash flows (coupon and principal) on the bond rather than the expected cash flows. In fact, if we wanted to estimate a risk premium for bonds, we would need to estimate the expected return based upon expected cash flows, allowing for the default risk. This would result in a lower default spread and equity risk premium. Both this approach and the last one use the standard deviation in equity of a market to make a judgment about country risk premium, but they measure it relative to different bases. This approach uses the country bond as a base, whereas the previous one uses the standard deviation in the U.S. market. This approach assumes that investors are more likely to choose between Brazilian bonds and Brazilian equity, whereas the previous approach assumes that the choice is across equity markets.

There are two potential measurement problems with using this approach. The first is that the relative standard deviation of equity is a volatile number, both across countries (ranging from 4.04 for India to 0.48 for the Phillipines) and across time (Brazil's relative volatility numbers have ranged from close to one to well above 2). The second is that computing the relative volatility requires us to estimate volatility in the government bond, which, in turn, presupposes that long-term government bonds not only exist but are also traded.[97] In countries where this data item is not available, we have three choices. One is to fall back on one of the other two approaches. The second is to use a different market measure of default risk, say the CDS spread, and compute the standard deviation in the spread; this number can be standardized by dividing the level of the spread. The third is to compute a cross sectional average of the ratio of stock market to bond market volatility across countries, where both items are available, and use that average. In 2015, for instance, there were 26 emerging markets, where both the equity market volatility and the government bond volatility numbers were available, at least for 100 trading weeks; the numbers are summarized in Appendix 5. The median ratio, across these markets, of equity market volatility to bond price volatility was approximately 1.88.[98] We also computed a second measure of relative volatility: equity volatility divided by the coefficient of variation in the CDS spread.

	$\sigma_{Equity} / \sigma_{Bond}$	$\sigma_{Equity} / \sigma_{CDS}$
Number of countries with data	26	46

[97] One indication that the government bond is not heavily traded is an abnormally low standard deviation on the bond yield.

[98] The ratio seems to be lowest in the markets with the highest default spreads and higher in markets with lower default spreads. The median ratio this year is higher than it has been historically. On my website, I continue to use a multiple of 1.50, reflecting the historical value for this ratio.

Average	1.86	2.11
Median	1.88	0.97
Maximum	4.04	23.49
Minimum	0.48	0.51

Looking at the descriptive statistics, the need to adjust default spreads seems to be smaller, at least in the cross section, if you use the CDS spread as your measure of the default spread for a country; the median ratio is close to one.

Choosing between the approaches

The three approaches to estimating country risk premiums will usually give you different estimates, with the bond default spread and relative equity standard deviation approaches generally yielding lower country risk premiums than the melded approach that uses both the country bond default spread and the equity and bond standard deviations. Table 16 summarizes the estimates of country equity and total risk premium using the three approaches for Brazil in March 2014:

Table 16: Country and Total Equity Risk Premium: Brazil in January 2013

Approach	*Mature Market Equity Premium*	*Brazil Country Risk Premium*	*Total Equity Risk Premium*
Country Bond Default Spread	5.75%	1.90%	7.65%
Relative Equity Market Standard Deviations	5.75%	6.02%	11.77%
Melded Approach (Bond default spread X Relative Standard $\text{Deviation}_{\text{Bond}}$)	5.75%	1.90%*1.86 = 3.53%	9.28%
Melded Approach (CDS X Relative Standard $\text{Deviation}_{\text{CDS}}$)	5.75%	3.37% *1.87= 6.30%	12.05%

The CDS and relative equity market approaches yield similar equity risk premiums, but that is more the exception than the rule. In particular, the melded CDS approach offers more promise going forward, as more countries have CDS traded on them. With all three approaches, just as companies mature and become less risky over time, countries can mature and become less risky as well.

One way to adjust country risk premiums over time is to begin with the premium that emerges from the melded approach and to adjust this premium down towards either the country bond default spread or the country premium estimated from equity standard deviations. Thus, the equity risk premium will converge to the country bond default spread as we look at longer term expected returns. As an illustration, the country risk premium for Brazil would be 3.53% for the next year but decline over time to 1.90% (country default spread) or perhaps even lower, depending upon your assessment of how Brazil's economy will evolve over time.

Implied Equity Premiums

The problem with any historical premium approach, even with substantial modifications, is that it is backward looking. Given that our objective is to estimate an updated, forward-looking premium, it seems foolhardy to put your faith in mean reversion and past data. In this section, we will consider three approaches for estimating equity risk premiums that are more forward looking.

1. DCF Model Based Premiums

When investors price assets, they are implicitly telling you what they require as an expected return on that asset. Thus, if an asset has expected cash flows of \$15 a year in perpetuity, and an investor pays \$75 for that asset, he is announcing to the world that his required rate of return on that asset is 20% (15/75). In this section, we expand on this intuition and argue that the current market prices for equity, in conjunction with expected cash flows, should yield an estimate on the equity risk premium.

A Stable Growth DDM Premium

It is easiest to illustrated implied equity premiums with a dividend discount model (DDM). In the DDM, the value of equity is the present value of expected dividends from the investment. In the special case where dividends are assumed to grow at a constant rate forever, we get the classic stable growth (Gordon) model:

$$\text{Value of equity} = \frac{\text{Expected Dividends Next Period}}{\text{(Required Return on Equity - Expected Growth Rate)}}$$

This is essentially the present value of dividends growing at a constant rate. Three of the four inputs in this model can be obtained or estimated - the current level of the market (value), the expected dividends next period and the expected growth rate in earnings and dividends in the long term. The only "unknown" is then the required return on equity; when we solve for it, we get an implied expected return on stocks. Subtracting out the riskfree rate will yield an implied equity risk premium.

To illustrate, assume that the current level of the S&P 500 Index is 900, the expected dividend yield on the index is 2% and the expected growth rate in earnings and dividends in the long term is 7%. Solving for the required return on equity yields the following:

$$900 = (.02*900) /(r - .07)$$

Solving for r,

$$r = (18+63)/900 = 9\%$$

If the current riskfree rate is 6%, this will yield a premium of 3%.

In fact, if we accept the stable growth dividend discount model as the base model for valuing equities and assume that the expected growth rate in dividends should equate to the riskfree rate in the long term, the dividend yield on equities becomes a measure of the equity risk premium:

$$\text{Value of equity} = \frac{\text{Expected Dividends Next Period}}{\text{(Required Return on Equity - Expected Growth Rate)}}$$

$$\begin{aligned}\text{Dividends/ Value of Equity} &= \text{Required Return on Equity – Expected Growth rate}\\ \text{Dividend Yield} &= \text{Required Return on Equity – Riskfree rate}\\ &= \text{Equity Risk Premium}\end{aligned}$$

Rozeff (1984) made this argument[99] and empirical support has been claimed for dividend yields as predictors of future returns in many studies since.[100] Note that this simple equation will break down if (a) companies do not pay out what they can afford to in dividends, i.e., they hold back cash or (b) if earnings are expected to grow at extraordinary rates for the short term.

There is another variant of this model that can be used, where we focus on earnings instead of dividends. To make this transition, though, we have to state the expected growth rate as a function of the payout ratio and return on equity (ROE) :[101]

$$\begin{aligned}\text{Growth rate} &= \text{(1 – Dividends/ Earnings) (Return on equity)}\\ &= \text{(1 – Payout ratio) (ROE)}\end{aligned}$$

Substituting back into the stable growth model,

---

[99] Rozeff, M. S. 1984. *Dividend yields are equity risk premiums*, Journal of Portfolio Management, v11, 68-75.

[100] Fama, E. F., and K. R. French. 1988. *Dividend yields and expected stock returns*. Journal of Financial Economics, v22, 3-25.

[101] This equation for sustainable growth is discussed more fully in Damodaran, A., 2002, Investment Valuation, John Wiley and Sons.

$$\text{Value of equity} = \frac{\text{Expected Earnings Next Period (Payout ratio)}}{\text{(Required Return on Equity - (1-Payout ratio) (ROE))}}$$

If we assume that the return on equity (ROE) is equal to the required return on equity (cost of equity), i.e., that the firm does not earn excess returns, this equation simplifies as follows:

$$\text{Value of equity} = \frac{\text{Expected Earnings Next Period}}{\text{Required Return on Equity}}$$

In this case, the required return on equity can be written as:

$$\text{Required return on equity} = \frac{\text{Expected Earnings Next Period}}{\text{Value of Equity}}$$

In effect, the inverse of the PE ratio (also referenced as the earnings yield) becomes the required return on equity, <u>if firms are in stable growth and earning no excess returns</u>. Subtracting out the riskfree rate should yield an implied premium:

Implied premium (EP approach) = Earnings Yield on index – Riskfree rate

In January 2015, the first of these approaches would have delivered a very low equity risk premium for the US market.

Dividend Yield = 1.87%

The second approach of netting the earnings yield against the risk free rate would have generated a more plausible number[102]:

Earnings Yield = 5.57%:
Implied premium = Earnings yield – 10-year US Treasury Bond rate
= 5.57% - 2.17% = 3.40%

Both approaches, though, draw on the dividend discount model and make strong assumptions about firms being in stable growth and/or long-term excess returns.

A Generalized Model: Implied Equity Risk Premium

To expand the model to fit more general specifications, we would make the following changes: Instead of looking at the actual dividends paid as the only cash flow to equity, we would consider <u>potential dividends instead of actual dividends</u>. In my earlier work (2002, 2006), the free cash flow to equity (FCFE), i.e, the cash flow left over after taxes, reinvestment needs and debt repayments, was offered as a measure of potential dividends.[103] Over the last decade, for instance, firms have paid out only about half their FCFE as dividends. If this poses too much of an estimation challenge, there is a

---

[102] The earnings yield in January 2015 is estimated by dividing the aggregated earnings for the index by the index level.

[103] Damodaran, A., 2002, *Investment Valuation*, John Wiley and Sons; Damodaran, A., 2006, *Damodaran on Valuation*, John Wiley and Sons.

simpler alternative. Firms that hold back cash build up large cash balances that they use over time to fund stock buybacks. Adding stock buybacks to aggregate dividends paid should give us a better measure of total cash flows to equity. The model can also be expanded to allow for a high growth phase, where earnings and dividends can grow at rates that are very different (usually higher, but not always) than stable growth values. With these changes, the value of equity can be written as follows:

$$\text{Value of Equity} = \sum_{t=1}^{t=N} \frac{E(FCFE_t)}{(1+k_e)^t} + \frac{E(FCFE_{N+1})}{(k_e\text{-}g_N)\,(1+k_e)^N}$$

In this equation, there are N years of high growth, $E(FCFE_t)$ is the expected free cash flow to equity (potential dividend) in year t, k_e is the rate of return expected by equity investors and g_N is the stable growth rate (after year N). We can solve for the rate of return equity investors need, given the expected potential dividends and prices today. Subtracting out the riskfree rate should generate a more realistic equity risk premium.

In a variant of this approach, the implied equity risk premium can be computed from excess return or residual earnings models. In these models, the value of equity today can be written as the sum of capital invested in assets in place and the present value of future excess returns:[104]

$$\text{Value of Equity} = \text{Book Equity today} + \sum_{t=1}^{t=\infty} \frac{\text{Net Income}_t - k_e(\text{Book Equity}_{t-1})}{(1+k_e)^t}$$

If we can make estimates of the book equity and net income in future periods, we can then solve for the cost of equity and use that number to back into an implied equity risk premium. Claus and Thomas (2001) use this approach, in conjunction with analyst forecasts of earnings growth, to estimate implied equity risk premiums of about 3% for the market in 2000.[105] Easton (2007) provides a summary of possible limitations of models that attempt to extract costs of equity from accounting data including the unreliability of book value numbers and the use of optimistic estimates of growth from analysts.[106]

Implied Equity Risk Premium: S&P 500

Given its long history and wide following, the S&P 500 is a logical index to use to try out the implied equity risk premium measure. In this section, we will begin by

---

[104] For more on excess return models, see Damodaran, A, 2006, *Valuation Approaches and Metrics: A Survey of the Theory and Evidence,* Working Paper, www.damodaran.com.

[105] Claus, J. and J. Thomas, 2001,'*Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets*, Journal of Finance 56(5), 1629–1666.

[106] Easton, P., 2007, *Estimating the cost of equity using market prices and accounting data*, Foundations and Trends in Accounting, v2, 241-364.

estimating implied equity risk premiums at the start of the years 2008-2015, and follow up by looking at the volatility in that estimate over time.

Implied Equity Risk Premiums: Annual Estimates from 2008 to 2015

On December 31, 2007, the S&P 500 Index closed at 1468.36, and the dividend yield on the index was roughly 1.89%. In addition, the consensus estimate of growth in earnings for companies in the index was approximately 5% for the next 5 years.[107] Since this is not a growth rate that can be sustained forever, we employ a two-stage valuation model, where we allow growth to continue at 5% for 5 years, and then lower the growth rate to 4.02% (the riskfree rate) after that.[108] Table 17 summarizes the expected dividends for the next 5 years of high growth, and for the first year of stable growth thereafter:

Table 17: Estimated Dividends on the S&P 500 Index – January 1, 2008

Year	*Dividends on Index*
1	29.12
2	30.57
3	32.10
4	33.71
5	35.39
6	36.81

[a]Dividends in the first year = 1.89% of 1468.36 (1.05)

If we assume that these are reasonable estimates of the expected dividends and that the index is correctly priced, the value can be written as follows:

$$1468.36 = \frac{29.12}{(1+r)} + \frac{30.57}{(1+r)^2} + \frac{32.10}{(1+r)^3} + \frac{33.71}{(1+r)^4} + \frac{35.39}{(1+r)^5} + \frac{36.81}{(r-.0402)(1+r)^5}$$

Note that the last term in the equation is the terminal value of the index, based upon the stable growth rate of 4.02%, discounted back to the present. Solving for required return in this equation yields us a value of 6.04%. Subtracting out the ten-year treasury bond rate (the riskfree rate) yields an implied equity premium of 2.02%.

The focus on dividends may be understating the premium, since the companies in the index have bought back substantial amounts of their own stock over the last few years. Table 18 summarizes dividends and stock buybacks on the index, going back to 2001.

Table 18: Dividends and Stock Buybacks: 2001- 2007

Year	*Dividend*	*Stock Buyback*	*Total Yield*

[107] We used the average of the analyst estimates for individual firms (bottom-up). Alternatively, we could have used the top-down estimate for the S&P 500 earnings.

[108] The treasury bond rate is the sum of expected inflation and the expected real rate. If we assume that real growth is equal to the real interest rate, the long term stable growth rate should be equal to the treasury bond rate.

	Yield	*Yield*	
2001	1.37%	1.25%	2.62%
2002	1.81%	1.58%	3.39%
2003	1.61%	1.23%	2.84%
2004	1.57%	1.78%	3.35%
2005	1.79%	3.11%	4.90%
2006	1.77%	3.39%	5.16%
2007[a]	1.89%	4.00%	5.89%
Average total yield between 2001-2007 =			4.02%

[a]Trailing 12-month data, from September 2006 through September 2007. In January 2008, this was the information that would have been available. The actual cash yield for all of 2007 was 6.49%.

In 2007, for instance, firms collectively returned more than twice as much in the form of buybacks than they paid out in dividends. Since buybacks are volatile over time, and 2007 may represent a high-water mark for the phenomenon, we recomputed the expected cash flows, in table 19, for the next 6 years using the average total yield (dividends + buybacks) of 4.11%, instead of the actual dividends, and the growth rates estimated earlier (5% for the next 5 years, 4.02% thereafter):

Table 19: Cashflows on S&P 500 Index

Year	*Dividends+ Buybacks on Index*
1	63.37
2	66.54
3	69.86
4	73.36
5	77.02

Using these cash flows to compute the expected return on stocks, we derive the following:

$$1468.36 = \frac{63.37}{(1+r)} + \frac{66.54}{(1+r)^2} + \frac{69.86}{(1+r)^3} + \frac{73.36}{(1+r)^4} + \frac{77.02}{(1+r)^5} + \frac{77.02(1.0402)}{(r-.0402)(1+r)^5}$$

Solving for the required return and the implied premium with the higher cash flows:

Required Return on Equity = 8.39%

Implied Equity Risk Premium = Required Return on Equity - Riskfree Rate

= 8.48% - 4.02% = 4.46%

This value (4.46%) would have been our estimate of the equity risk premium on January 1, 2008.

During 2008, the S&P 500 lost just over a third of its value and ended the year at 903.25 and the treasury bond rate plummeted to close at 2.21% on December 31, 2008. Firms also pulled back on stock buybacks and financial service firms in particular cut dividends during the year. The inputs to the equity risk premium computation reflect these changes:

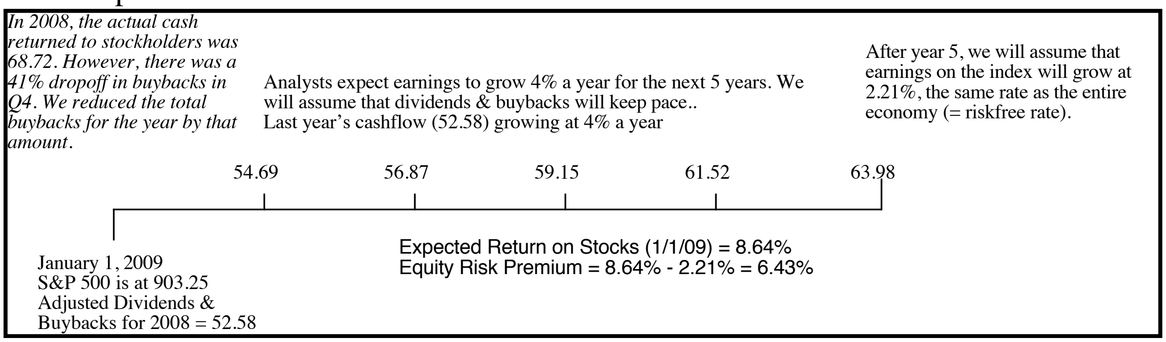
Level of the index = 903.25 (Down from 1468.36)

Treasury bond rate = 2.21% (Down from 4.02%)

Updated dividends and buybacks on the index = 52.58 (Down about 15%)

Expected growth rate = 4% for next 5 years (analyst estimates) and 2.21% thereafter (set equal to riskfree rate).

The computation is summarized below:



The resulting equation is below:

$$903.25 = \frac{54.69}{(1+r)} + \frac{56.87}{(1+r)^2} + \frac{59.15}{(1+r)^3} + \frac{61.52}{(1+r)^4} + \frac{63.98}{(1+r)^5} + \frac{63.98(1.0221)}{(r-.0221)(1+r)^5}$$

Solving for the required return and the implied premium with the higher cash flows:

Required Return on Equity = 8.64%

Implied Equity Risk Premium = Required Return on Equity - Riskfree Rate

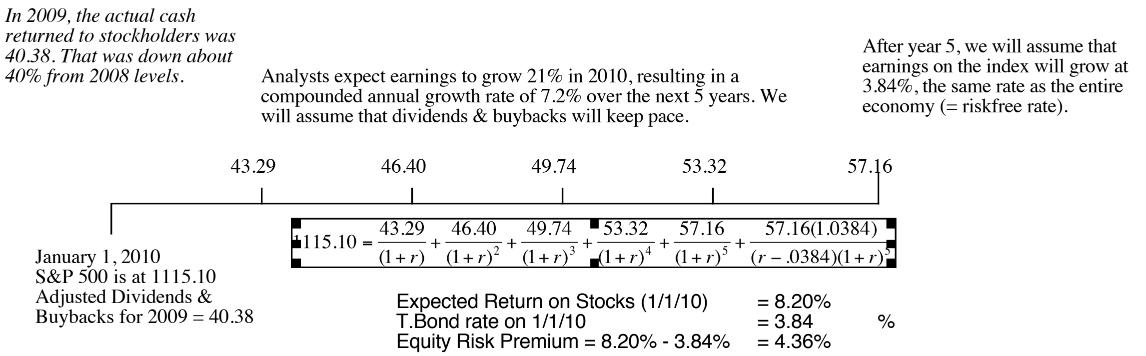
= 8.64% - 2.21% = 6.43%

The implied premium rose more than 2%, from 4.37% to 6.43%, over the course of the year, indicating that investors perceived more risk in equities at the end of the year, than they did at the start and were demanding a higher premium to compensate.

By January 2010, the fears of a banking crisis had subsided and the S&P 500 had recovered to 1115.10. However, a combination of dividend cuts and a decline in stock buybacks had combined to put the cash flows on the index down to 40.38 in 2009. That was partially offset by increasing optimism about an economic recovery and expected earnings growth for the next 5 years had bounced back to 7.2%.[109] The resulting equity risk premium is 4.36%:

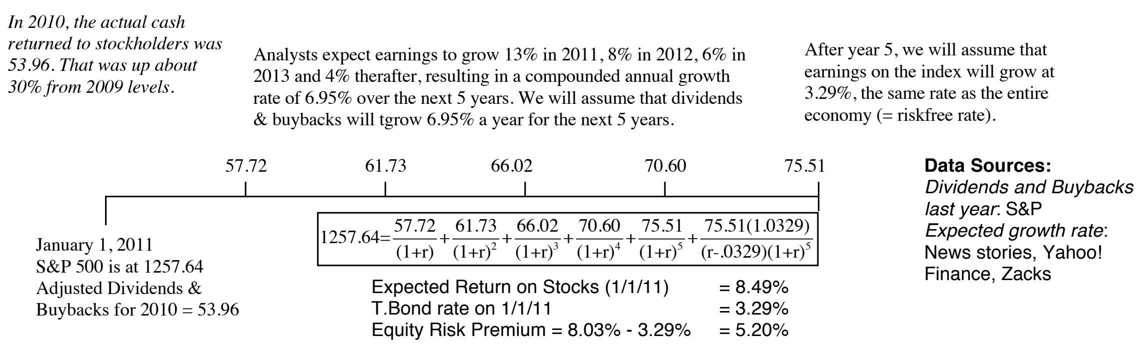
---

[109] The expected earnings growth for just 2010 was 21%, primarily driven by earnings bouncing back to pre-crisis levels, followed by a more normal 4% earnings growth in the following years. The compounded average growth rate is $((1.21)\ (1.04)^4)^{1/5}$-1= .072 or 7.2%.



In effect, equity risk premiums have reverted back to what they were before the 2008 crisis.

Updating the numbers to January 2011, the S&P 500 had climbed to 1257.64, but cash flows on the index, in the form of dividends and buybacks, made an even more impressive comeback, increasing to 53.96 from the depressed 2009 levels. The implied equity risk premium computation is summarized below:



The implied equity risk premium climbed to 5.20%, with the higher cash flows more than offsetting the rise in equity prices.

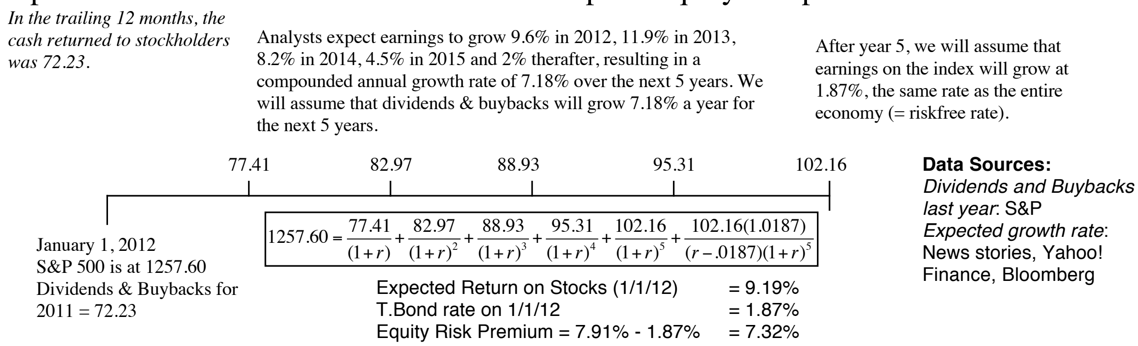
The S&P 500 ended 2011 at 1257.60, almost unchanged from the level at the start of the year. The other inputs into the implied equity risk premium equation changed significantly over the year:

a. The ten-year treasury bond rate dropped during the course of the year from 3.29% to 1.87%, as the European debt crisis caused a "flight to safety". The US did lose its AAA rating with Standard and Poor's during the course of the year, but we will continue to assume that the T.Bond rate is risk free.
b. Companies that had cut back dividends and scaled back stock buybacks in 2009, after the crisis, and only tentatively returned to the fray in 2010, returned to buying back stocks at almost pre-crisis levels. The total dividends and buybacks

for the trailing 12 months leading into January 2012 climbed to 72.23, a significant increase over the previous year.[110]

c. Analysts continued to be optimistic about earnings growth, in the face of signs of a pickup in the US economy, forecasting growth rate of 9.6% for 2012 (year 1), 11.9% in 2013, 8.2% in 2014, 4% in 2015 and 2.5% in 2016, leading to a compounded annual growth rate of 7.18% a year:.

Incorporating these inputs into the implied equity risk premium computation, we get an expected return on stocks of 9.29% and an implied equity risk premium of 7.32%:



Since the index level did not change over the course of the year, the jump in the equity risk premium from 5.20% on January 1, 2011 to 7.32% on January 1, 2012, was precipitated by two factors. The first was the drop in the ten-year treasury bond rate to a historic low of 1.87% and the second was the surge in the cash returned to stockholders, primarily in buybacks. With the experiences of the last decade fresh in our minds, we considered the possibility that the cash returned during the trailing 12 months may reflect cash that had built up during the prior two years, when firms were in their defensive posture. If that were the case, it is likely that buybacks will decline to a more normalized value in future years. To estimate this value, we looked at the total cash yield on the S&P 500 from 2002 to 2011 and computed an average value of 4.69% over the decade in table 20.

Table 20: Dividends and Buybacks on S&P 500 Index: 2002-2011

Year	Dividend Yield	Buybacks/Index	Yield
2002	1.81%	1.58%	3.39%
2003	1.61%	1.23%	2.84%
2004	1.57%	1.78%	3.35%
2005	1.79%	3.11%	4.90%
2006	1.77%	3.39%	5.16%

[110] These represented dividends and stock buybacks from October 1, 2010 to September 30, 2011, based upon the update from S&P on December 22, 2011. The data for the last quarter is not made available until late March of the following year.

2007	1.92%	4.58%	6.49%
2008	3.15%	4.33%	7.47%
2009	1.97%	1.39%	3.36%
2010	1.80%	2.61%	4.42%
2011	2.00%	3.53%	5.54%
Average: Last 10 years =			4.69%

Assuming that the cash returned would revert to this yield provides us with a lower estimate of the cash flow (4.69% of 1257.60= 59.01) and an equity risk premium of 6.01%:

In the trailing 12 months, the cash returned to stockholders was 72.23. Using the average cash yield of 4.69% for 2002-2011 the cash returned would have been 59.01.

Analysts expect earnings to grow 9.6% in 2012, 11.9% in 2013, 8.2% in 2014, 4.5% in 2015 and 2.5% therafter, resulting in a compounded annual growth rate of 7.18% over the next 5 years. We will assume that dividends & buybacks will grow 7.18% a year for the next 5 years.

After year 5, we will assume that earnings on the index will grow at 1.87%, the same rate as the entire economy (= riskfree rate).

63.24 67.78 72.65 77.87 83.46

January 1, 2012
S&P 500 is at 1257.60
Normalized Dividends & Buybacks for 2011 = 59.01

$$1257.60 = \frac{63.24}{(1+r)} + \frac{67.78}{(1+r)^2} + \frac{72.65}{(1+r)^3} + \frac{77.87}{(1+r)^4} + \frac{83.46}{(1+r)^5} + \frac{83.46(1.0287)}{(r-.0187)(1+r)^5}$$

Expected Return on Stocks (1/1/12) = 7.88%
T.Bond rate on 1/1/12 = 1.87%
Equity Risk Premium = 7.91% - 1.87% = 7.32%

Data Sources:
Dividends and Buybacks last year: S&P
Expected growth rate: News stories, Yahoo! Finance, Bloomberg

So, did the equity risk premium for the S&P 500 jump from 5.20% to 7.32%, as suggested by the raw cash yield, or from 5.20% to 6.01%, based upon the normalized yield? We would be more inclined to go with the latter, especially since the index remained unchanged over the year. Note, though, that if the cash returned by firms does not drop back in the next few quarters, we will revisit the assumption of normalization and the resulting lower equity risk premium.

By January 1, 2013, the S&P 500 climbed to 1426.19 and the treasury bond rate had dropped to 1.76%. The dividends and buybacks were almost identical to the prior year and the smoothed out cash returned (using the average yield over the prior 10 years) climbed to 69.46. Incorporating the lower growth expectations leading into 2013, the implied equity risk premium dropped to 5.78% on January 1, 2013:

In 2012, the actual cash returned to stockholders was 72.25. Using the average total yield for the last decade yields 69.46

Analysts expect earnings to grow 7.67% in 2013, 7.28% in 2014, scaling down to 1.76% in 2017, resulting in a compounded annual growth rate of 5.27% over the next 5 years. We will assume that dividends & buybacks will tgrow 5.27% a year for the next 5 years.

After year 5, we will assume that earnings on the index will grow at 1.76%, the same rate as the entire economy (= riskfree rate).

73.12 76.97 81.03 85.30 89.80

January 1, 2013
S&P 500 is at 1426.19
Adjusted Dividends & Buybacks for base year = 69.46

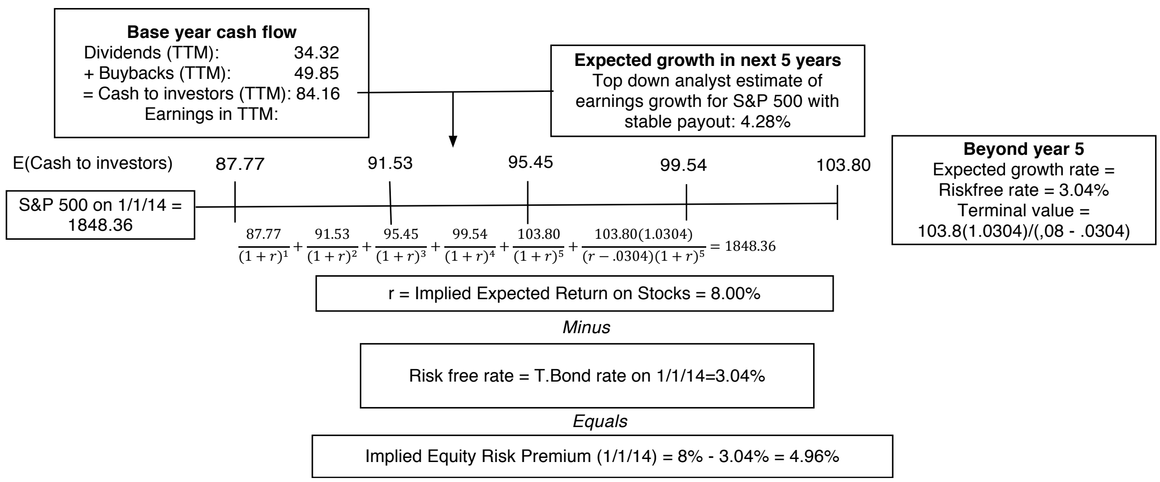
$$1426.19 = \frac{73.12}{(1+r)} + \frac{76.97}{(1+r)^2} + \frac{81.03}{(1+r)^3} + \frac{85.30}{(1+r)^4} + \frac{89.80}{(1+r)^5} + \frac{89.80(1.0176)}{(r-.0176)(1+r)^5}$$

Expected Return on Stocks (1/1/13) = 7.54%
T.Bond rate on 1/1/13 = 1.76%
Equity Risk Premium = 7.54% - 1.76% = 5.78%

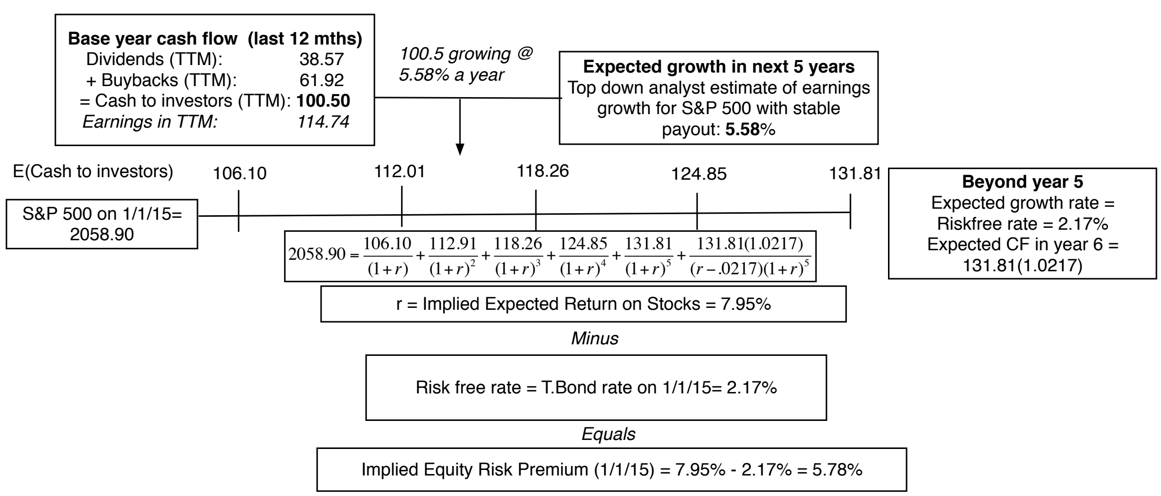
Data Sources:
Dividends and Buybacks last year: S&P
Expected growth rate: S&P, Media reports, Factset, Thomson-Reuters

Note that the chasm between the trailing 12-month cash flow premium and the smoother cash yield premium that had opened up at the start of 2012 had narrowed. The trailing 12-month cash flow premium was 6%, just 0.22% higher than the 5.78% premium obtained with the smoothed out cash flow.

After a good year for stocks, the S&P 500 was at 1848.36 on January 1, 2014, up 29.6% over the prior year, and cash flows also jumped to 84.16 over the trailing 12 months (ending September 30, 2013), up 16.48% over the prior year. Incorporating an increase in the US ten-year treasury bond rate to 3.04%, the implied equity risk premium at the start of 2014 was 4.96%.



During 2014, stocks continued to rise, albeit at a less frenetic pace, and the US ten-year treasury bond rate dropped back again to 2.17%. Since buybacks and dividends grew at higher rate than prices, the net effect was an increase in the implied equity risk premium to 5.78% at the start of 2015:



A Term Structure for Equity Risk Premiums

When we estimate an implied equity risk premium, from the current level of the index and expected future cash flows, we are estimating a compounded average equity risk premium over the long term. Thus, the 5.78% estimate of the equity risk premium at the start of 2015 is the geometric average of the annualized equity risk premiums in future years and is analogous to the yield to maturity on a long term bond.

But is it possible that equity risk premiums have a term structure, just as interest rates do? Absolutely. In a creative attempt to measure the slope of the term structure of equity risk premiums, Binsberger, Brandt and Koijen (2012) use dividend strips, i.e., short term assets that pay dividends for finite time periods (and have no face value), to extract equity risk premiums for the short term as opposed to the long term. Using dividend strips on the S&P 500 to extract expected returns from 1996 to 2009, they find that equity risk premiums are higher for shorter term claims than for longer term claims, by approximately 2.75%.[111] Their findings are contested by Boguth, Carlson, Fisher and Simutin (2011), who note that small market pricing frictions are amplified when valuing synthetic dividend strips and that using more robust return measures results in no significant differences between short term and longer term equity risk premiums.[112]

While this debate will undoubtedly continue, the relevance to valuation and corporate finance practice is questionable. Even if you could compute period-specific equity risk premiums, the effect on value of using these premiums (instead of the compounded average premium) would be small in most valuations. To illustrate, your valuation of an asset, using an equity risk premium of 7% for the first 3 years and 5.5% thereafter[113], at the start of 2015, would be very similar to the value you would have obtained using 5.78% as your equity risk premium for all time periods. The only scenario where using year-specific premiums would make a material difference would be in the valuation of an asset or investment with primarily short-term cash flows, where using a higher short term premium will yield a lower (and perhaps more realistic) value for the asset.

---

111 Binsbergen, J. H. van, Michael W. Brandt, and Ralph S. J. Koijen, 2012, *On the timing and pricing of dividends*, American Economic Review, v102, 1596-1618.

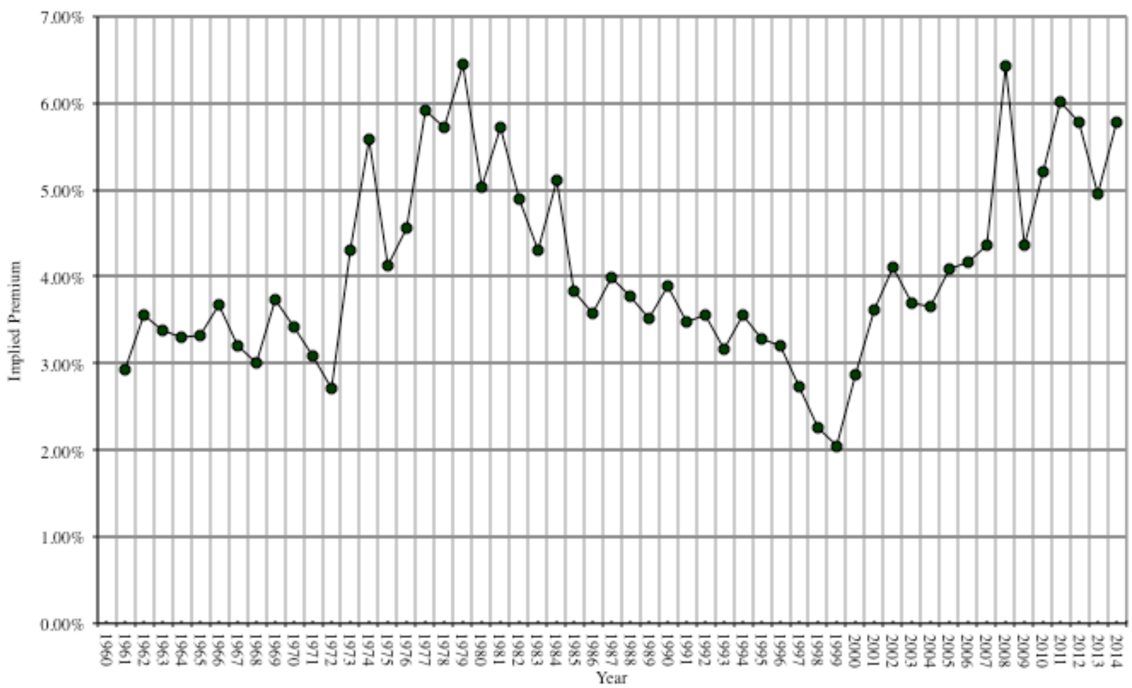
112 Boguth, O., M. Carlson, A. Fisher and M. Simutin, 2011, *Dividend Strips and the Term Structure of Equity Risk Premia: A Case Study of Limits to Arbitrage*, Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1931105. In a response, Binsbergen, Brandt and Koijen argue that their results hold even if traded dividend strips (rather than synthetic strips) are used.

113 The compounded average premium over time, using a 7% equity risk premium for the first 3 years and 5.88% thereafter, is roughly 6.01%.

Time Series Behavior for S&P 500 Implied Premium

As the inputs to the implied equity risk premium, it is quite clear that the value for the premium will change not just from day to day but from one minute to the next. In particular, movements in the index will affect the equity risk premium, with higher (lower) index values, other things remaining equal, translating into lower (higher) implied equity risk premiums. In Figure 9, we chart the implied premiums in the S&P 500 from 1960 to 2014 (year ends):

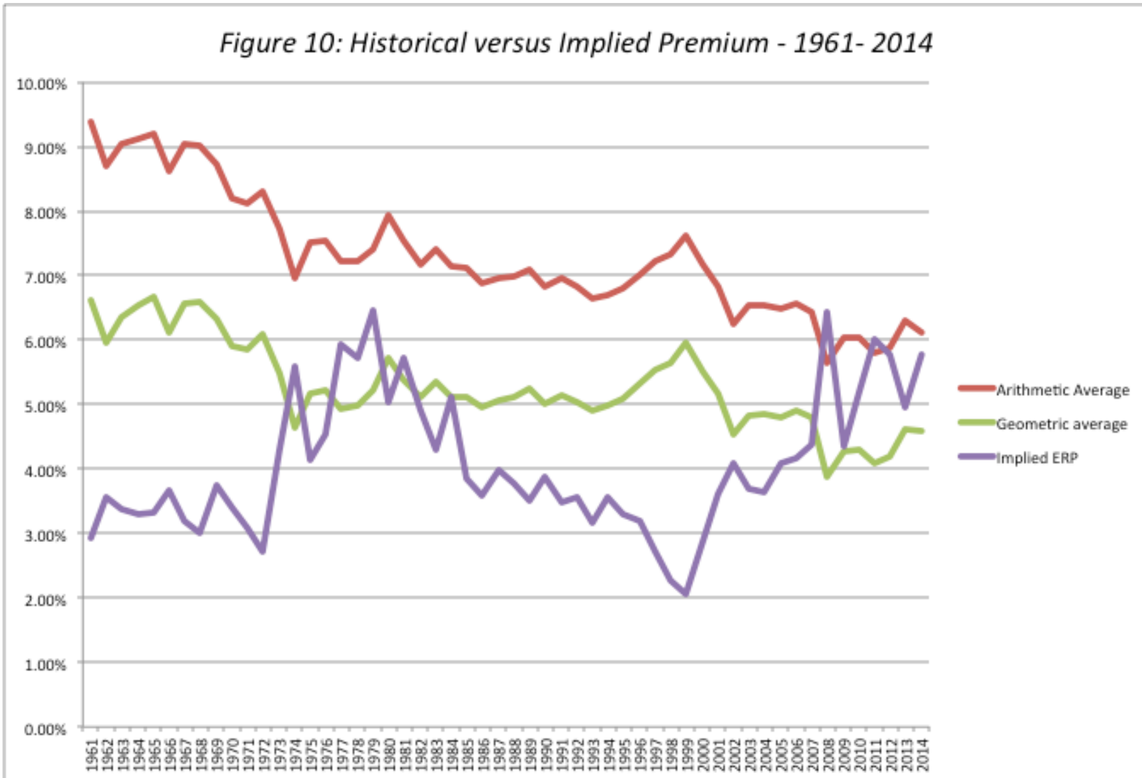
Figure 9: Implied Premium for US Equity Market: 1960-2014



In terms of mechanics, we used potential dividends (including buybacks) as cash flows, and a two-stage discounted cash flow model; the estimates for each year are in appendix 6.[114] Looking at these numbers, we would draw the following conclusions:

- The implied equity premium has generally been lower than the historical risk premium for the US equity market for most of the last few decades. To provide a contrast, we compare the implied equity risk premiums each year to the historical risk premiums for stocks over treasury bonds, using both geometric and arithmetic averages, each year from 1961 to 2014 in figure 10:

[114] We used analyst estimates of growth in earnings for the 5-year growth rate after 1980. Between 1960 and 1980, we used the historical growth rate (from the previous 5 years) as the projected growth, since analyst estimates were difficult to obtain. Prior to the late 1980s, the dividends and potential dividends were very similar, because stock buybacks were uncommon. In the last 20 years, the numbers have diverged.



The arithmetic average premium, which is used by many practitioners, has been significantly higher than the implied premium over almost the entire fifty-year period (with 2009 and 2011 being the only exceptions). The geometric premium does provide a more interesting mix of results, with implied premiums exceeding historical premiums in the mid-1970s and again since 2008.

- The implied equity premium did increase during the seventies, as inflation increased. This does have interesting implications for risk premium estimation. Instead of assuming that the risk premium is a constant, and unaffected by the level of inflation and interest rates, which is what we do with historical risk premiums, would it be more realistic to increase the risk premium if expected inflation and interest rates go up? We will come back and address this question in the next section.
- While historical risk premiums have generally drifted down for the last few decades, there is a strong tendency towards mean reversion in implied equity premiums. Thus, the premium, which peaked at 6.5% in 1978, moved down towards the average in the 1980s. By the same token, the premium of 2% that we observed at the end of the dot-com boom in the 1990s quickly reverted back to the average, during the market correction from 2000-2003.[115] Given this tendency, it is possible that we can end up

[115] Arnott, Robert D., and Ronald Ryan, 2001, *The Death of the Risk Premium: Consequences of the*

with a far better estimate of the implied equity premium by looking at not just the current premium, but also at historical trend lines. We can use the average implied equity premium over a longer period, say ten to fifteen years. Note that we do not need as many years of data to make this estimate as we do with historical premiums, because the standard errors tend to be smaller.

Finally, the crisis of 2008 was unprecedented in terms of its impact on equity risk premiums. Implied equity risk premiums rose more during 2008 than in any one of the prior 50 years, with much of the change happening in a fifteen week time period towards the end of the year. While much of that increase dissipated in 2009, as equity risk premiums returned to pre-crisis levels, equity risk premiums have remained more volatile since 2008. In the next section, we will take a closer look at this time period.

Implied Equity Risk Premiums during a Market Crisis and Beyond

When we use historical risk premiums, we are, in effect, assuming that equity risk premiums do not change much over short periods and revert back over time to historical averages. This assumption was viewed as reasonable for mature equity markets like the United States, but was put under a severe test during the market crisis that unfolded with the fall of Lehman Brothers on September 15, and the subsequent collapse of equity markets, first in the US, and then globally.

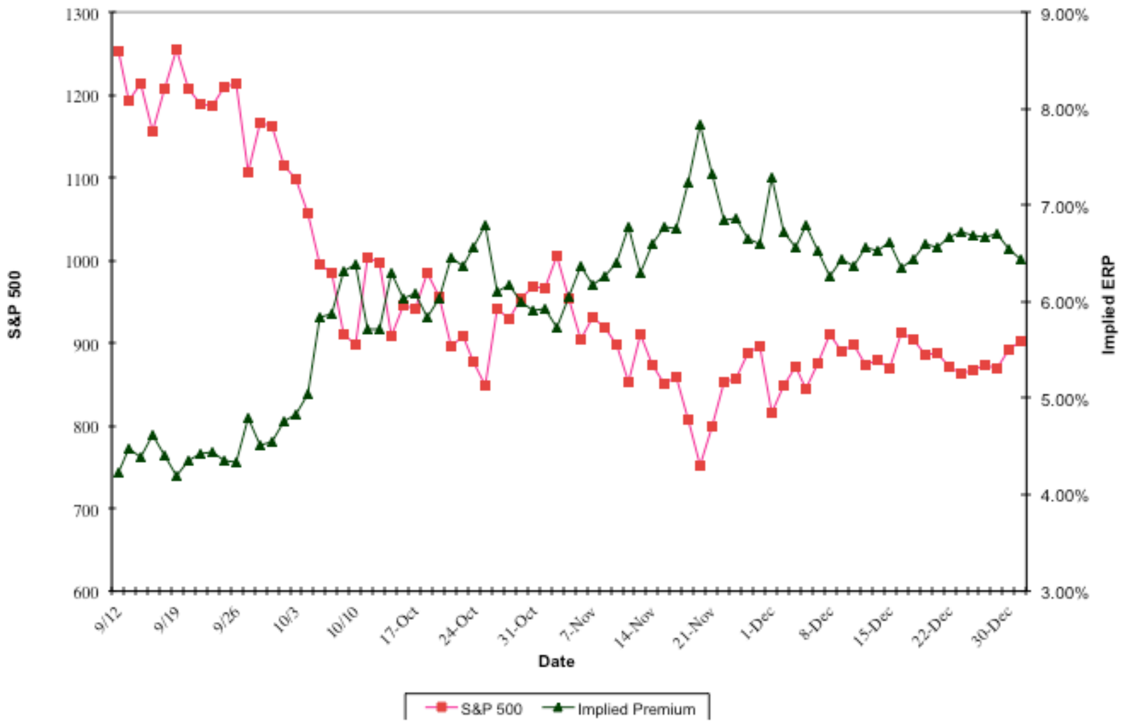
Since implied equity risk premiums reflect the current level of the index, the 75 trading days between September 15, 2008, and December 31, 2008, offer us an unprecedented opportunity to observe how much the price charged for risk can change over short periods. In figure 11, we depict the S&P 500 on one axis and the implied equity risk premium on the other. To estimate the latter, we used the level of the index and the treasury bond rate at the end of each day and used the total dollar dividends and buybacks over the trailing 12 months to compute the cash flows for the most recent year.[116] We also updated the expected growth in earnings for the next 5 years, but that number changed only slowly over the period. For example, the total dollar dividends and buybacks on the index for the trailing 12 months of 52.58 resulted in a dividend yield of 4.20% on September 12 (when the index closed at 1252) but jumped to 4.97% on October 6, when the index closed at 1057.[117]

1990s, Journal of Portfolio Management, v27, 61-74. They make the same point about reduction in implied equity risk premiums that we do. According to their calculations, though, the implied equity risk premium in the late 1990s was negative.

[116] This number, unlike the index and treasury bond rate, is not updated on a daily basis. We did try to modify the number as companies in the index announced dividend suspensions or buyback modifications.

[117] It is possible, and maybe even likely, that the banking crisis and resulting economic slowdown was leading some companies to reassess policies on buybacks. Alcoa, for instance, announced that it was terminating stock buybacks. However, other companies stepped up buybacks in response to lower stock

Figure 11: Implied Equity Risk Premium - 9/12- 12/31/08



In a period of a month, the implied equity risk premium rose from 4.20% on September 12 to 6.39% at the close of trading of October 10 as the S&P moved from 1250 down to 903. Even more disconcertingly, there were wide swings in the equity risk premium within a day; in the last trading hour just on October 10, the implied equity risk premium ranged from a high of 6.6% to a low of 6.1%. Over the rest of the year, the equity risk premium gyrated, hitting a high of 8% in late November, before settling into the year-end level of 6.43%.

The volatility captured in figure 12 was not restricted to just the US equity markets. Global equity markets gyrated with and sometimes more than the US, default spreads widened considerably in corporate bond markets, commercial paper and LIBOR rates soared while the 3-month treasury bill rate dropped close to zero and the implied volatility in option markets rose to levels never seen before. Gold surged but other commodities, such as oil and grains, dropped. Not only did we discover how intertwined equity markets are around the globe but also how markets for all risky assets are tied together. We will explicitly consider these linkages as we go through the rest of the paper.

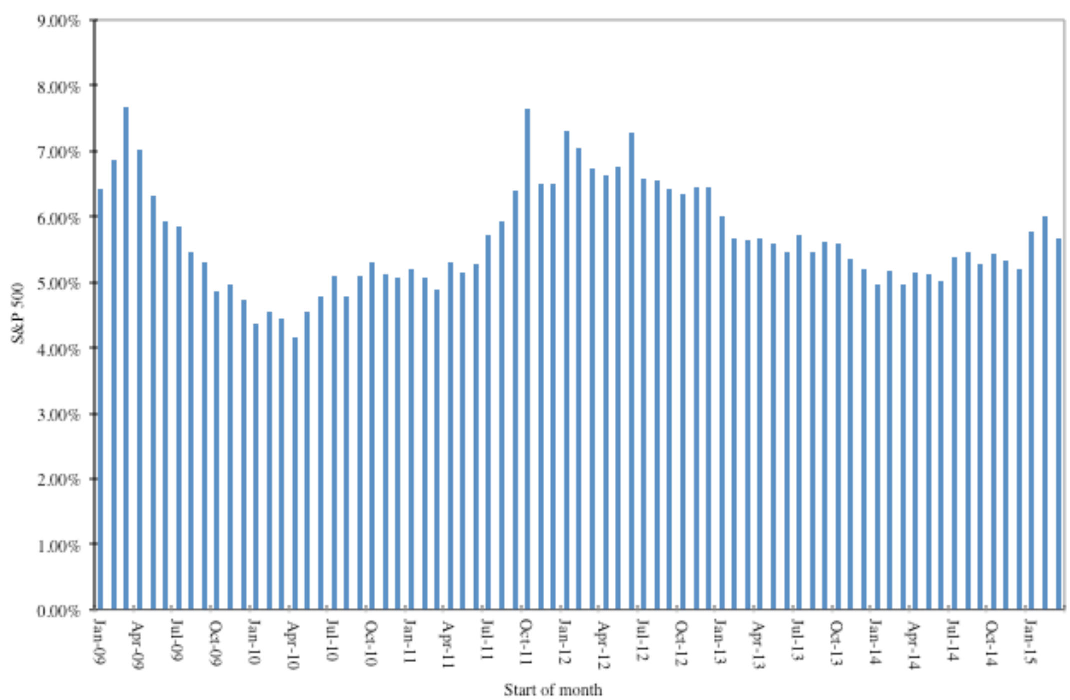
---

prices. If the total cash return was dropping, as the market was, the implied equity risk premiums should be lower than the numbers that we have computed.

There are two ways in which we can view this volatility. One the one side, proponents of using historical averages (either of actual or implied premiums) will use the day-to-day volatility in market risk premiums to argue for the stability of historical averages. They are implicitly assuming that when the crisis passes, markets will return to the status quo. On the other hand, there will be many who point to the unprecedented jump in implied premiums over a few weeks and note the danger of sticking with a "fixed" premium. They will argue that there are sometimes structural shifts in markets, i.e. big events that change market risk premiums for long periods, and that we should be therefore be modifying the risk premiums that we use in valuation as the market changes around us. In January 2009, in the context of equity risk premiums, the first group would have argued we should ignore history (both in terms of historical returns and implied equity risk premiums) and move to equity risk premiums of 6%+ for mature markets (and higher for emerging markets whereas the second would have made a case for sticking with a historical average, which would have been much lower than 6.43%.

The months since the crisis ended in 2008 have seen ups and downs in the implied premium, with clear evidence that the volatility in the equity risk premium has increased over the last few years. In figure 12, we report on the monthly equity risk premiums for the S&P 500 from January 2009 through March 2015:

Figure 12: Implied ERP by month: S&P 500
January 2009- Current



Note that the equity risk premium dropped from its post-crisis highs in 2010 but climbed back in 2011 to 6% or higher, before dropping back to 5% in 2013, before rising again in the last year.

On a personal note, I believe that the very act of valuing companies requires taking a stand on the appropriate equity risk premium to use. For many years prior to September 2008, I used 4% as my mature market equity risk premium when valuing companies, and assumed that mean reversion to this number (the average implied premium over time) would occur quickly and deviations from the number would be small. Though mean reversion is a powerful force, I think that the banking and financial crisis of 2008 has created a new reality, i.e., that equity risk premiums can change quickly and by large amounts even in mature equity markets. Consequently, I have forsaken my practice of staying with a fixed equity risk premium for mature markets, and I now vary it year-to-year, and even on an intra-year basis, if conditions warrant. After the crisis, in the first half of 2009, I used equity risk premiums of 6% for mature markets in my valuations. As risk premiums came down in 2009, I moved back to using a 4.5% equity risk premium for mature markets in 2010. With the increase in implied premiums at the start of 2011, my valuations for the year were based upon an equity risk premium of 5% for mature markets and I increased that number to 6% for 2012. In 2015, I will be using a lower equity risk premium (5.75%), reflecting the implied premium at the start o the year but will remain vigilant by computing the premium on a monthly basis. While some may view this shifting equity risk premium as a sign of weakness, I would frame it differently. When valuing individual companies, I want my valuations to reflect my assessments of the company and not my assessments of the overall equity market. Using equity risk premiums that are very different from the implied premium will introduce a market view into individual company valuations.

Determinants of Implied Premiums

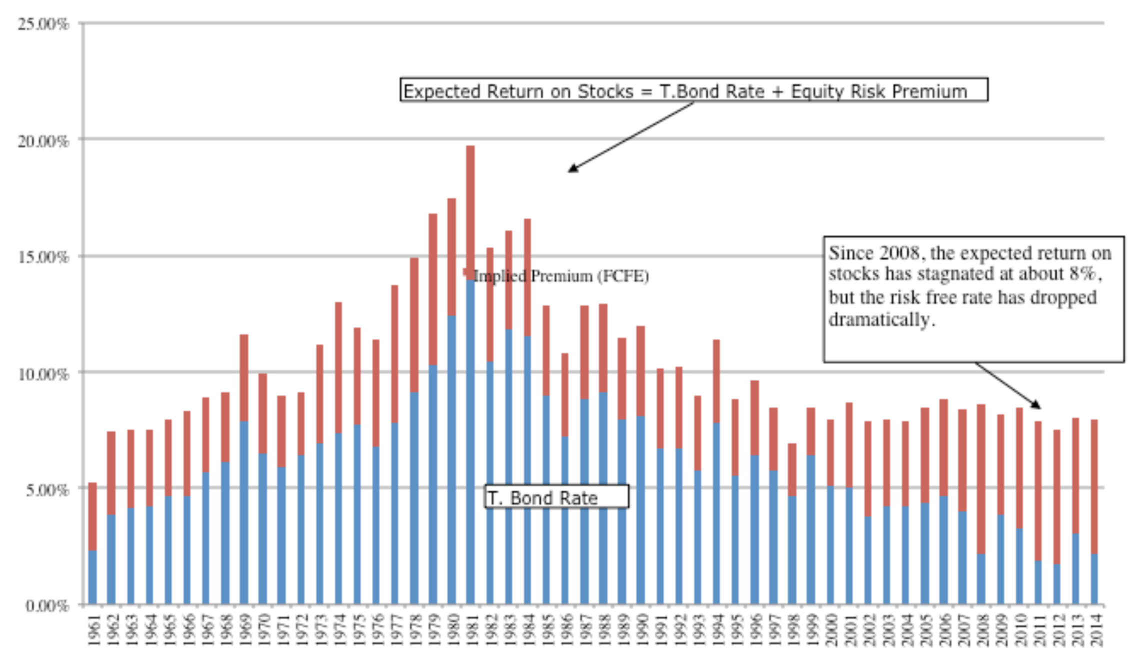
One of the advantages of estimating implied equity risk premiums, by period, is that we can track year to year changes in that number and relate those changes to shifts in interest rates, the macro environment or even to company characteristics. By doing so, not only can we get a better understanding of what causes equity risk premiums to change over time, but we are also able to come up with better estimates of future premiums.

Implied ERP and Interest rates

In much of valuation and corporate finance practice, we assume that the equity risk premium that we compute and use is unrelated to the level of interest rates. In particular, the use of historical risk premiums, where the premium is based upon an average premium earned over shifting risk free rates, implicitly assumes that the level of the premium is unchanged as the risk free rate changes. Thus, we use the same equity risk premium of 4.2% (the historical average for 1928-2012) on a risk free rate of 1.76% in 2012, as we would have, if the risk free rate had been 10%.

But is this a reasonable assumption? How much of the variation in the premium over time can be explained by changes in interest rates? Put differently, do equity risk premiums increase as the risk free rate increases or are they unaffected? To answer this question, we looked at the relationship between the implied equity risk premium and the treasury bond rate (risk free rate). As can be seen in figure 13, the implied equity risk premiums were highest in the 1970s, when interest rates and inflation were also high. However, there is contradictory evidence between 2008 and 2014, when high equity risk premiums accompanied low risk free rates.

Figure 13: Implied ERP and Risk free Rates



To examine the relationship between equity risk premiums and risk free rates, we ran a regression of the implied equity risk premium against both the level of long-term rates (the treasury bond rate) and the slope of the yield curve (captured as the difference between the 10-year treasury bond rate and the 3-month T.Bill rate), with the t statistics reported in brackets below each coefficient:

$$\begin{array}{lllll} \text{Implied ERP} = & 3.62\% & + 0.0570\ (\text{T.Bond Rate}) & + 0.0731\ (\text{T.Bond} - \text{T.Bill}) & R^2 = 2.54\% \\ & (8.45) & \quad (1.05) & \quad (0.37) & \end{array}$$

There is a mildly positive relationship between the T.Bond rate and implied equity risk premiums: every 1% increase in the treasury bond rate increases the equity risk premium by 0.06%. The slope of the yield curve seems to have little impact on the implied equity risk premium. Removing the latter variable and running the regression again:

Implied ERP = 3.74% + 0.0531 (T.Bond Rate) R^2=1.88%
(10.27) (1.00)

This regression does provide very weak support for the view that equity risk premiums should not be constant but should be linked to the level of interest rates. In fact, the regression can be used to estimate an equity risk premium, conditional on current interest rates. On March 14, 2015, for instance, when the 10-year treasury bond rate was 2.75%, the implied equity risk premium would have been computed as follows:

Implied ERP = 3.74% + 0.0531 (2.25%) = 3.86%

This would have been below the observed implied equity risk premium of about 5.78% and the average implied equity risk premium of 4.1% between 1960 and 2014. Put differently, given the low level of risk free rates in 2015 and the historical relationship between equity risk premiums and risk free rates, we would have expected the equity risk premium to be a much lower number (3.86%) than the actual number (5.78%).

Implied ERP and Macroeconomic variables

While we considered the interaction between equity risk premiums and interest rates in the last section, the analysis can be expanded to include other macroeconomic variables including economic growth, inflation rates and exchange rates. Doing so may give us a way of estimating an "intrinsic' equity risk premium, based upon macro economic variables, that is less susceptible to market moods and perceptions.

To explore the relationship, we estimated the correlation, between the implied equity risk premiums that we estimated for the S&P 500 and three macroeconomic variables – real GDP growth for the US, inflation rates (CPI) and exchange rates (trade weighted dollar), using data from 1973 to 2014, in table 21 (t statistics in brackets):

Table 21: Correlation Matrix: ERP and Macroeconomic variables: 1973-2015

	ERP	*Weighted Dollar*	*Real GDP*	*CPI*
ERP	1.0000			
Weighted dollar	-0.3492 (2.33)**	1.0000		
Real GDP	0.3883 (2.63)**	-0.1608 (01.02)	1.0000	
CPI	0.1452	-0.1550	0.0123	1.0000

	(0.92)	(0.98)	(0.08)	

** Statistically significant

The implied equity risk premium is positively correlated with GDP growth, decreasing as GDP growth increases and negatively correlated with the US dollar, with a stronger dollar going with lower implied equity risk premiums. The ERP is also mildly affected by inflation, with higher inflation going hand-in-hand with higher equity risk premiums.[118]

Following up on this analysis, we regressed equity risk premiums against the inflation rate, the weighted dollar and GDP growth, using data from 1974 to 2014:

Implied ERP = 4.21% - 0.1419 Real GDP growth + 0.1204 CPI + 0.0149 Weighted \$ R^2= 30.68%
(12.13) (1.90) (2.36) (0.67)

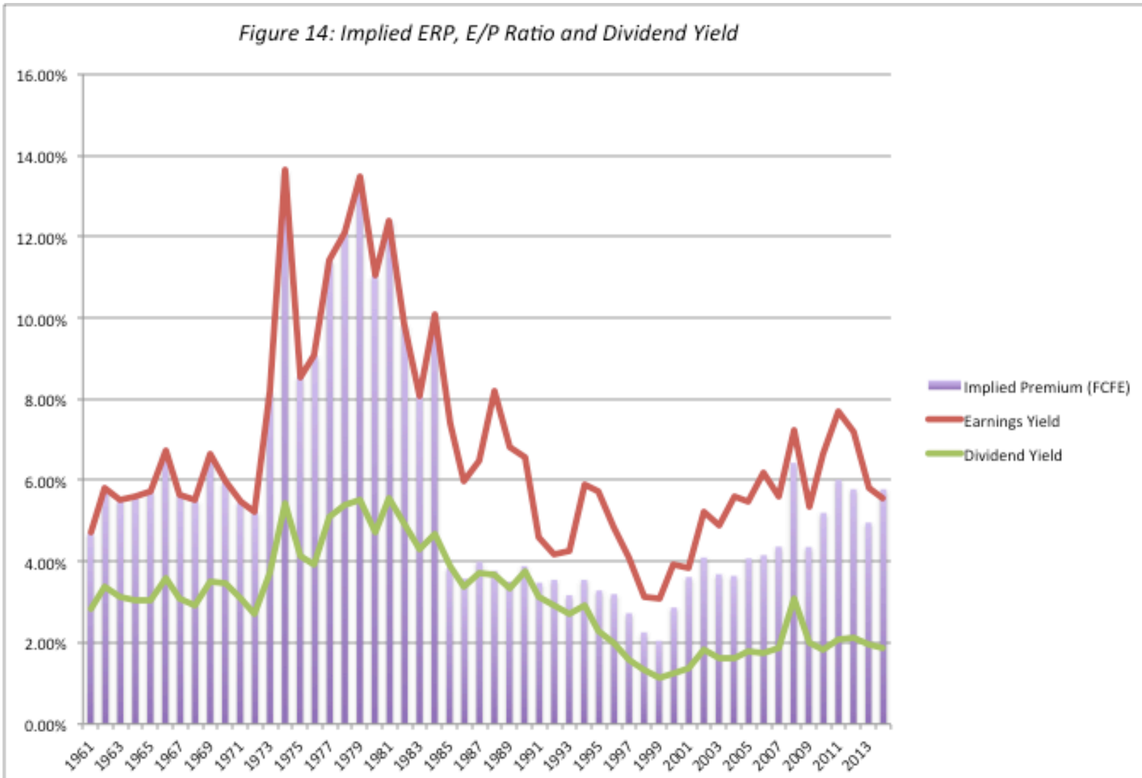
Based on this regression, every 1% increase in the inflation rate increases the equity risk premium by approximately 0.1204%, whereas every 1% increase in the growth rate in real GDP decreases the implied equity risk premium by 0.1419%.

From a risk perspective, it is not the level of GDP growth that matters, but uncertainty about that level; you can have low and stable economic growth and high and unstable economic growth. Since 2008, the economies of both developed and emerging markets have become more unstable over time and upended long held beliefs about developed economies. It will be interesting to see if equity risk premiums become more sensitive to real economic growth in this environment.

Implied ERP, Earnings Yields and Dividend Yields

Earlier in the paper, we noted that the dividend yield and the earnings yield (net of the risk free rate) can be used as proxies for the equity risk premium, if we make assumptions about future growth (stable growth, with the dividend yield) or expected excess returns (zero, with the earnings yield). In figure 14, we compare the implied equity risk premiums that we computed to the earnings and dividend yields for the S&P 500 from 1961 to 2014:

[118] The correlation was also computed for lagged and leading versions of these variables, with two material differences: the equity risk premium is negatively correlated with leading inflation rates and positively correlated with a leading weighted dollar.



Note that the dividend yield is a very close proxy for the implied equity risk premium until the late 1980s, when the two measures decoupled, a phenomenon that is best explained by the rise of stock buybacks as an alternative way of returning cash to stockholders.

The earnings yield, with the riskfree rate netted out, has generally not been a good proxy for the implied equity risk premium and would have yielded negative values for the equity risk premium (since you have to subtract out the risk free rate from it) through much of the 1990s. However, it does move with the implied equity risk premium. The difference between the earnings to price measure and the implied ERP can be attributed to a combination of higher earnings growth and excess returns that investors expect companies to deliver in the future. Analysts and academic researchers who use the earnings to price ratio as a proxy for forward-looking equity risk premiums may therefore end up with significant measurement error in their analyses.

Implied ERP and Technical Indicators

Earlier in the paper, we noted that any market timing forecast can be recast as a view on the future direction of the equity risk premium. Thus, a view that the market is under (over) priced and likely to go higher (lower is consistent with a belief that equity risk premiums will decline (increase) in the future. Many market timers do rely on technical indicators, such as moving averages and momentum measures, to make their judgment about market direction. To evaluate whether these approaches have a basis, you would need to look at how these measures are correlated with changes in equity risk premiums.

In a test of the efficacy of technical indicators, Neely, Rapach, Tu and Zhou (2011) compare the predictive power of macroeconomic/fundamental indications (including the interest rate, inflation, GDP growth and earnings/dividend yield numbers) with those of technical indicators (moving average, momentum and trading volume) and conclude that the latter better explain movements in stock returns.[119] They conclude that a composite prediction, that incorporates both macroeconomic and technical indicators, is superior to using just one set or the other of these variables. Note, however, that their study focused primarily on the predictability of stock returns over the next year and not on longer term equity risk premiums.

Extensions of Implied Equity Risk Premium

The process of backing out risk premiums from current prices and expected cashflows is a flexible one. It can be expanded into emerging markets to provide estimates of risk premiums that can replace the country risk premiums we developed in the last section. Within an equity market, it can be used to compute implied equity risk premiums for individual sectors or even classes of companies.

Other Equity Markets

The advantage of the implied premium approach is that it is market-driven and current, and does not require any historical data. Thus, it can be used to estimate implied equity premiums in any market, no matter how short its history, It is, however, bounded by whether the model used for the valuation is the right one and the availability and reliability of the inputs to that model. Earlier in this paper, we estimated country risk premiums for Brazil, using default spreads and equity market volatile. To provide a contrast, we estimated the implied equity risk premium for the Brazilian equity market in September 2009, from the following inputs.

- The index (Bovespa) was trading at 61,172 on September 30, 2009, and the dividend yield on the index over the previous 12 months was approximately 2.2%. While stock buybacks represented negligible cash flows, we did compute the FCFE for companies in the index, and the aggregate FCFE yield across the companies was 4.95%.

---

[119] Neely, C.J., D.E. Rapach, J. Tu and G. Zhou, 2011, *Forecasting the Equity Risk Premium: The Role of Technical Indicators*, Working Paper, http://ssrn.com/abstract=1787554.

- Earnings in companies in the index are expected to grow 6% (in US dollar terms) over the next 5 years, and 3.45% (set equal to the treasury bond rate) thereafter.
- The riskfree rate is the US 10-year treasury bond rate of 3.45%.

The time line of cash flows is shown below:

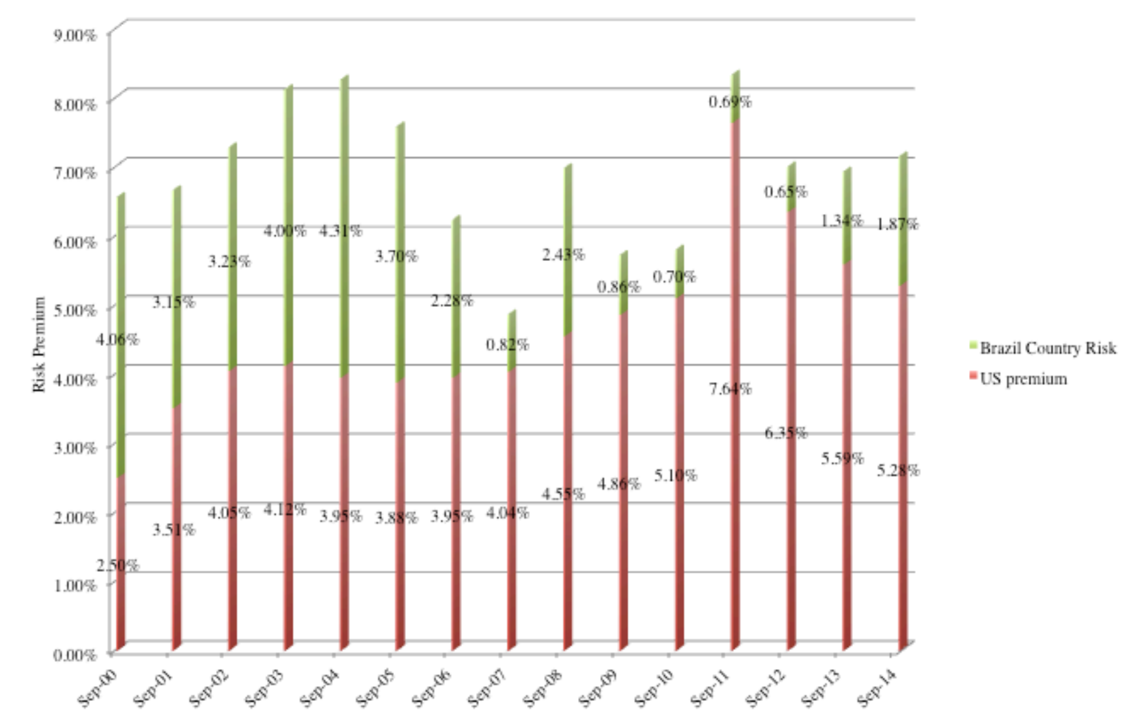
$$61{,}272 = \frac{3210}{(1+r)} + \frac{3{,}402}{(1+r)^2} + \frac{3{,}606}{(1+r)^3} + \frac{3{,}821}{(1+r)^4} + \frac{4{,}052}{(1+r)^5} + \frac{4{,}052(1.0345)}{(r-.0345)(1+r)^5}$$

These inputs yield a required return on equity of 9.17%, which when compared to the treasury bond rate of 3.45% on that day results in an implied equity premium of 5.72%. For simplicity, we have used nominal dollar expected growth rates[120] and treasury bond rates, but this analysis could have been done entirely in the local currency.

One of the advantages of using implied equity risk premiums is that that they are more sensitive to changing market conditions. The implied equity risk premium for Brazil in September 2007, when the Bovespa was trading at 73512, was 4.63%, lower than the premium in September 2009, which in turn was much lower than the premium prevailing in September 2014. In figure 15, we trace the changes in the implied equity risk premium in Brazil from September 2000 to September 2014 and compare them to the implied premium in US equities:

[120] The input that is most difficult to estimate for emerging markets is a long-term expected growth rate. For Brazilian stocks, I used the average consensus estimate of growth in earnings for the largest Brazilian companies which have ADRs listed on them. This estimate may be biased, as a consequence.

Figure 15: Implied Equity Risk Premium - Brazil



Implied equity risk premiums in Brazil declined steadily from 2003 to 2007, with the September 2007 numbers representing a historic low. They surged in September 2008, as the crisis unfolded, fell back in 2009 and 2010 but increased again in 2011. In fact, the Brazil portion of the implied equity risk premium fell to its lowest level in ten years in September 2010, a phenomenon that remained largely unchanged in 2011 and 2012. Political turmoil and corruptions scandals have combined to push the premium back up again in the last year or two.

Computing and comparing implied equity risk premiums across multiple equity markets allows us to pinpoint markets that stand out, either as over priced (because their implied premiums are too low, relative to other markets) or under priced (because their premiums at too high, relative to other markets). In September 2007, for instance, the implied equity risk premiums in India and China were roughly equal to or even lower than the implied premium for the United States, computed at the same time. Even an optimist on future growth these countries would be hard pressed to argue that equity markets in these markets and the United States were of equivalent risk, which would lead us to conclude that these stocks were overvalued relative to US companies.

One final note is worth making. Over the last decade, the implied equity risk premiums in the largest emerging markets – India, China and Brazil- have all declined substantially, relative to developed markets. In table 22, we summarize implied equity risk premiums for developed and emerging markets from 2001 and 2013, making simplistic assumptions about growth and stable growth valuation models:[121]

Table 22: Developed versus Emerging Market Equity Risk Premiums

Start of year	*PBV Developed*	*PBV Emerging*	*ROE (Dev)*	*ROE (Emerg)*	*US T.Bond*	*Cost of Equity (Developed)*	*Cost of Equity (Emerging)*	*Differential ERP*
2004	2.00	1.19	10.81%	11.65%	4.25%	7.28%	10.63%	3.35%
2005	2.09	1.27	11.12%	11.93%	4.22%	7.26%	10.50%	3.24%
2006	2.03	1.44	11.32%	12.18%	4.39%	7.55%	10.11%	2.56%
2007	1.67	1.67	10.87%	12.88%	4.70%	8.19%	10.00%	1.81%
2008	0.87	0.83	9.42%	11.12%	4.02%	10.30%	12.37%	2.07%
2009	1.20	1.34	8.48%	11.02%	2.21%	7.35%	9.04%	1.69%
2010	1.39	1.43	9.14%	11.22%	3.84%	7.51%	9.30%	1.79%
2011	1.12	1.08	9.21%	10.04%	3.29%	8.52%	9.61%	1.09%
2012	1.17	1.18	9.10%	9.33%	1.88%	7.98%	8.35%	0.37%
2013	1.56	1.63	8.67%	10.48%	1.76%	6.02%	7.50%	1.48%
2014	1.95	1.50	9.27%	9.64%	3.04%	6.00%	7.77%	1.77%
2015	1.88	1.56	9.69%	9.75%	2.17%	5.94%	7.39%	1.45%

The trend line from 2004 to 2012 is clear as the equity risk premiums, notwithstanding a minor widening in 2008, have converged in developed and emerging markets, suggesting that globalization has put "emerging market risk" into developed markets, while creating "developed markets stability factors" (more predictable government policies, stronger legal and corporate governance systems, lower inflation and stronger currencies) in emerging markets. In the last two years, we did see a correction in emerging markets that pushed the premium back up, albeit to a level that was still lower than it was prior to 2010.

Sector premiums

Using current prices and expected future cash flows to back out implied risk premiums is not restricted to market indices. We can employ the approach to estimate the implied equity risk premium for a specific sector at a point in time. In September 2008,

[121] We start with the US treasury bond rate as the proxy for global nominal growth (in US dollar terms), and assume that the expected growth rate in developed markets is 0.5% lower than that number and the expected growth rate in emerging markets is 1% higher than that number. The equation used to compute the ERP is a simplistic one, based on the assumptions that the countries are in stable growth and that the return on equity in each country is a predictor of future return on equity:

PBV = (ROE – g)/ (Cost of equity –g)

Cost of equity = (ROE –g + PBV(g))/ PBV

for instance, there was a widely held perception that investors were attaching much higher equity risk premiums to commercial bank stocks, in the aftermath of the failures of Fannie Mae, Freddie Mac, Bear Stearns and Lehman. To test this proposition, we took a look at the S&P Commercial Bank index, which was trading at 318.26 on September 12, 2008, with an expected dividend yield of 5.83% for the next 12 months. Assuming that these dividends will grow at 4% a year for the next 5 years and 3.60% (the treasury bond rate) thereafter, well below the nominal growth rate in the overall economy, we arrived at the following equation:

$$318.26 = \frac{19.30}{(1+r)} + \frac{20.07}{(1+r)^2} + \frac{20.87}{(1+r)^3} + \frac{21.71}{(1+r)^4} + \frac{22.57}{(1+r)^5} + \frac{22.57(1.036)}{(r-.036)(1+r)^5}$$

Solving for the expected return yields a value of 9.74%, which when netted out against the riskfree rate at the time (3.60%) yields an implied premium for the sector:

Implied ERP for Banking in September 2008 = 9.74% - 3.60% = 6.14%

How would we use this number? One approach would be to compare it to the average implied premium in this sector over time, with the underlying assumption that the value will revert back to the historical average for the sector. The implied equity risk premium for commercial banking stocks was close to 4% between 2005 and 2007, which would lead to the conclusion that banking stocks were undervalued in September 2008. The other is to assume that the implied equity premium for a sector is reflective of perceptions of future risk in that sector; in September 2008, there can be no denying that financial service companies faced unique risks and the market was reflecting these risks in prices. As a postscript, the implied equity risk premium for financial service firms was 5.80% in January 2012, just below the market-implied premium at the time (6.01%), suggesting that some of the post-crisis fear about banking stocks had receded.

A note of caution has to be added to about sector-implied premiums. Since these risk premiums consolidate both sector risk and market risk, it would be inappropriate to multiply these premiums by conventional betas, which are measures of sector risk. Thus, multiplying the implied equity risk premium for the technology sector (which will yield a high value) by a market beta for a technology company (which will also be high for the same reason) will result in double counting risk.[122]

Firm Characteristics

Earlier in this paper, we talked about the small firm premium and how it has been estimated using historical data, resulting in backward looking estimates with substantial

[122] You could estimate betas for technology companies against a technology index (rather than the market index) and use these betas with the implied equity risk premium for technology companies.

standard error. We could use implied premiums to arrive at more forward looking estimates, using the following steps:

Step 1: Compute the implied equity risk premium for the overall market, using a broad index such as the S&P 500. Earlier in this paper, we estimated this, as of January 2015, to be 5.78%.

Step 2: Compute the implied equity risk premium for an index containing primarily or only small cap firms, such as the S&P 600 Small Cap Index. On January 1, 2015, the index was trading at 695.08, with an aggregated FCFE yield of about 3.76% (yielding a FCFE for the most recent year of 26.14), and an expected growth rate in earnings of 10.25% for the next 5 years. Using these values, in conjunction with the prevailing riskfree rate of 2.17%, yields the following equation:

$$695.08 = \frac{28.81}{(1+r)} + \frac{31.77}{(1+r)^2} + \frac{35.02}{(1+r)^3} + \frac{38.61}{(1+r)^4} + \frac{42.57}{(1+r)^5} + \frac{42.57\ (1.0217)}{(r-.0217)(1+r)^5}$$

Solving for the expected return, we get:

Expected return on small cap stocks = 7.61%

Implied equity risk premium for small cap stocks = 7.61% -2.17% = 5.44%

Step 3: The forward-looking estimate of the small cap premium should be the difference between the implied premium for small cap stocks (in step 2) and the implied premium for the market (in step 1).

Small cap premium = 5.44% - 5.78% = -0.34%

With the numbers in January 2015, small caps are priced to generate an expected return that is lower than the rest of the market, thus putting into question the wisdom of using the 4-5% small cap premium in computing costs of equity.

This approach to estimating premiums can be extended to other variables. For instance, one of the issues that has challenged analysts in valuation is how to incorporate the illiquidity of an asset into its estimated value. While the conventional approach is to attach an illiquidity discount, an alternative is to adjust the discount rate upwards for illiquid assets. If we compute the implied equity risk premiums for stocks categorized by illiquidity, we may be able to come up with an appropriate adjustment. For instance, you could estimate the implied equity risk premium for the stocks that rank in the lowest decile in terms of illiquidity, defined as turnover ratio.[123] Comparing this value to the implied premium for the S&P 500 of 5.78% should yield an implied illiquidity risk premium. Adding this premium to the cost of equity for relatively illiquid investments will then discount the value of these investments for illiquidity.

---

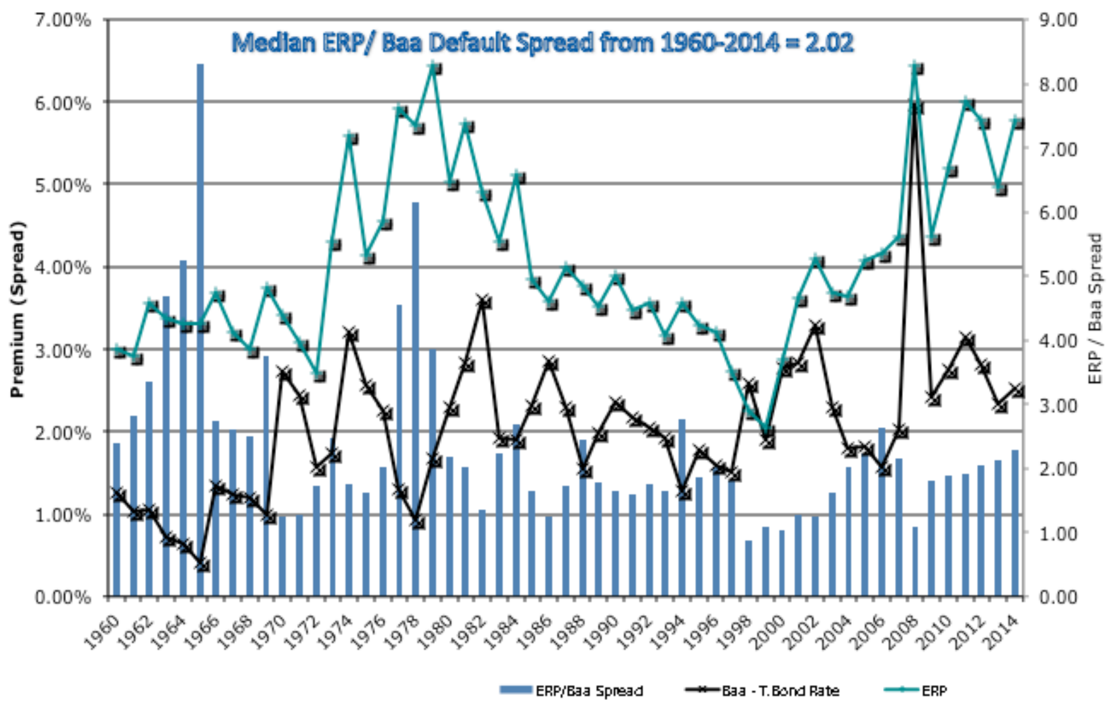
[123] The turnover ratio is obtained by dividing $ trading volume in a stock by its market capitalization at that time.

2. Default Spread Based Equity Risk Premiums

While we think of corporate bonds, stocks and real estate as different asset classes, it can be argued that they are all risky assets and that they should therefore be priced consistently. Put another way, there should be a relationship across the risk premiums in these asset classes that reflect their fundamental risk differences. In the corporate bond market, the default spread, i.e, the spread between the interest rate on corporate bonds and the treasury bond rate, is used as the risk premium. In the equity market, as we have seen through this paper, historical and implied equity premiums have tussled for supremacy as the measure of the equity risk premium. In the real estate market, no mention is made of an explicit risk premium, but real estate valuations draw heavily on the "capitalization rate", which is the discount rate applied to a real estate property's earnings to arrive at an estimate of value. The use of higher (lower) capitalization rates is the equivalent of demanding a higher (lower) risk premium.

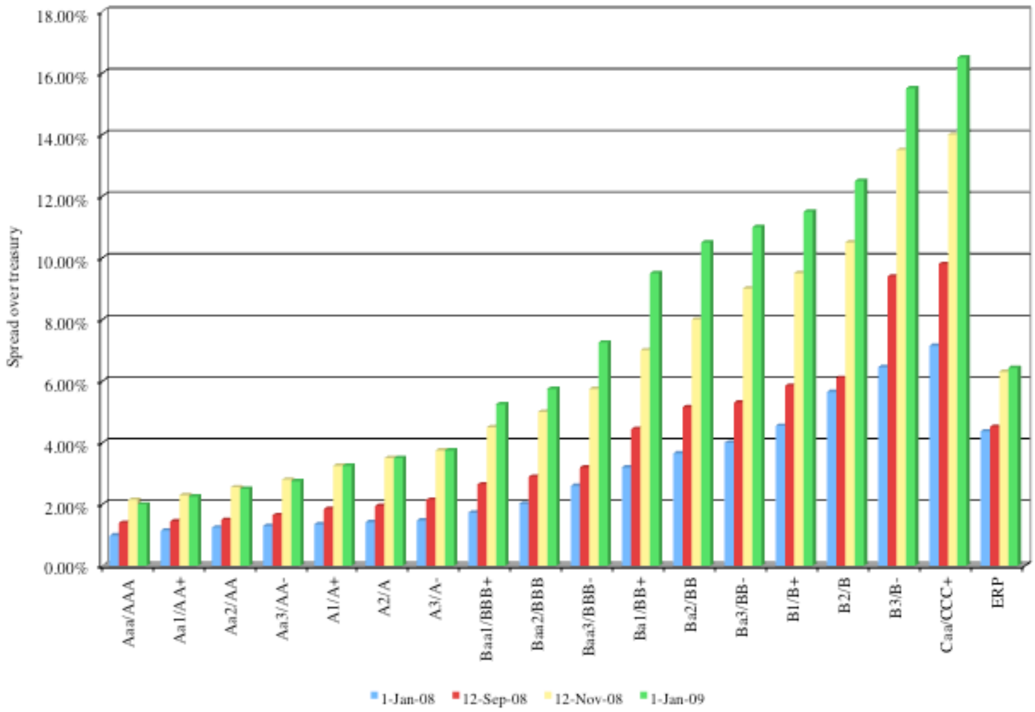
Of these three premiums, the default spread is the less complex and the most widely accessible data item. If equity risk premiums could be stated in terms of the default spread on corporate bonds, the estimation of equity risk premiums would become immeasurably simpler. For instance, assume that the default spread on Baa rated corporate bonds, relative to the ten-year treasury bond, is 2.2% and that equity risk premiums are routinely twice as high as Baa bonds, the equity risk premium would be 4.4%. Is such a rule of thumb even feasible? To answer this question, we looked at implied equity risk premiums and Baa-rated corporate bond default spreads from 1960 to 2014 in Figure 16.

Figure 16: Equity Risk Premiums and Bond Default Spreads



Note that both default spreads and equity risk premiums jumped in 2008, with the former increasing more on a proportionate basis. The ratio of 1.08 (ERP/ Baa Default Spread) at the end of 2008 was close to the lowest value in the entire series, suggesting that either equity risk premiums were too low or default spreads were too high. At the end of 2013, both the equity risk premium and the default spread increased, and the ratio moved back to 2.12, a little higher than the median value of 2.02 for the entire time period. The connection between equity risk premiums and default spreads was most obvious during 2008, where changes in one often were accompanied by changes in the other. Figure 17 graphs out changes in default spreads and ERP over the tumultuous year:

Figure 17: Default Spreads on Ratings Classes



How could we use the historical relationship between equity risk premiums and default spreads to estimate a forward-looking equity risk premium? On January 1, 2015, the default spread on a Baa rated bond was 2.52%. Applying the median ratio of 2.02, estimated from 1960-2014 numbers, to the Baa default spread of 2.52% results in the following estimate of the ERP:

Default Spread on Baa bonds (over treasury) on 1/1/2015 = 2.52%

Imputed Equity Risk Premium = Default Spread * Median ratio or ERP/Spread

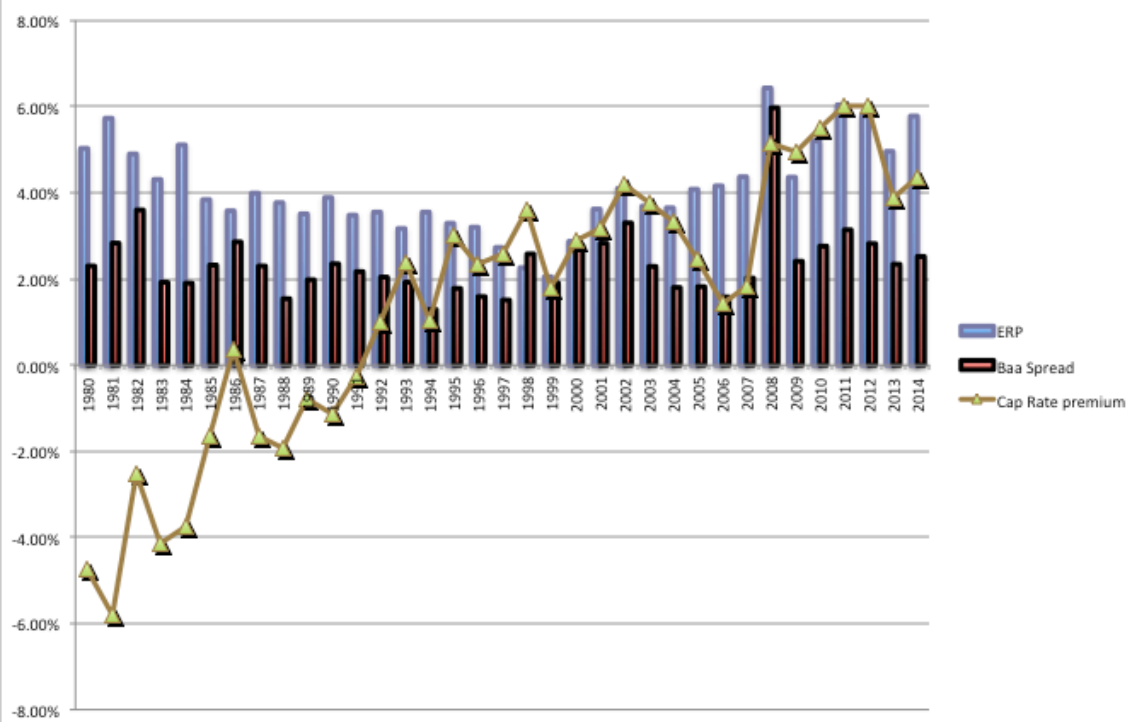
= 2.52%* 2.02 = 5.10%

This is a little lower than the implied equity risk premium of 5.78% that we computed in January 2015. Note that there is significant variation in the ratio (of ERP to default spreads) over time, with the ratio dropping below one at the peak of the dot.com boom (when equity risk premiums dropped to 2%) and rising to as high as 2.63 at the end of 2006; the standard error in the estimate is 0.20. Whenever the ratio has deviated significantly from the average, though, there is reversion back to that median over time.

The capitalization rate in real estate, as noted earlier, is a widely used number in the valuation of real estate properties. For instance, a capitalization rate of 10%, in conjunction with an office building that generates income of $ 10 million, would result in a property value of $ 100 million ($10/.10). The difference between the capitalization ratio and the treasury bond rate can be considered a real estate market risk premium, In

Figure 18, we used the capitalization rate in real estate ventures and compared the risk premiums imputed for real estate with both bond default spreads and implied equity risk premiums between 1980 and 2014.

Figure 18: Equity Risk Premiums, Cap Rates and Bond Spreads



The story in this graph is the convergence of the real estate and financial asset risk premiums. In the early 1980s, the real estate market seems to be operating in a different risk/return universe than financial assets, with the cap rates being less than the treasury bond rate. For instance, the cap rate in 1980 was 8.1%, well below the treasury bond rate of 12.8%, resulting in a negative risk premium for real estate. The risk premiums across the three markets - real estate, equity and bonds - starting moving closer to each other in the late 1980s and the trend accelerated in the 1990s. We would attribute at least some of this increased co-movement to the securitization of real estate in this period. In 2008, the three markets moved almost in lock step, as risk premiums in the markets rose and prices fell. The housing bubble of 2004-2008 is manifested in the drop in the real estate equity risk premium during those years, bottoming out at less than 2% at the 2006. The correction in housing prices since has pushed the premium back up. Both equity and bond premiums have adjusted quickly to pre-crisis levels in 2009 and 2010, and real estate premiums are following, albeit at a slower pace.

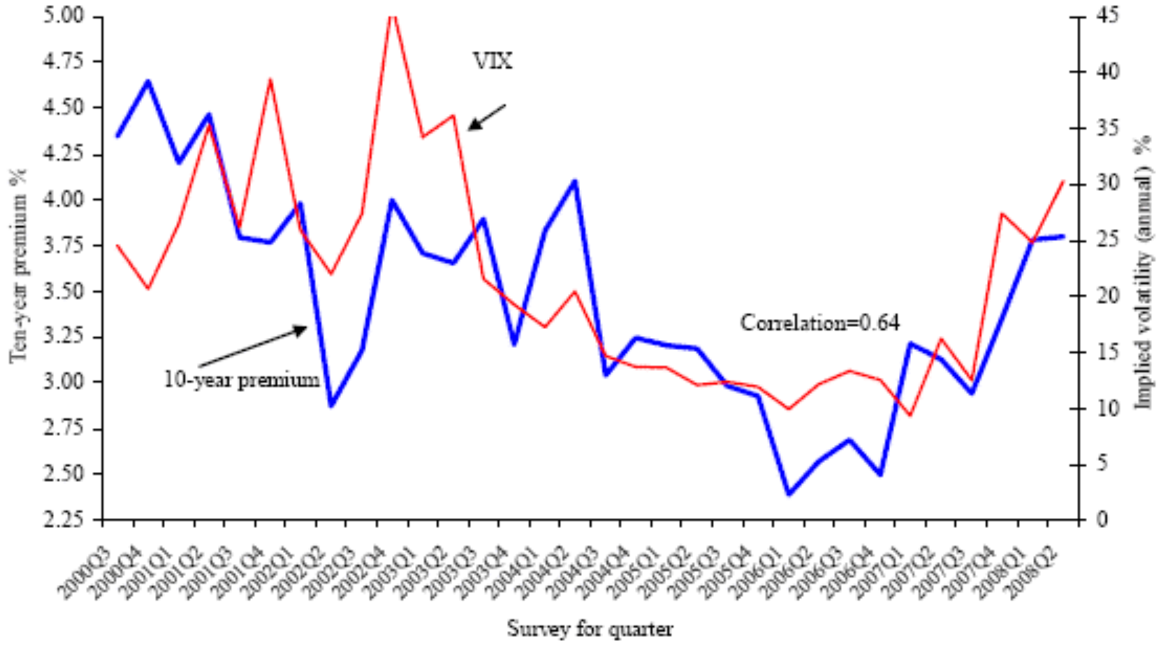
While the noise in the ratios (of ERP to default spreads and cap rates) is too high for us to develop a reliable rule of thumb, there is enough of a relationship here that we would suggest using this approach as a secondary one to test to see whether the equity risk premiums that we are using in practice make sense, given how risky assets are being priced in other markets. Thus, using an equity risk premium of 2%, when the Baa default spread is approximately at the same level strikes us as imprudent, given history. For macro strategists, there is a more activist way of using these premiums. When risk premiums in markets diverge, there is information in the relative pricing. Thus, the drop in equity risk premiums in the late 1990s, as default spreads stayed stable, would have signaled that the equity markets were overvalued (relative to bonds), just as the drop in default spreads between 2004 and 2007, while equity risk premiums were stagnant, would have suggested the opposite.

3. Option Pricing Model based Equity Risk Premium

There is one final approach to estimating equity risk premiums that draws on information in the option market. In particular, option prices can be used to back out implied volatility in the equity market. To the extent that the equity risk premium is our way of pricing in the risk of future stock price volatility, there should be a relationship between the two.

The simplest measure of volatility from the options market is the volatility index (VIX), which is a measure of 30—day volatility constructed using the implied volatilities in traded S&P 500 index options. The CFO survey premium from Graham and Harvey that we referenced earlier in the paper found a high degree of correlation between the premiums demanded by CFOs and the VIX value (see figure 19 below):

Figure 19: Volatility Index (VIX) and Survey Risk Premiums



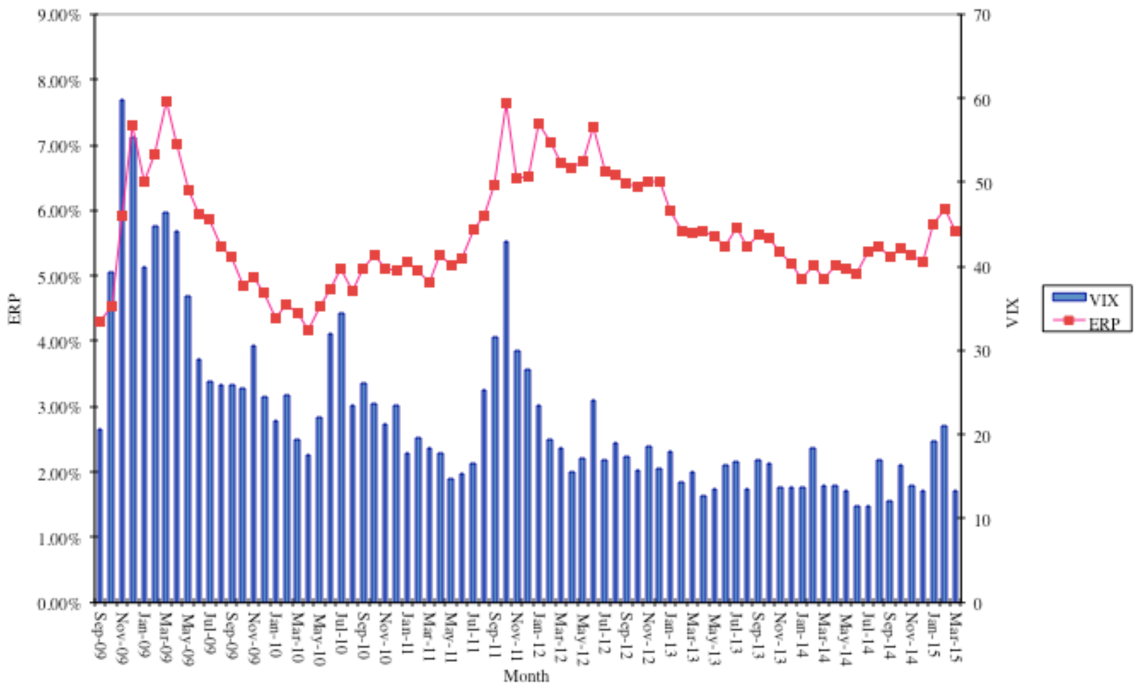
Santa-Clara and Yan (2006) use options on the S&P 500 to estimate the ex-ante risk assessed by investors from 1996 and 2002 and back out an implied equity risk premium on that basis.[124] To estimate the ex-ante risk, they allow for both continuous and discontinuous (or jump) risk in stocks, and use the option prices to estimate the probabilities of both types of risk. They then assume that investors share a specific utility function (power utility) and back out a risk premium that would compensate for this risk. Based on their estimates, investors should have demanded an equity risk premium of 11.8% for their perceived risk and that the perceived risk was about 70% higher than the realized risk over this period.

The link between equity market volatility and the equity risk premium also became clearer during the market meltdown in the last quarter of 2008. Earlier in the paper, we noted the dramatic shifts in the equity risk premiums, especially in the last year, as the financial crisis has unfolded. In Figure 20, we look at the implied equity risk premium each month from September 2008 to March 2014 and the volatility index (VIX) for the S&P 500:

---

[124] Santa-Clara, P. and S. Yan, 2006, *Crashes, Volatility, and the Equity Premium: Lessons from S&P 500 Options*, Review of Economics and Statistics, v92, pg 435-451.

Figure 20: ERP versus VIX



Note that the surge in equity risk premiums between September 2008 and December 2008 coincided with a jump in the volatility index and that both numbers have declined in the years since the crisis. The drop in the VIX between September 2011 and March 2012 was not accompanied by a decrease in the implied equity risk premium, but equity risk premiums drifted down in the year after. While the VIX stayed low for much of 2014, equity risk premiums climbed through the course of the year.

In a paper referenced earlier, Bollerslev, Tauchen and Zhou (2009) take a different tack and argue that it is not the implied volatility per se, but the variance risk, i.e., the difference between the implied variance (in option prices) and the actual variance, that drives expected equity returns.[125] Thus, if the realized variance in a period is far higher (lower) than the implied variance, you should expect to see higher (lower) equity risk premiums demanded for subsequent periods. While they find evidence to back this proposition, they also note the relationship is strongest for short term returns (next quarter) and are weaker for longer-term returns. Bekaert and Hoerova (2013) decomposed the squared VIX into two components, a conditional variance of the stock

[125] Bollerslev, T. G. Tauchen and H. Zhou, 2009, *Expected Stock Returns and Variance Risk Premia*, Review of Financial Studies, v22, 4463-4492.

market and an equity variance premium, and conclude that while the latter is a significant predictor of stock returns but the former is not.[126]

Choosing an Equity Risk Premium

We have looked at three different approaches to estimating risk premiums, the survey approach, where the answer seems to depend on who you ask and what you ask them, the historical premium approach, with wildly different results depending on how you slice and dice historical data and the implied premium approach, where the final number is a function of the model you use and the assumptions you make about the future. Ultimately, thought, we have to choose a number to use in analysis and that number has consequences. In this section, we consider why the approaches give you different numbers and a pathway to use to devise which number is best for you.

Why do the approaches yield different values?

The different ways of estimating equity risk premium provide cover for analysts by providing justification for almost any number they choose to use in practice. No matter what the premium used by an analyst, whether it be 3% or 12%, there is back-up evidence offered that the premium is appropriate. While this may suffice as a legal defense, it does not pass muster on common sense grounds since not all risk premiums are equally justifiable. To provide a measure of how the numbers vary, the values that we have attached to the US equity risk premium, using different approaches, in January 2013 are summarized in table 23.

Table 23: Equity Risk Premium (ERP) for the United States – January 2013

Approach Used	*ERP*	*Additional information*
Survey: CFOs	3.73%	Campbell and Harvey survey of CFOs (2014); Average estimate. Median was 3.4%.
Survey: Global Fund Managers	4.60%	Merrill Lynch (January 2014) survey of global managers
Historical - US	4.60%	Geometric average - Stocks over T.Bonds: 1928-2014
Historical – Multiple Equity Markets	2.80%	Average premium across 20 markets from 1900-2014: Dimson, Marsh and Staunton (2015)
Current Implied premium	5.78%	From S&P 500 – January 1, 2015
Average Implied premium	4.13%	Average of implied equity risk premium: 1960-2014

[126] Bekaert, G. and M. Hoerova, 2013, *The VIX, Variance Premium and Stock Market Volatility*, SSRN Working Paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2342200.

Implied premium adjusted for T.Bond rate and term structure	3.86%	Using regression of implied premium on T.Bond rate
Default spread based premium	5.10%	Baa Default Spread * Median value of (ERP/ Default Spread)

The equity risk premiums, using the different approaches, yield a range, with the lowest value being 2.80% and the highest being 5.78%. Note that the range would have been larger if we used other measures of historical risk premiums: different time periods, arithmetic instead of geometric averages.

There are several reasons why the approaches yield different answers much of time and why they converge sometimes.

1. When stock prices enter an extended phase of upward (downward) movement, the historical risk premium will climb (drop) to reflect past returns. Implied premiums will tend to move in the opposite direction, since higher (lower) stock prices generally translate into lower (higher) premiums. In 1999, for instance, after the technology induced stock price boom of the 1990s, the implied premium was 2% but the historical risk premium was almost 6%.
2. Survey premiums reflect historical data more than expectations. When stocks are going up, investors tend to become more optimistic about future returns and survey premiums reflect this optimism. In fact, the evidence that human beings overweight recent history (when making judgments) and overreact to information can lead to survey premiums overshooting historical premiums in both good and bad times. In good times, survey premiums are even higher than historical premiums, which, in turn, are higher than implied premiums; in bad times, the reverse occurs.
3. When the fundamentals of a market change, either because the economy becomes more volatile or investors get more risk averse, historical risk premiums will not change but implied premiums will. Shocks to the market are likely to cause the two numbers to deviate. After the terrorist attack on the World Trade Center in September 2001, for instance, implied equity risk premiums jumped almost 0.50% but historical premiums were unchanged (at least until the next update).

In summary, we should not be surprised to see large differences in equity risk premiums as we move from one approach to another, and even within an approach, as we change estimation parameters.

Which approach is the "best" approach?

If the approaches yield different numbers for the equity risk premium, and we have to choose one of these numbers, how do we decide which one is the "best" estimate? The answer to this question will depend upon several factors:

a. Predictive Power: In corporate finance and valuation, what we ultimately care about is the equity risk premium for the future. Consequently, the approach that has the best predictive power, i.e. yields forecasts of the risk premium that are closer to realized premiums, should be given more weight. So, which of the approaches does best on this count?

   Campbell and Shiller (1988) suggested that the dividend yield, a simplistic measure of the implied equity risk premium, had significant predictive power for future returns.[127] However, Goyal and Welch (2007) examined many of the measures suggested as predictors of the equity risk premium in the literature, including the dividend yield and the earnings to price ratio, and find them all wanting.[128] Using data from 1926 to 2005, they conclude that while the measures do reasonably well in sample, they perform poorly out of sample, suggesting that the relationships in the literature are either spurious or unstable. Campbell and Thompson (2008) disagree, noting that putting simple restrictions on the predictive regressions improve out of sample performance for many predictive variables.[129]

   To answer this question, we looked at the implied equity risk premiums from 1960 to 2014 and considered four predictors of this premium – the historical risk premium through the end of the prior year, the implied equity risk premium at the end of the prior year, the average implied equity risk premium over the previous five years and the premium implied by the Baa default spread. Since the survey data does not go back very far, we could not test the efficacy of the survey premium. Our results are summarized in table 24:

[127] Campbell, J. Y. and R. J. Shiller. 1988, *The Dividend-Price Ratio And Expectations Of Future Dividends And Discount Factors*, Review of Financial Studies, v1(3), 195-228.

[128] Goyal, A. and I. Welch, 2007, *A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,* Review of Financial Studies, v21, 1455-1508.

[129] Campbell, J.Y., and S.B. Thompson, 2008, *Predictive Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?* Review of Financial Studies, v21, 150-9-1531.

Table 24: Predictive Power of different estimates- 1960 - 2014

Predictor	*Correlation with implied premium next year*	*Correlation with actual return- next 5 years*	*Correlation with actual return – next 10 years*[130]
Current implied premium	0.736	0.352	0.500
Average implied premium: Last 5 years	0.684	0.238	0.449
Historical Premium	-0.460	-0.365	-0.466
Default Spread based premium	0.047	0.148	0.165

Over this period, the implied equity risk premium at the end of the prior period was the best predictor of the implied equity risk premium in the next period, whereas historical risk premiums did worst. If we extend our analysis to make forecasts of the actual return premium earned by stocks over bonds for the next five or ten years, the current implied equity risk premium still yields the best forecast for the future, though default spread based premiums improve as predictors. Historical risk premiums perform even worse as forecasts of actual risk premiums over the next 5 or 10 years. If predictive power were the only test, historical premiums clearly fail the test.

b. Beliefs about markets: Implicit in the use of each approach are assumptions about market efficiency or lack thereof. If you believe that markets are efficient in the aggregate, or at least that you cannot forecast the direction of overall market movements, the current implied equity premium is the most logical choice, since it is estimated from the current level of the index. If you believe that markets, in the aggregate, can be significantly overvalued or undervalued, the historical risk premium or the average implied equity risk premium over long periods becomes a better choice. If you have absolutely no faith in markets, survey premiums will be the choice.

c. Purpose of the analysis: Notwithstanding your beliefs about market efficiency, the task for which you are using equity risk premiums may determine the right risk premium to use. In acquisition valuations and equity research, for instance, you are

130 I computed the compounded average return on stocks in the following five (ten) years and netted out the compounded return earned on T.Bonds over the following five (ten) years. This was a switch from the simple arithmetic average of returns over the next 10 years that I was using until last year's survey.

asked to assess the value of an individual company and not take a view on the level of the overall market. This will require you to use the current implied equity risk premium, since using any other number will bring your market views into the valuation. To see why, assume that the current implied premium is 4% and you decide to use a historical premium of 6% in your company valuation. Odds are that you will find the company to be over valued, but a big reason for your conclusion is that you started off with the assumption that the market itself is over valued by about 25-30%.[131] To make yourself market neutral, you will have to stick with the current implied premium. In corporate finance, where the equity risk premium is used to come up with a cost of capital, which in turn determines the long-term investments of the company, it may be more prudent to build in a long-term average (historical or implied) premium.

In conclusion, there is no one approach to estimating equity risk premiums that will work for all analyses. If predictive power is critical or if market neutrality is a pre-requisite, the current implied equity risk premium is the best choice. For those more skeptical about markets, the choices are broader, with the average implied equity risk premium over a long time period having the strongest predictive power. Historical risk premiums are very poor predictors of both short-term movements in implied premiums or long-term returns on stocks.

As a final note, there are papers that report consensus premiums, often estimated by averaging across approaches. I remain skeptical about these estimates, since the approaches vary not only in terms of accuracy and predictive power but also in their philosophy. Averaging a historical risk premium with an implied premium may give an analyst a false sense of security but it really makes no sense since they represent different views of the world and push in different directions.

Five myths about equity risk premiums

There are widely held misconceptions about equity risk premiums that we would like to dispel in this section.

1. Services "know" the risk premium: When Ibbotson and Sinquefield put together the first database of historical returns on stocks, bonds and bills in the 1970s, the data that they used was unique and not easily replicable, even for professional money managers. The niche they created, based on proprietary data, has led some to believe that Ibbotson Associates, and data services like them, have the capacity to read the

[131] If the current implied premium is 4%, using a 6% premium on the market will reduce the value of the index by about 25-30%.

historical data better than the rest of us, and therefore come up with better estimates. Now that the access to data has been democratized, and we face a much more even playing field, there is no reason to believe that any service has an advantage over any other, when it comes to historical premiums. Analysts should no longer be allowed to hide behind the defense that the equity risk premiums they use come from a reputable service and are thus beyond questioning.

2. There is no right risk premium: The flip side of the "services know it best" argument is that the data is so noisy that no one knows what the right risk premium is, and that any risk premium within a wide range is therefore defensible. As we have noted in this paper, it is indeed possible to arrive at outlandishly high or low premiums, but only if you use estimation approaches that do not hold up to scrutiny. The arithmetic average premium from 2005 to 2014 for stocks over treasury bonds is an equity risk premium estimate, but it is not a good one.
3. The equity risk premium does not change much over time: Equity risk premiums reflect both economic fundamentals and investor risk aversion and they do change over time, sometimes over very short intervals, as evidenced by what happened in the last quarter of 2008. Shocks to the system – a collapse of a large company or sovereign entity or a terrorist attack – can cause premiums to shoot up overnight. A failure to recognize this reality will lead to analyses that lag reality.
4. Using the same premium is more important than using the right premium: Within many investment banks, corporations and consulting firms, the view seems to be that getting all analysts to use the same number as the risk premium is more important than testing to see whether that number makes sense. Thus, if all equity research analysts use 5% as the equity risk premium, the argument is that they are all being consistent. There are two problems with this argument. The first is that using a premium that is too high or low will lead to systematic errors in valuation. For instance, using a 5% risk premium across the board, when the implied premium is 4%, will lead you to find that most stocks are overvalued. . The second is that the impact of using too high a premium can vary across stocks, with growth stocks being affected more negatively than mature companies. A portfolio manager who followed the recommendations of these analysts would then be over invested in mature companies and under invested in growth companies.
5. If you adjust the cash flows for risk, there is no need for a risk premium: While statement is technically correct, adjusting cash flows for risk has to go beyond reflecting the likelihood of negative scenarios in the expected cash flow. The risk adjustment to expected cash flows to make them certainty equivalent cash flows

requires us to answer exactly the same questions that we deal with when adjusting discount rates for risk.

Summary

The risk premium is a fundamental and critical component in portfolio management, corporate finance and valuation. Given its importance, it is surprising that more attention has not been paid in practical terms to estimation issues. In this paper, we began by looking at the determinants of equity risk premiums including macro economic volatility, investor risk aversion and behavioral components. We then looked at the three basic approaches used to estimate equity risk premiums – the survey approach, where investors or managers are asked to provide estimates of the equity risk premium for the future, the historical return approach, where the premium is based upon how well equities have done in the past and the implied approach, where we use future cash flows or observed bond default spreads to estimate the current equity risk premium.

The premiums we estimate can vary widely across approaches, and we considered two questions towards the end of the paper. The first is why the numbers vary across approaches and the second is how to choose the "right" number to use in analysis. For the latter question, we argued that the choice of a premium will depend upon the forecast period, whether your believe markets are efficient and whether you are required to be market neutral in your analysis.

Appendix 1: Historical Returns on Stocks, Bonds and Bills – United States

Year	*Stocks*	*T.Bills*	*T.Bonds*	*Stocks - T. Bills*	*Stocks - T.Bonds*	*Arithmetic Average: Stocks versus T. Bonds*	*Geometric average: Stocks vs T.Bonds*
1928	43.81%	3.08%	0.84%	40.73%	42.98%	42.98%	42.98%
1929	-8.30%	3.16%	4.20%	-11.46%	-12.50%	15.24%	12.33%
1930	-25.12%	4.55%	4.54%	-29.67%	-29.66%	0.27%	-3.60%
1931	-43.84%	2.31%	-2.56%	-46.15%	-41.28%	-10.12%	-15.42%
1932	-8.64%	1.07%	8.79%	-9.71%	-17.43%	-11.58%	-15.81%
1933	49.98%	0.96%	1.86%	49.02%	48.13%	-1.63%	-7.36%
1934	-1.19%	0.32%	7.96%	-1.51%	-9.15%	-2.70%	-7.61%
1935	46.74%	0.18%	4.47%	46.57%	42.27%	2.92%	-2.49%
1936	31.94%	0.17%	5.02%	31.77%	26.93%	5.59%	0.40%
1937	-35.34%	0.30%	1.38%	-35.64%	-36.72%	1.36%	-4.22%
1938	29.28%	0.08%	4.21%	29.21%	25.07%	3.51%	-1.87%
1939	-1.10%	0.04%	4.41%	-1.14%	-5.51%	2.76%	-2.17%
1940	-10.67%	0.03%	5.40%	-10.70%	-16.08%	1.31%	-3.30%
1941	-12.77%	0.08%	-2.02%	-12.85%	-10.75%	0.45%	-3.88%
1942	19.17%	0.34%	2.29%	18.84%	16.88%	1.54%	-2.61%
1943	25.06%	0.38%	2.49%	24.68%	22.57%	2.86%	-1.18%
1944	19.03%	0.38%	2.58%	18.65%	16.45%	3.66%	-0.21%
1945	35.82%	0.38%	3.80%	35.44%	32.02%	5.23%	1.35%
1946	-8.43%	0.38%	3.13%	-8.81%	-11.56%	4.35%	0.63%
1947	5.20%	0.57%	0.92%	4.63%	4.28%	4.35%	0.81%
1948	5.70%	1.02%	1.95%	4.68%	3.75%	4.32%	0.95%
1949	18.30%	1.10%	4.66%	17.20%	13.64%	4.74%	1.49%
1950	30.81%	1.17%	0.43%	29.63%	30.38%	5.86%	2.63%
1951	23.68%	1.48%	-0.30%	22.20%	23.97%	6.61%	3.46%
1952	18.15%	1.67%	2.27%	16.48%	15.88%	6.98%	3.94%
1953	-1.21%	1.89%	4.14%	-3.10%	-5.35%	6.51%	3.57%
1954	52.56%	0.96%	3.29%	51.60%	49.27%	8.09%	4.98%
1955	32.60%	1.66%	-1.34%	30.94%	33.93%	9.01%	5.93%
1956	7.44%	2.56%	-2.26%	4.88%	9.70%	9.04%	6.07%
1957	-10.46%	3.23%	6.80%	-13.69%	-17.25%	8.16%	5.23%
1958	43.72%	1.78%	-2.10%	41.94%	45.82%	9.38%	6.39%
1959	12.06%	3.26%	-2.65%	8.80%	14.70%	9.54%	6.66%
1960	0.34%	3.05%	11.64%	-2.71%	-11.30%	8.91%	6.11%
1961	26.64%	2.27%	2.06%	24.37%	24.58%	9.37%	6.62%
1962	-8.81%	2.78%	5.69%	-11.59%	-14.51%	8.69%	5.97%
1963	22.61%	3.11%	1.68%	19.50%	20.93%	9.03%	6.36%
1964	16.42%	3.51%	3.73%	12.91%	12.69%	9.13%	6.53%
1965	12.40%	3.90%	0.72%	8.50%	11.68%	9.20%	6.66%

Year	*Stocks*	*T.Bills*	*T.Bonds*	*Stocks - T. Bills*	*Stocks - T.Bonds*	*Arithmetic Average: Stocks versus T. Bonds*	*Geometric average: Stocks vs T.Bonds*
1966	-9.97%	4.84%	2.91%	-14.81%	-12.88%	8.63%	6.11%
1967	23.80%	4.33%	-1.58%	19.47%	25.38%	9.05%	6.57%
1968	10.81%	5.26%	3.27%	5.55%	7.54%	9.01%	6.60%
1969	-8.24%	6.56%	-5.01%	-14.80%	-3.23%	8.72%	6.33%
1970	3.56%	6.69%	16.75%	-3.12%	-13.19%	8.21%	5.90%
1971	14.22%	4.54%	9.79%	9.68%	4.43%	8.12%	5.87%
1972	18.76%	3.95%	2.82%	14.80%	15.94%	8.30%	6.08%
1973	-14.31%	6.73%	3.66%	-21.03%	-17.97%	7.73%	5.50%
1974	-25.90%	7.78%	1.99%	-33.68%	-27.89%	6.97%	4.64%
1975	37.00%	5.99%	3.61%	31.01%	33.39%	7.52%	5.17%
1976	23.83%	4.97%	15.98%	18.86%	7.85%	7.53%	5.22%
1977	-6.98%	5.13%	1.29%	-12.11%	-8.27%	7.21%	4.93%
1978	6.51%	6.93%	-0.78%	-0.42%	7.29%	7.21%	4.97%
1979	18.52%	9.94%	0.67%	8.58%	17.85%	7.42%	5.21%
1980	31.74%	11.22%	-2.99%	20.52%	34.72%	7.93%	5.73%
1981	-4.70%	14.30%	8.20%	-19.00%	-12.90%	7.55%	5.37%
1982	20.42%	11.01%	32.81%	9.41%	-12.40%	7.18%	5.10%
1983	22.34%	8.45%	3.20%	13.89%	19.14%	7.40%	5.34%
1984	6.15%	9.61%	13.73%	-3.47%	-7.59%	7.13%	5.12%
1985	31.24%	7.49%	25.71%	23.75%	5.52%	7.11%	5.13%
1986	18.49%	6.04%	24.28%	12.46%	-5.79%	6.89%	4.97%
1987	5.81%	5.72%	-4.96%	0.09%	10.77%	6.95%	5.07%
1988	16.54%	6.45%	8.22%	10.09%	8.31%	6.98%	5.12%
1989	31.48%	8.11%	17.69%	23.37%	13.78%	7.08%	5.24%
1990	-3.06%	7.55%	6.24%	-10.61%	-9.30%	6.82%	5.00%
1991	30.23%	5.61%	15.00%	24.62%	15.23%	6.96%	5.14%
1992	7.49%	3.41%	9.36%	4.09%	-1.87%	6.82%	5.03%
1993	9.97%	2.98%	14.21%	6.98%	-4.24%	6.65%	4.90%
1994	1.33%	3.99%	-8.04%	-2.66%	9.36%	6.69%	4.97%
1995	37.20%	5.52%	23.48%	31.68%	13.71%	6.80%	5.08%
1996	23.82%	5.02%	1.43%	18.79%	22.39%	7.02%	5.32%
1997	31.86%	5.05%	9.94%	26.81%	21.92%	7.24%	5.53%
1998	28.34%	4.73%	14.92%	23.61%	13.42%	7.32%	5.63%
1999	20.89%	4.51%	-8.25%	16.38%	29.14%	7.63%	5.96%
2000	-9.03%	5.76%	16.66%	-14.79%	-25.69%	7.17%	5.51%
2001	-11.85%	3.67%	5.57%	-15.52%	-17.42%	6.84%	5.17%
2002	-21.97%	1.66%	15.12%	-23.62%	-37.08%	6.25%	4.53%
2003	28.36%	1.03%	0.38%	27.33%	27.98%	6.54%	4.82%
2004	10.74%	1.23%	4.49%	9.52%	6.25%	6.53%	4.84%

Year	*Stocks*	*T.Bills*	*T.Bonds*	*Stocks - T. Bills*	*Stocks - T.Bonds*	*Arithmetic Average: Stocks versus T. Bonds*	*Geometric average: Stocks vs T.Bonds*
2005	4.83%	3.01%	2.87%	1.82%	1.97%	6.47%	4.80%
2006	15.61%	4.68%	1.96%	10.94%	13.65%	6.57%	4.91%
2007	5.48%	4.64%	10.21%	0.84%	-4.73%	6.42%	4.79%
2008	-36.55%	1.59%	20.10%	-38.14%	-56.65%	5.65%	3.88%
2009	25.94%	0.14%	-11.12%	25.80%	37.05%	6.03%	4.29%
2010	14.82%	0.13%	8.46%	14.69%	6.36%	6.03%	4.31%
2011	2.10%	0.03%	16.04%	2.07%	-13.94%	5.79%	4.10%
2012	15.89%	0.05%	2.97%	15.84%	12.92%	5.88%	4.20%
2013	32.15%	0.07%	-9.10%	32.08%	41.25%	6.29%	4.62%
2014	13.48%	0.05%	10.75%	13.43%	2.73%	6.11%	4.60%

Appendix 2: Sovereign Ratings by Country- January 2015

Sovereign	*Foreign Currency Rating*	*Local Currency Rating*	*Sovereign*	*Foreign Currency Rating*	*Local Currency Rating*
Abu Dhabi	Aa2	Aa2	Czech Republic	A1	A1
Albania	B1	B1	Democratic Republic of the Congo	B3	B3
Angola	Ba2	Ba2	Denmark	Aaa	Aaa
Argentina	Caa1	Caa1	Dominican Republic	B1	B1
Armenia	Ba2	Ba2	Ecuador	B3	-
Australia	Aaa	Aaa	Egypt	Caa1	Caa1
Austria	Aaa	Aaa	El Salvador	Ba3	-
Azerbaijan	Baa3	Baa3	Estonia	A1	A1
Bahamas	Baa2	Baa2	Ethiopia	B1	B1
Bahrain	Baa2	Baa2	Fiji	B1	B1
Bangladesh	Ba3	Ba3	Finland	Aaa	Aaa
Barbados	B3	B3	France	Aa1	Aa1
Belarus	B3	B3	Gabon	Ba3	Ba3
Belgium	Aa3	Aa3	Georgia	Ba3	Ba3
Belize	Caa2	Caa2	Germany	Aaa	Aaa
Bermuda	A1	A1	Ghana	B2	B2
Bolivia	Ba3	Ba3	Greece	Caa1	Caa1
Bosnia and Herzegovina	B3	B3	Guatemala	Ba1	Ba1
Botswana	A2	A2	Honduras	B3	B3
Brazil	Baa2	Baa2	Hong Kong	Aa1	Aa1
Bulgaria	Baa2	Baa2	Hungary	Ba1	Ba1
Cambodia	B2	B2	Iceland	Baa3	Baa3
Canada	Aaa	Aaa	India	Baa3	Baa3
Cayman Islands	Aa3	-	Indonesia	Baa3	Baa3
Chile	Aa3	Aa3	Ireland	Baa1	Baa1
China	Aa3	Aa3	Isle of Man	Aa1	Aa1
Colombia	Baa2	Baa2	Israel	A1	A1
Costa Rica	Ba1	Ba1	Italy	Baa2	Baa2
Côte d'Ivoire	B1	B1	Jamaica	Caa3	Caa3
Croatia	Ba1	Ba1	Japan	A1	A1
Cuba	Caa2	-	Jordan	B1	B1
Cyprus	B3	B3	Kazakhstan	Baa2	Baa2

Appendix 2: Sovereign Ratings by Country- January 2015 (Continued)

Sovereign	*Foreign Currency Rating*	*Local Currency Rating*	*Sovereign*	*Foreign Currency Rating*	*Local Currency Rating*
Kenya	B1	B1	Qatar	Aa2	Aa2
Korea	Aa3	Aa3	Republic of the Congo	Ba3	Ba3
Kuwait	Aa2	Aa2	Romania	Baa3	Baa3
Latvia	Baa1	Baa1	Russia	Baa2	Baa2
Lebanon	B2	B2	Saudi Arabia	Aa3	Aa3
Lithuania	Baa1	Baa1	Senegal	B1	B1
Luxembourg	Aaa	Aaa	Serbia	B1	B1
Macao	Aa2	Aa2	Sharjah	A3	A3
Malaysia	A3	A3	Singapore	Aaa	Aaa
Malta	A3	A3	Slovakia	A2	A2
Mauritius	Baa1	Baa1	Slovenia	Ba1	Ba1
Mexico	A3	A3	South Africa	Baa2	Baa2
Moldova	B3	B3	Spain	Baa2	Baa2
Mongolia	B2	B2	Sri Lanka	B1	-
Montenegro	Ba3	-	St. Maarten	Baa1	Baa1
Morocco	Ba1	Ba1	St. Vincent & the Grenadines	B3	B3
Mozambique	B1	B1	Suriname	Ba3	Ba3
Namibia	Baa3	Baa3	Sweden	Aaa	Aaa
Netherlands	Aaa	Aaa	Switzerland	Aaa	Aaa
New Zealand	Aaa	Aaa	Taiwan	Aa3	Aa3
Nicaragua	B3	B3	Thailand	Baa1	Baa1
Nigeria	Ba3	Ba3	Trinidad and Tobago	Baa1	Baa1
Norway	Aaa	Aaa	Tunisia	Ba3	Ba3
Oman	A1	A1	Turkey	Baa3	Baa3
Pakistan	Caa1	Caa1	Uganda	B1	B1
Panama	Baa2	-	Ukraine	Caa3	Caa3
Papua New Guinea	B1	B1	United Arab Emirates	Aa2	Aa2
Paraguay	Ba2	Ba2	UK	Aa1	Aa1
Peru	A3	A3	USA	Aaa	Aaa
Philippines	Baa2	Baa2	Uruguay	Baa2	Baa2
Poland	A2	A2	Venezuela	Caa1	Caa1
Portugal	Ba1	Ba1	Vietnam	B1	B1
			Zambia	B1	B1

Appendix 3: Country Risk Scores from the PRS Group – January 2015

Country	*PRS Composite Risk Score*	*Country*	*PRS Composite Risk Score*
Albania	66.3	Egypt	59.0
Algeria	68.3	El Salvador	66.8
Angola	65.8	Estonia	69.5
Argentina	63.8	Ethiopia	59.3
Armenia	63.0	Finland	79.0
Australia	78.5	France	70.8
Austria	79.5	Gabon	71.3
Azerbaijan	75.8	Gambia	62.8
Bahamas	75.8	Germany	84.5
Bahrain	70.5	Ghana	61.3
Bangladesh	64.0	Greece	64.3
Belarus	59.3	Guatemala	66.8
Belgium	76.0	Guinea	47.8
Bolivia	73.8	Guinea-Bissau	62.5
Botswana	79.5	Guyana	61.8
Brazil	67.5	Haiti	61.0
Brunei	87.0	Honduras	64.8
Bulgaria	69.3	Hong Kong	81.0
Burkina Faso	63.0	Hungary	72.3
Cameroon	63.5	Iceland	79.8
Canada	82.0	India	68.8
Chile	75.8	Indonesia	67.3
China, Peoples' Rep.	71.8	Iran	61.3
Colombia	68.5	Iraq	61.8
Congo, Dem. Republic	55.3	Ireland	78.5
Congo, Republic	68.8	Israel	72.3
Costa Rica	73.5	Italy	72.5
Cote d'Ivoire	62.3	Jamaica	68.5
Croatia	68.5	Japan	78.8
Cuba	65.5	Jordan	65.0
Cyprus	69.3	Kazakhstan	70.5
Czech Republic	78.3	Kenya	63.3
Denmark	81.3	Korea, D.P.R.	55.8
Dominican Republic	71.5	Korea, Republic	81.5
Ecuador	67.0	Kuwait	81.5

Appendix 3: Country Risk Scores from the PRS Group – January 2015 (Continued)

Country	*PRS Composite Risk Score*	*Country*	*PRS Composite Risk Score*
Latvia	69.0	Russia	64.3
Lebanon	58.5	Saudi Arabia	78.8
Liberia	50.0	Senegal	62.8
Libya	59.3	Serbia	63.0
Lithuania	76.0	Sierra Leone	61.5
Luxembourg	87.5	Singapore	87.0
Madagascar	63.5	Slovakia	74.3
Malawi	61.0	Slovenia	70.0
Malaysia	78.8	Somalia	37.3
Mali	60.5	South Africa	67.3
Malta	75.8	Spain	70.5
Mexico	68.8	Sri Lanka	62.3
Moldova	63.8	Sudan	50.0
Mongolia	64.3	Suriname	72.0
Morocco	67.3	Sweden	82.0
Mozambique	56.0	Switzerland	89.5
Myanmar	62.8	Syria	41.5
Namibia	75.8	Taiwan	83.0
Netherlands	81.0	Tanzania	62.3
New Zealand	83.0	Thailand	67.0
Nicaragua	64.8	Togo	60.3
Niger	55.8	Trinidad & Tobago	76.8
Nigeria	62.5	Tunisia	63.5
Norway	90.0	Turkey	61.5
Oman	81.0	Uganda	58.0
Pakistan	58.5	Ukraine	54.3
Panama	71.8	United Arab Emirates	82.8
Papua New Guinea	64.8	United Kingdom	78.8
Paraguay	69.5	United States	77.3
Peru	71.5	Uruguay	72.0
Philippines	72.3	Venezuela	54.8
Poland	75.3	Vietnam	70.0
Portugal	73.3	Yemen, Republic	59.5
Qatar	82.3	Zambia	67.0
Romania	71.5	Zimbabwe	54.5

Appendix 4: Equity Market volatility, relative to S&P 500: Total Equity Risk Premiums and Country Risk Premiums (Weekly returns from 2/13-2/15)

Country	*Std deviation in Equities (weekly)*	*Relative Volatility (to US)*	*Total Equity Risk Premium*	*Country risk premium*
Argentina	35.50%	3.27	18.78%	13.03%
Bahrain	7.59%	0.70	4.01%	-1.74%
Bangladesh	16.24%	1.49	8.59%	2.84%
Bosnia	8.99%	0.83	4.76%	-0.99%
Botswana	4.19%	0.39	2.22%	-3.53%
Brazil	22.25%	2.05	11.77%	6.02%
Bulgaria	15.33%	1.41	8.11%	2.36%
Chile	13.91%	1.28	7.36%	1.61%
China	17.82%	1.64	9.43%	3.68%
Colombia	16.00%	1.47	8.46%	2.71%
Costa Rica	8.78%	0.81	4.64%	-1.11%
Croatia	7.42%	0.68	3.93%	-1.82%
Cyprus	36.97%	3.40	19.56%	13.81%
Czech Republic	15.39%	1.42	8.14%	2.39%
Egypt	25.47%	2.34	13.47%	7.72%
Estonia	10.26%	0.94	5.43%	-0.32%
Ghana	9.09%	0.84	4.81%	-0.94%
Greece	40.49%	3.72	21.42%	15.67%
Hungary	17.21%	1.58	9.10%	3.35%
Iceland	10.89%	1.00	5.76%	0.01%
India	14.09%	1.30	7.45%	1.70%
Indonesia	16.49%	1.52	8.72%	2.97%
Ireland	16.07%	1.48	8.50%	2.75%
Israel	8.33%	0.77	4.41%	-1.34%
Italy	20.74%	1.91	10.97%	5.22%
Jamaica	10.04%	0.92	5.31%	-0.44%
Jordan	9.88%	0.91	5.23%	-0.52%
Kazakhastan	28.17%	2.59	14.90%	9.15%
Kenya	10.09%	0.93	5.34%	-0.41%
Korea	11.20%	1.03	5.92%	0.17%
Kuwait	10.47%	0.96	5.54%	-0.21%
Laos	14.18%	1.30	7.50%	1.75%
Latvia	12.11%	1.11	6.41%	0.66%
Lebanon	5.89%	0.54	3.12%	-2.63%
Lithuania	8.54%	0.79	4.52%	-1.23%
Macedonia	13.64%	1.25	7.22%	1.47%

Malaysia	8.61%	0.79	4.55%	-1.20%
Malta	6.91%	0.64	3.66%	-2.09%
Mauritius	5.42%	0.50	2.87%	-2.88%
Mexico	14.81%	1.36	7.83%	2.08%
Mongolia	20.05%	1.84	10.61%	4.86%
Montenegro	13.26%	1.22	7.01%	1.26%
Morocco	8.26%	0.76	4.37%	-1.38%
Namibia	15.33%	1.41	8.11%	2.36%
Nigeria	24.07%	2.21	12.73%	6.98%
Oman	17.68%	1.63	9.35%	3.60%
Pakistan	15.07%	1.39	7.97%	2.22%
Palestine	14.08%	1.30	7.45%	1.70%
Panama	6.18%	0.57	3.27%	-2.48%
Peru	16.15%	1.49	8.54%	2.79%
Philippines	14.69%	1.35	7.77%	2.02%
Poland	15.08%	1.39	7.98%	2.23%
Portugal	21.66%	1.99	11.46%	5.71%
Qatar	20.25%	1.86	10.71%	4.96%
Romania	12.29%	1.13	6.50%	0.75%
Russia	21.02%	1.93	11.12%	5.37%
Saudi Arabia	19.02%	1.75	10.06%	4.31%
Serbia	8.58%	0.79	4.54%	-1.21%
Singapore	9.68%	0.89	5.12%	-0.63%
Slovakia	17.07%	1.57	9.03%	3.28%
Slovenia	15.26%	1.40	8.07%	2.32%
South Africa	13.79%	1.27	7.29%	1.54%
Spain	19.38%	1.78	10.25%	4.50%
Sri Lanka	12.40%	1.14	6.56%	0.81%
Taiwan	10.97%	1.01	5.80%	0.05%
Tanzania	18.22%	1.68	9.64%	3.89%
Thailand	16.87%	1.55	8.92%	3.17%
Tunisia	8.23%	0.76	4.35%	-1.40%
Turkey	25.06%	2.31	13.26%	7.51%
UAE	32.50%	2.99	17.19%	11.44%
Ukraine	27.07%	2.49	14.32%	8.57%
US	**10.87%**	**1.00**	**5.75%**	**0.00%**
Venezuela	40.04%	3.68	21.18%	15.43%
Vietnam	16.75%	1.54	8.86%	3.11%

Appendix 5: Equity Market Volatility versus Bond Market/CDS volatility

Standard deviation in equity index (σ_{Equity}) and government bond price (σ_{Bond}) was computed, using 100 trading weeks, where available. To compute the σ_{CDS}, we first computed the standard deviation of the CDS in basis points over 100 weeks and then divided by the level of the CDS to get a coefficient of variation.

Country	σ_{Equity}	σ_{Bond}	$\sigma_{Equity}/\sigma_{Bond}$	σ_{CDS}	$\sigma_{Equity}/\sigma_{CDS}$
Argentina	35.50%		NA	2.95%	12.05
Bahrain	7.59%		NA	14.65%	0.66
Bangladesh	16.24%		NA	NA	NA
Bosnia	8.99%		NA	NA	NA
Botswana	4.19%		NA	NA	NA
Brazil	22.25%	11.97%	1.86	12.78%	1.87
Bulgaria	15.33%	17.49%	0.88	18.69%	1.01
Chile	13.91%	6.66%	2.09	32.46%	0.75
China	17.82%		NA	28.11%	0.92
Colombia	16.00%	6.67%	2.40	23.79%	0.91
Costa Rica	8.78%		NA	11.91%	0.86
Croatia	7.42%		NA	1.05%	7.07
Cyprus	36.97%		NA	16.74%	2.38
Czech Republic	15.39%	7.26%	2.12	5.19%	3.02
Egypt	25.47%		NA	1.08%	23.49
Estonia	10.26%		NA	54.97%	0.74
Ghana	9.09%		NA	NA	NA
Greece	40.49%	56.23%	0.72	12.17%	3.45
Hungary	17.21%		NA	24.13%	0.95
Iceland	10.89%	4.04%	2.70	16.14%	0.84
India	14.09%	3.49%	4.04	11.35%	1.35
Indonesia	16.49%	9.45%	1.74	18.87%	1.06
Ireland	16.07%	5.00%	3.21	7.19%	2.31
Israel	8.33%	5.90%	1.41	220.40%	2.24
Italy	20.74%	7.40%	2.80	31.74%	0.97
Jamaica	10.04%		NA	NA	NA
Jordan	9.88%		NA	NA	NA
Kazakhastan	28.17%		NA	16.96%	1.83
Kenya	10.09%		NA	NA	NA
Korea	11.20%	6.59%	1.70	49.83%	0.72
Kuwait	10.47%		NA	NA	NA
Laos	14.18%		NA	NA	NA
Latvia	12.11%		NA	20.87%	0.79
Lebanon	5.89%	4.44%	1.33	11.82%	0.62
Lithuania	8.54%		NA	21.35%	0.61
Macedonia	13.64%		NA	NA	NA
Malaysia	8.61%		NA	30.24%	0.59

Malta	6.91%		NA	NA	NA
Mauritius	5.42%		NA		
Mexico	14.81%	9.51%	1.56	21.85%	0.90
Mongolia	20.05%		NA	NA	NA
Montenegro	13.26%		NA	NA	NA
Morocco	8.26%		NA	17.27%	0.65
Namibia	15.33%		NA	NA	NA
Nigeria	24.07%		NA	NA	NA
Oman	17.68%		NA	NA	NA
Pakistan	15.07%		NA	15.93%	1.11
Palestine	14.08%		NA	NA	NA
Panama	6.18%		NA	19.13%	0.51
Peru	16.15%	8.51%	1.90	20.04%	1.01
Philippines	14.69%	30.36%	0.48	33.29%	0.77
Poland	15.08%	11.71%	1.29	30.94%	0.80
Portugal	21.66%	10.18%	2.13	36.42%	0.96
Qatar	20.25%		NA	26.85%	1.02
Romania	12.29%		NA	21.61%	0.78
Russia	21.02%	40.10%	0.52	22.87%	1.15
Saudi Arabia	19.02%		NA	36.45%	0.89
Serbia	8.58%		NA	NA	NA
Singapore	9.68%		NA	NA	NA
Slovakia	17.07%	7.91%	2.16	23.18%	0.97
Slovenia	15.26%	13.06%	1.17	8.18%	1.95
South Africa	13.79%		NA	14.78%	1.08
Spain	19.38%	7.30%	2.65	49.92%	0.89
Sri Lanka	12.40%		NA	NA	NA
Taiwan	10.97%		NA	NA	NA
Tanzania	18.22%		NA	NA	NA
Thailand	16.87%	6.49%	2.60	26.79%	0.90
Tunisia	8.23%		NA	13.41%	0.75
Turkey	25.06%	13.17%	1.90	14.83%	1.84
UAE	32.50%		NA	NA	NA
Ukraine	27.07%		NA	6.66%	4.13
US	10.87%		NA	283.38%	2.87
Venezuela	40.04%	36.25%	1.10	10.62%	3.88
Vietnam	16.75%		NA	11.81%	1.54

Appendix 6: Year-end Implied Equity Risk Premiums: 1961-2013

Year	*S&P 500*	*Earnings*[a]	*Dividends*[a]	*T.Bond Rate*	*Estimated Growth*	*Implied Premium*
1961	71.55	3.37	2.04	2.35%	2.41%	2.92%
1962	63.1	3.67	2.15	3.85%	4.05%	3.56%
1963	75.02	4.13	2.35	4.14%	4.96%	3.38%
1964	84.75	4.76	2.58	4.21%	5.13%	3.31%
1965	92.43	5.30	2.83	4.65%	5.46%	3.32%
1966	80.33	5.41	2.88	4.64%	4.19%	3.68%
1967	96.47	5.46	2.98	5.70%	5.25%	3.20%
1968	103.86	5.72	3.04	6.16%	5.32%	3.00%
1969	92.06	6.10	3.24	7.88%	7.55%	3.74%
1970	92.15	5.51	3.19	6.50%	4.78%	3.41%
1971	102.09	5.57	3.16	5.89%	4.57%	3.09%
1972	118.05	6.17	3.19	6.41%	5.21%	2.72%
1973	97.55	7.96	3.61	6.90%	8.30%	4.30%
1974	68.56	9.35	3.72	7.40%	6.42%	5.59%
1975	90.19	7.71	3.73	7.76%	5.99%	4.13%
1976	107.46	9.75	4.22	6.81%	8.19%	4.55%
1977	95.1	10.87	4.86	7.78%	9.52%	5.92%
1978	96.11	11.64	5.18	9.15%	8.48%	5.72%
1979	107.94	14.55	5.97	10.33%	11.70%	6.45%
1980	135.76	14.99	6.44	12.43%	11.01%	5.03%
1981	122.55	15.18	6.83	13.98%	11.42%	5.73%
1982	140.64	13.82	6.93	10.47%	7.96%	4.90%
1983	164.93	13.29	7.12	11.80%	9.09%	4.31%
1984	167.24	16.84	7.83	11.51%	11.02%	5.11%
1985	211.28	15.68	8.20	8.99%	6.75%	3.84%
1986	242.17	14.43	8.19	7.22%	6.96%	3.58%
1987	247.08	16.04	9.17	8.86%	8.58%	3.99%
1988	277.72	24.12	10.22	9.14%	7.67%	3.77%
1989	353.4	24.32	11.73	7.93%	7.46%	3.51%
1990	330.22	22.65	12.35	8.07%	7.19%	3.89%
1991	417.09	19.30	12.97	6.70%	7.81%	3.48%
1992	435.71	20.87	12.64	6.68%	9.83%	3.55%
1993	466.45	26.90	12.69	5.79%	8.00%	3.17%
1994	459.27	31.75	13.36	7.82%	7.17%	3.55%
1995	615.93	37.70	14.17	5.57%	6.50%	3.29%
1996	740.74	40.63	14.89	6.41%	7.92%	3.20%
1997	970.43	44.09	15.52	5.74%	8.00%	2.73%
1998	1229.23	44.27	16.20	4.65%	7.20%	2.26%

1999	1469.25	51.68	16.71	6.44%	12.50%	2.05%
2000	1320.28	56.13	16.27	5.11%	12.00%	2.87%
2001	1148.09	38.85	15.74	5.05%	10.30%	3.62%
2002	879.82	46.04	16.08	3.81%	8.00%	4.10%
2003	1111.91	54.69	17.88	4.25%	11.00%	3.69%
2004	1211.92	67.68	19.407	4.22%	8.50%	3.65%
2005	1248.29	76.45	22.38	4.39%	8.00%	4.08%
2006	1418.3	87.72	25.05	4.70%	12.50%	4.16%
2007	1468.36	82.54	27.73	4.02%	5.00%	4.37%
2008	903.25	65.39	28.05	2.21%	4.00%	6.43%
2009	1115.10	59.65	22.31	3.84%	7.20%	4.36%
2010	1257.64	83.66	23.12	3.29%	6.95%	5.20%
2011	1257.60	97.05	26.02	1.87%	7.18%	6.01%
2012	1426.19	102.47	30.44	1.76%	5.27%	5.78%
2013	1848.36	107.45	36.28	3.04%	4.28%	4.96%
2014	2058.90	114.74	38.57	2.17%	5.58%	5.78%

[a] The earnings and dividend numbers for the S&P 500 represent the estimates that would have been available at the start of each of the years and thus may not match up to the actual numbers for the year. For instance, in January 2011, the estimated earnings for the S&P 500 index included actual earnings for three quarters of 2011 and the estimated earnings for the last quarter of 2011. The actual earnings for the last quarter would not have been available until March of 2011.

CHAPTER 13

DIVIDEND DISCOUNT MODELS

In the strictest sense, the only cash flow you receive from a firm when you buy publicly traded stock is the dividend. The simplest model for valuing equity is the dividend discount model -- the value of a stock is the present value of expected dividends on it. While many analysts have turned away from the dividend discount model and viewed it as outmoded, much of the intuition that drives discounted cash flow valuation is embedded in the model. In fact, there are specific companies where the dividend discount model remains a useful took for estimating value.

This chapter explores the general model as well as specific versions of it tailored for different assumptions about future growth. It also examines issues in using the dividend discount model and the results of studies that have looked at its efficacy.

The General Model

When an investor buys stock, she generally expects to get two types of cashflows - dividends during the period she holds the stock and an expected price at the end of the holding period. Since this expected price is itself determined by future dividends, the value of a stock is the present value of dividends through infinity.

$$\text{Value per share of stock} = \sum_{t=1}^{t=\infty} \frac{E(DPS_t)}{(1+k_e)^t}$$

where,

DPS_t = Expected dividends per share

k_e = Cost of equity

The rationale for the model lies in the present value rule - the value of any asset is the present value of expected future cash flows discounted at a rate appropriate to the riskiness of the cash flows.

There are two basic inputs to the model - expected dividends and the cost on equity. To obtain the expected dividends, we make assumptions about expected future growth rates in earnings and payout ratios. The required rate of return on a stock is determined by its riskiness, measured differently in different models - the market beta in the CAPM, and the factor betas in the arbitrage and multi-factor models. The model is flexible enough to allow for time-varying discount rates, where the time variation is caused by expected changes in interest rates or risk across time.

Versions of the model

Since projections of dollar dividends cannot be made through infinity, several versions of the dividend discount model have been developed based upon different assumptions about future growth. We will begin with the simplest – a model designed to value stock in a stable-growth firm that pays out what it can afford in dividends and then look at how the model can be adapted to value companies in high growth that may be paying little or no dividends.

I. The Gordon Growth Model

The Gordon growth model can be used to value a firm that is in 'steady state' with dividends growing at a rate that can be sustained forever.

The Model

The Gordon growth model relates the value of a stock to its expected dividends in the next time period, the cost of equity and the expected growth rate in dividends.

$$\text{Value of Stock} = \frac{\text{DPS}_1}{k_e - \text{g}}$$

where,

DPS_1 = Expected Dividends one year from now (next period)

k_e= Required rate of return for equity investors

g = Growth rate in dividends forever

What is a stable growth rate?

While the Gordon growth model is a simple and powerful approach to valuing equity, its use is limited to firms that are growing at a stable rate. There are two insights worth keeping in mind when estimating a 'stable' growth rate. First, since the growth rate in the firm's dividends is expected to last forever, the firm's other measures of performance (including earnings) can also be expected to grow at the same rate. To see why, consider the consequences in the long term of a firm whose earnings grow 6% a year forever, while its dividends grow at 8%. Over time, the dividends will exceed earnings. On the other hand, if a firm's earnings grow at a faster rate than dividends in the long term, the payout ratio, in the long term, will converge towards zero, which is also not a steady state. Thus, though the model's requirement is for the expected growth rate in dividends, analysts should be able to substitute in the expected growth rate in earnings and get precisely the same result, if the firm is truly in steady state.

The second issue relates to what growth rate is reasonable as a 'stable' growth rate. As noted in Chapter 12, this growth rate has to be less than or equal to the growth rate of the economy in which the firm operates. This does not, however, imply that analysts will always agree about what this rate should be even if they agree that a firm is a stable growth firm for three reasons.

- Given the uncertainty associated with estimates of expected inflation and real growth in the economy, there can be differences in the benchmark growth rate used by different analysts, i.e., analysts with higher expectations of inflation in the long term may project a nominal growth rate in the economy that is higher.
- The growth rate of a company may not be greater than that of the economy but it can be less. Firms can becomes smaller over time relative to the economy.
- There is another instance in which an analyst may be stray from a strict limit imposed on the 'stable growth rate'. If a firm is likely to maintain a few years of 'above-stable' growth rates, an approximate value for the firm can be obtained by adding a premium to the stable growth rate, to reflect the above-average growth in the initial years. Even in this case, the flexibility that the analyst has is limited. The sensitivity of the model to growth implies that the stable growth rate cannot be more than 1% or 2% above the growth rate in the economy. If the deviation becomes larger, the analyst will be better served using a two-stage or a three-stage model to capture the 'super-normal' or 'above-average' growth and restricting the Gordon growth model to when the firm becomes truly stable.

Does a stable growth rate have to be constant over time?

The assumption that the growth rate in dividends has to be constant over time is a difficult assumption to meet, especially given the volatility of earnings. If a firm has an average growth rate that is close to a stable growth rate, the model can be used with little real effect on value. Thus, a cyclical firm that can be expected to have year-to-year swings in growth rates, but has an average growth rate that is 5%, can be valued using the Gordon growth model, without a significant loss of generality. There are two reasons for this result. First, since dividends are smoothed even when earnings are volatile, they are less likely to be affected by year-to-year changes in earnings growth. Second, the mathematical effects of using an average growth rate rather than a constant growth rate are small.

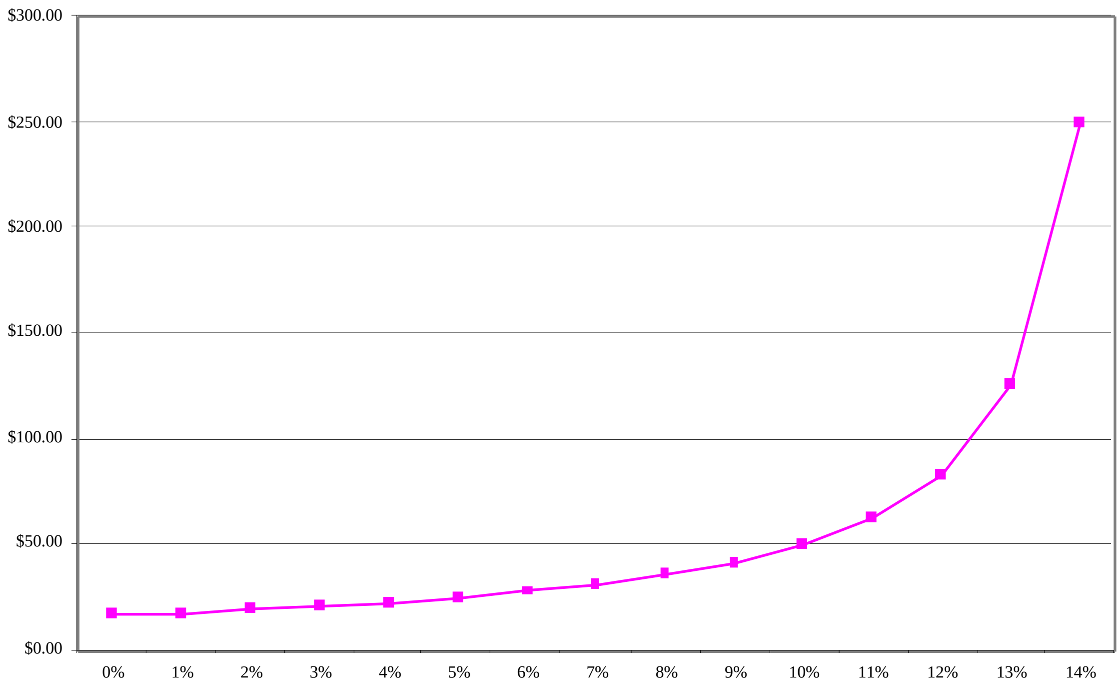
Limitations of the model

The Gordon growth model is a simple and convenient way of valuing stocks but it is extremely sensitive to the inputs for the growth rate. Used incorrectly, it can yield misleading or even absurd results, since, as the growth rate converges on the discount rate, the value goes to infinity. Consider a stock, with an expected dividend per share next period of $2.50, a cost of equity of 15%, and an expected growth rate of 5% forever. The value of this stock is:

$$\text{Value} = \frac{2.50}{0.15 - 0.05} = \$\ 25.00$$

Note, however, the sensitivity of this value to estimates of the growth rate in Figure 13.1.

Figure 13.1: Value Per Share and Expected Growth Rate



As the growth rate approaches the cost of equity, the value per share approaches infinity. If the growth rate exceeds the cost of equity, the value per share becomes negative.

This issue is tied to the question of what comprises a stable growth rate. If an analyst follows the constraints discussed in the previous chapter in estimating stable growth rates, this will never happen. In this example, for instance, an analyst who uses a 14% growth rate and obtains a $250 value would have been violating a basic rule on what comprises stable growth.

Works best for:

In summary, the Gordon growth model is best suited for firms growing at a rate comparable to or lower than the nominal growth in the economy and which have well established dividend payout policies that they intend to continue into the future. The dividend payout of the firm has to be consistent with the assumption of stability, since stable firms generally pay substantial dividends[1]. In particular, this model will under estimate the value of the stock in firms that consistently pay out less than they can afford and accumulate cash in the process.

.DDMst.xls: This spreadsheet allows you to value a stable growth firm, with stable firm characteristics (beta and retun on equity) and dividends that roughly match cash flows.

Illustration 13.1: Value a regulated firm: Consolidated Edison in May 2001

Consolidated Edison is the electric utility that supplies power to homes and businesses in New York and its environs. It is a monopoly whose prices and profits are regulated by the State of New York.

Rationale for using the model

- The firm is in stable growth; based upon size and the area that it serves. Its rates are also regulated. It is unlikely that the regulators will allow profits to grow at extraordinary rates.
- The firm is in a stable business and regulation is likely to restrict expansion into new businesses.
- The firm is in stable leverage.
- The firm pays out dividends that are roughly equal to FCFE.
    - Average Annual FCFE between 1996 and 2000 = $551 million
    - Average Annual Dividends between 1996 and 2000 = $506 million
    - Dividends as % of FCFE = 91.54%

Background Information

Earnings per share in 2000 = $3.13

Dividend Payout Ratio in 1994 = 69.97%

Dividends per share in 2000 = $2.19

Return on equity = 11.63%

Estimates

We first estimate the cost of equity, using a bottom-up levered beta for electric utilities of 0.90, a riskfree rate of 5.40% and a market risk premium of 4%.

Con Ed Beta = 0.90

Cost of Equity = 5.4% + 0.90*4% = 9%

We estimate the expected growth rate from fundamentals.

Expected growth rate = (1- Payout ratio) Return on equity

= (1-0.6997)(0.1163) = 3.49%

[1] The average payout ratio for large stable firms in the United States is about 60%.

Valuation

We now use the Gordon growth model to value the equity per share at Con Ed:

$$\text{Value of Equity} = \frac{\text{Expected dividends next year}}{\text{Cost of equity - Expected growth rate}} = \frac{(\$2.19)(1.0349)}{0.09 - 0.0349} = \$41.15$$

Con Ed was trading for $36.59 on the day of this analysis (May 14, 2001). Based upon this valuation, the stock would have been under valued.

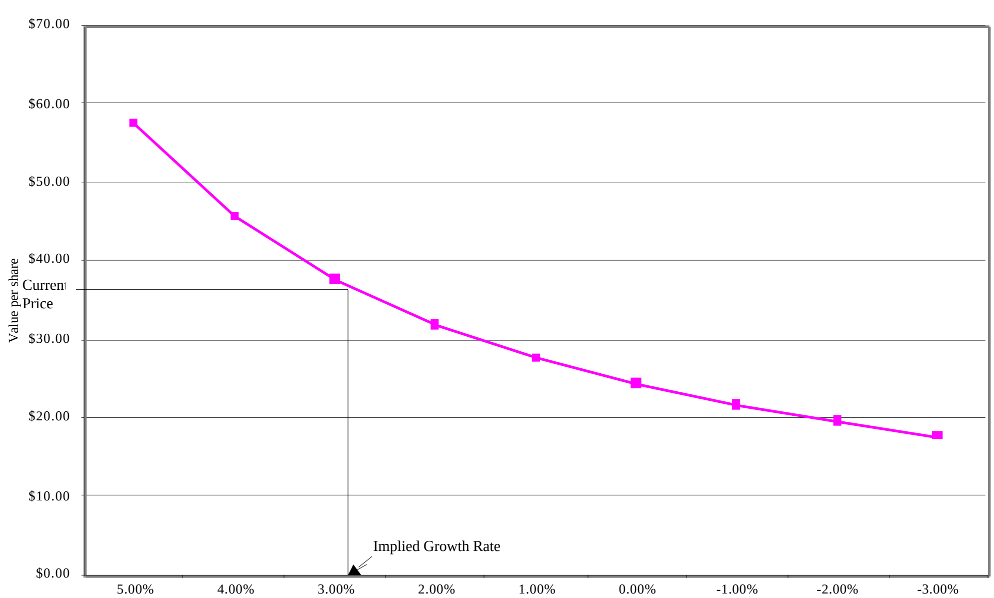
.DDMst.xlss: This spreadsheet allows you to value a stable growth firm, with stable firm characteristics (beta and return on equity) and dividends that roughly match cash flows.

Implied Growth Rate

Our value for Con Ed is different from the market price and this is likely to be the case with almost any company that you value. There are three possible explanations for this deviation. One is that you are right and the market is wrong. While this may be the correct explanation, you should probably make sure that the other two explanations do not hold – that the market is right and you are wrong or that the difference is too small to draw any conclusions. [

To examine the magnitude of the difference between the market price and your estimate of value, you can hold the other variables constant and change the growth rate in your valuation until the value converges on the price. Figure 13.2 estimates value as a function of the expected growth rate (assuming a beta of 0.90 and current dividends per share of $2.19).

Figure 13.2: Value per share versus Growth



Solving for the expected growth rate that provides the current price,

$$\$36.59 = \frac{\$2.19(1+g)}{0.09 - g}$$

The growth rate in earnings and dividends would have to be 2.84% a year to justify the stock price of \$36.59. This growth rate is called an **implied growth rate**. Since we estimate growth from fundamentals, this allows us to estimate an implied return on equity.

$$\text{Implied return on equity} = \frac{\text{Implied growth rate}}{\text{Retention ratio}} = \frac{0.0284}{0.3003} = 9.47\%$$

Illustration 13.2: Value a real estate investment trust: Vornado REIT

Real estate investment trusts were created in the early 1970s by a law that allowed these entities to invest in real estate and pass the income, tax-free, to their investors. In return for the tax benefit, however, REITs are required to return at least 95% of their earnings as dividends. Thus, they provide an interesting case study in dividend discount model valuation. Vornado Realty Trust owns and has investments in real estate in the New York area including Alexander's, the Hotel Pennsylvania and other ventures.

Rationale for using the model

Since the firm is required to pay out 95% of its earnings as dividends, the growth in earnings per share will be modest,[2] making it a good candidate for the Gordon growth model.

Background Information

In 2000, Vornado paid dividends per share of \$2.12 on earnings per share of \$2.22. The estimated payout ratio is:

$$\text{Expected payout ratio} = \frac{2.12}{2.22} = 95.50\%$$

The firm had a return on equity of 12.29%.

Estimates

We use the average beta for real estate investment trusts of 0.69, a riskfree rate of 5.4% and a risk premium of 4% to estimate a cost of equity:

Cost of equity = 5.4% + 0.69 (4%) = 8.16%

The expected growth rate is estimated from the dividend payout ratio and the return on equity:

[2] Growth in net income may be much higher, since REITs can still issue new equity for investing in new ventures.

Expected growth rate = (1- 0.955) (0.1229) = 0.55%

Valuation

Value per share = $\frac{2.12(1.0055)}{0.0816 - 0.0055} = \$28.03$

It is particularly important with REITs that we steer away from net income growth, which may be much higher. On May 14, 2001, Vornado Realty was trading at $36.57, which would make it overvalued.

II. Two-stage Dividend Discount Model

The two-stage growth model allows for two stages of growth - an initial phase where the growth rate is not a stable growth rate and a subsequent steady state where the growth rate is stable and is expected to remain so for the long term. While, in most cases, the growth rate during the initial phase is higher than the stable growth rate, the model can be adapted to value companies that are expected to post low or even negative growth rates for a few years and then revert back to stable growth.

The Model

The model is based upon two stages of growth, an extraordinary growth phase that lasts n years and a stable growth phase that lasts forever afterwards.

Extraordinary growth rate: g% each year for n yearsStable growth: g_n forever

|____|_____________________>

Value of the Stock = PV of Dividends during extraordinary phase + PV of terminal price

$$P_0 = \sum_{t=1}^{t=n} \frac{DPS_t}{(1+k_{e,hg})^t} + \frac{P_n}{(1+k_{e,hg})^n} \text{ where } P_n = \frac{DPS_{n+1}}{(k_{e,st} - g_n)}$$

where,

DPS_t = Expected dividends per share in year t

k_e = Cost of Equity (hg: High Growth period; st: Stable growth period)

P_n = Price (terminal value) at the end of year n

g = Extraordinary growth rate for the first n years

g_n = Steady state growth rate forever after year n

In the case where the extraordinary growth rate (g) and payout ratio are unchanged for the first n years, this formula can be simplified.

$$P_0 = \frac{DPS_0 * (1+g) * \left(1 - \frac{(1+g)^n}{(1+k_{e,hg})^n}\right)}{k_{e,hg} - g} + \frac{DPS_{n+1}}{(k_{e,st} - g_n)(1+k_{e,hg})^n}$$

where the inputs are as defined above.

Calculating the terminal price

The same constraint that applies to the growth rate for the Gordon Growth Rate model, i.e., that the growth rate in the firm is comparable to the nominal growth rate in the economy, applies for the terminal growth rate (g_n) in this model as well.

In addition, the payout ratio has to be consistent with the estimated growth rate. If the growth rate is expected to drop significantly after the initial growth phase, the payout ratio should be higher in the stable phase than in the growth phase. A stable firm can pay out more of its earnings in dividends than a growing firm. One way of estimating this new payout ratio is to use the fundamental growth model described in Chapter 12.

Expected Growth = Retention ratio * Return on equity

Algebraic manipulation yields the following stable period payout ratio:

$$\text{Stable Payout ratio} = \frac{\text{Stable growth rate}}{\text{Stable period return on equity}}$$

Thus, a firm with a 5% growth rate and a return on equity of 15% will have a stable period payout ratio of 33.33%.

The other characteristics of the firm in the stable period should be consistent with the assumption of stability. For instance, it is reasonable to assume that a high growth firm has a beta of 2.0, but unreasonable to assume that this beta will remain unchanged when the firm becomes stable. In fact, the rule of thumb that we developed in the last chapter – that stable period betas should be between 0.8 and 1.2 – is worth repeating here. Similarly, the return on equity, which can be high during the initial growth phase, should come down to levels commensurate with a stable firm in the stable growth phase. What is a reasonable stable period return on equity? The industry average return on equity and the firm's own stable period cost of equity provide useful information to make this judgment.

Limitations of the model

There are three problems with the two-stage dividend discount model – the first two would apply to any two-stage model and the third is specific to the dividend discount model.

- The first practical problem is in defining the length of the extraordinary growth period. Since the growth rate is expected to decline to a stable level after this period, the value of an investment will increase as this period is made longer. While we did develop criteria that might be useful in making this judgment in Chapter 12, it is difficult in practice to convert these qualitative considerations into a specific time period.

- The second problem with this model lies in the assumption that the growth rate is high during the initial period and is transformed overnight to a lower stable rate at the end of the period. While these sudden transformations in growth can happen, it is much more realistic to assume that the shift from high growth to stable growth happens gradually over time.
- The focus on dividends in this model can lead to skewed estimates of value for firms that are not paying out what they can afford in dividends. In particular, we will under estimate the value of firms that accumulate cash and pay out too little in dividends.

Works best for:

Since the two-stage dividend discount model is based upon two clearly delineated growth stages, high growth and stable growth, it is best suited for firms which are in high growth and expect to maintain that growth rate for a specific time period, after which the sources of the high growth are expected to disappear. One scenario, for instance, where this may apply is when a company has patent rights to a very profitable product for the next few years and is expected to enjoy super-normal growth during this period. Once the patent expires, it is expected to settle back into stable growth. Another scenario where it may be reasonable to make this assumption about growth is when a firm is in an industry which is enjoying super-normal growth because there are significant barriers to entry (either legal or as a consequence of infra-structure requirements), which can be expected to keep new entrants out for several years.

The assumption that the growth rate drops precipitously from its level in the initial phase to a stable rate also implies that this model is more appropriate for firms with modest growth rates in the initial phase. For instance, it is more reasonable to assume that a firm growing at 12% in the high growth period will see its growth rate drops to 6% afterwards than it is for a firm growing at 40% in the high growth period.

Finally, the model works best for firms that maintain a policy of paying out most of residual cash flows – i.e, cash flows left over after debt payments and reinvestment needs have been met – as dividends.

Illustration 13.3: Valuing a firm with the two-stage dividend discount model: Procter & Gamble

Procter & Gamble (P&G) manufactures and markets consumer products all over the world. Some of its best known brand names include Pampers diapers, Tide detergent, Crest toothpaste and Vicks cough/cold medicines.

A Rationale for using the Model

- *Why two-stage?* While P&G is a firm with strong brand names and an impressive track record on growth, it faces two problems. The first is the saturation of the domestic U.S. market, which represents about half of P&G's revenues. The second is the increased competition from generics across all of its product lines. We will assume that the firm will continue to grow but restrict the growth period to 5 years.
- *Why dividends?* P&G has a reputation for paying high dividends and it has not accumulated large amounts of cash over the last decade.

Background Information

- Earnings per share in 2000 = \$3.00
- Dividends per share in 2000 = \$1.37
- Payout ratio in 2000 = $\frac{1.37}{3.00} = 45.67\%$
- Return on Equity in 2000 = 29.37%

Estimates

We will first estimate the cost of equity for P&G, based upon a bottom-up beta of 0.85 (estimated using the unlevered beta for consumer product firms and P&G's debt to equity ratio), a riskfree rate of 5.4% and a risk premium of 4%.

Cost of equity = 5.4% + 0.85 (4%) = 8.8%

To estimate the expected growth in earnings per share over the five-year high growth period, we use the retention ratio in the most recent financial year (2000) but lower the return on equity to 25% from the current value.

Expected growth rate = Retention ratio * Return on Equity

$$= (1 - 0.4567)(0.25) = 13.58\%$$

In stable growth, we will estimate that the beta for the stock will rise to 1, leading to a cost of equity of 9.40%.

Cost of equity in stable growth = 5.4% + 1 (4%) = 9.40%

The expected growth rate will be assumed to be equal to the growth rate of the economy (5%) and the return on equity will drop to 15%, which is lower than the current industry average (17.4%) but higher than the cost of equity estimated above. The retention ratio in stable growth during the stable growth period is calculated.

Retention ratio in stable growth = $\frac{g}{ROE} = \frac{5\%}{15\%} = 33.33\%$

The payout ratio in stable growth is therefore 66.67%.

Estimating the value:

The first component of value is the present value of the expected dividends during the high growth period. Based upon the current earnings ($3.00), the expected growth rate (13.58%) and the expected dividend payout ratio (45.67%), the expected dividends can be computed for each year in the high growth period.

Table 13.1: Expected Dividends per share: P&G

Year	*EPS*	*DPS*	*Present Value*
1	$3.41	$1.56	$1.43
2	$3.87	$1.77	$1.49
3	$4.40	$2.01	$1.56
4	$4.99	$2.28	$1.63
5	$5.67	$2.59	$1.70
Sum			$7.81

The present value is computed using the cost of equity of 8.8% for the high growth period.

Cumulative Present Value of Dividends during high growth (@8.8%) = $7.81

The present value of the dividends can also be computed in short hand using the following computation:

$$\text{PV of Dividends} = \frac{\$1.37(1.1358)\left(1 - \frac{(1.1358)^5}{(1.088)^5}\right)}{0.088 - 0.1358} = \$7.81$$

The price (terminal value) at the end of the high growth phase (end of year 5) can be estimated using the constant growth model.

$$\text{Terminal price} = \frac{\text{Expected Dividends per share}_{n+1}}{k_{e,st} - g_n}$$

Expected Earnings per share$_6$ = 3.00 *1.1358^5*1.05 = $5.96

Expected Dividends per share$_6$ = EPS$_6$*Stable period payout ratio

= $5.96 * 0.6667 = $3.97

$$\text{Terminal price} = \frac{\text{Dividends}_6}{k_{e,st} - g} = \frac{\$3.97}{0.094 - 0.05} = \$90.23$$

The present value of the terminal price –is:

$$\text{PV of Terminal Price} = \frac{\$90.23}{(1.088)^5} = \$59.18$$

The cumulated present value of dividends and the terminal price can then be calculated.

$$P_0 = \frac{\$1.37(1.1358)\left(1 - \frac{(1.1358)^5}{(1.088)^5}\right)}{0.088 - 0.1358} + \frac{\$90.23}{(1.088)^5} = \$7.81 + \$59.18 = \$66.99$$

P&G was trading at $63.90 at the time of this analysis on May 14, 2001.

.*DDM2st.xlss*: This spreadsheet allows you to value a firm with a temporary period of high earnings followed by stable growth.

A Trouble Shooting Guide: What is wrong with this valuation? DDM 2 St

If this is your 'problem'	*this may be the s*
· If you get a extremely low value from the 2-stage DDM, the likely culprits are	
- the stable period payout ratio is too low for a stable firm (< 40%)	If using fundame
	If entering direct
- the beta in the stable period is too high for a stable firm	Use a beta close
- the use of the two-stage model when the three-stage model is more appropriate	Use a three-stag
· If you get an extremely high value,	
- the growth rate in the stable growth period is too high for stable firm	Use a growth rat

Modifying the model to include stock buybacks

In recent years, firms in the United States have increasingly turned to stock buybacks as a way of returning cash to stockholders. Figure 13.3 presents the cumulative amounts paid out by firms in the form of dividends and stock buybacks from 1960 to 1998.

Figure 13.3: Stock Buybacks and Dividends: Aggregate for US Firms - 1989-98

The trend towards stock buybacks is very strong, especially in the 1990s.

What are the implications for the dividend discount model? Focusing strictly on dividends paid as the only cash returned to stockholders exposes us to the risk that we might be missing significant cash returned to stockholders in the form of stock buybacks. The simplest way to incorporate stock buybacks into a dividend discount model is to add them on to the dividends and compute a modified payout ratio:

$$\text{Modified dividend payout ratio} = \frac{\text{Dividends} + \text{Stock Buybacks}}{\text{Net Income}}$$

While this adjustment is straightforward, the resulting ratio for any one year can be skewed by the fact that stock buybacks, unlike dividends, are not smoothed out. In other words, a firm may buy back $ 3billion in stock in one year and not buy back stock for the next 3 years. Consequently, a much better estimate of the modified payout ratio can be obtained by looking at the average value over a four or five year period. In addition, firms may sometimes buy back stock as a way of increasing financial leverage. We could adjust for this by netting out new debt issued from the calculation above:

$$\text{Modified dividend payout} = \frac{\text{Dividends} + \text{Stock Buybacks} - \text{Long Term Debt issues}}{\text{Net Income}}$$

Adjusting the payout ratio to include stock buybacks will have ripple effects on the estimated growth and the terminal value. In particular, the modified growth rate in earnings per share can be written as:

Modified growth rate = (1 – Modified payout ratio) * Return on equity

Even the return on equity can be affected by stock buybacks. Since the book value of equity is reduced by the market value of equity bought back, a firm that buys backs stock can reduce its book equity (and increase its return on equity) dramatically. If we use this return on equity as a measure of the marginal return on equity (on new investments), we will overstate the value of a firm. Adding back stock buybacks in recent year to the book equity and re-estimating the return on equity can sometimes yield a more reasonable estimate of the return on equity on investments.

Illustration 13.4: Valuing a firm with modified dividend discount mode: Procter & Gamble

Consider our earlier valuation of Procter and Gamble where we used the current dividends as the basis for our projections. Note that over the last four years, P&G has had significant stock buybacks each period. Table 13.2 summarizes the dividends and buybacks over the period.

Table 13.2: Dividends and Stock Buybacks: P&G

	1997	*1998*	*1999*	*2000*	*Total*
Net Income	3415	3780	3763	3542	14500
Dividends	1329	1462	1626	1796	6213
Buybacks	2152	391	1881	-1021	3403
Dividends+Buybacks	3481	1853	3507	775	9616
Payout ratio	38.92%	38.68%	43.21%	50.71%	42.85%
Modified payout ratio	101.93%	49.02%	93.20%	21.88%	66.32%
Buybacks	1652	1929	2533	1766	
Net LT Debt issued	-500	1538	652	2787	
Buybacks net of debt	2152	391	1881	-1021	

Over the five-year period, P&G had significant buybacks but it also increased its leverage dramatically in the last three years. Summing up the total cash returned to stockholders over the last 4 years, we arrive at a modified payout ratio of 66.32%. If we substitute this payout ratio into the valuation in Illustration 13.3, the expected growth rate over the next 5 years drops to 8.42%:

Expected growth rate = (1- Modified payout ratio) ROE = (1-0.6632)(0.25) = 8.42%

We will still assume a five year high growth period and that the parameters in stable growth remain unchanged. The value per share can be estimated.

$$P_0 = \frac{\$3.00(0.6632)(1.0842)\left(1 - \frac{(1.0842)^5}{(1.0880)^5}\right)}{0.0880 - 0.0842} + \frac{\$71.50}{(1.0880)^5} = \$56.75$$

Note that the drop in growth rate in earnings during the high growth period reduces earnings in the terminal year, and the terminal value per share drops to $71.50.

This value is lower than that obtained in Illustration 13.3 and it reflects our expectation that P&G does not have as many new profitable new investments (earning a return on equity of 25%).

Valuing an entire market using the dividend discount model

All our examples of the dividend discount model so far have involved individual companies, but there is no reason why we cannot apply the same model to value a sector or even the entire market. The market price of the stock would be replaced by the cumulative market value of all of the stocks in the sector or market. The expected dividends would be the cumulated dividends of all these stocks and could be expanded to include stock buybacks by all firms. The expected growth rate would be the growth rate in cumulated earnings of the index. There would be no need for a beta or betas, since you are looking at the entire market (which should have a beta of 1) and you could add the risk premium (or premiums) to the riskfree rate to estimate a cost of equity. You could use a two-stage model, where this growth rate is greater than the growth rate of the economy, but you should be cautious about setting the growth rate too high or the growth period too long because it will be difficult for cumulated earnings growth of all firms in an economy to run ahead of the growth rate in the economy for extended periods.

Consider a simple example. Assume that you have an index trading at 700 and that the average dividend yield of stocks in the index is 5%. Earnings and dividends can be expected to grow at 4% a year forever and the riskless rate is 5.4%. If you use a market risk premium of 4%, the value of the index can be estimated.

Cost of equity = Riskless rate + Risk premium = 5.4% + 4% = 9.4%

Expected dividends next year = (Dividend yield * Value of the index)(1+ expected growth rate) = (0.05*700) (1.04) = 36.4

$$\text{Value of the index} = \frac{\text{Expected dividends next year}}{\text{Cost of equity - Expected growth rate}} = \frac{36.4}{0.094 - 0.04} = 674$$

At its existing level of 700, the market is slightly over priced.

Illustration 13.5: Valuing the S&P 500 using a dividend discount model: January 1, 2001

On January 1, 2001, the S&P 500 index was trading at 1320. The dividend yield on the index was only 1.43%, but including stock buybacks increases the modified dividend yield to 2.50%. Analysts were estimating that the earnings of the stocks in the index would increase 7.5% a year for the next 5 years. Beyond year 5, the expected growth rate is expected to be 5%, the nominal growth rate in the economy. The treasury bond rate was 5.1% and we will use a market risk premium of 4%, leading to a cost of equity of 9.1%:

Cost of equity = 5.1% + 4% = 9.1%

The expected dividends (and stock buybacks) on the index for the next 5 years can be estimated from the current dividends and expected growth of 7.50%.

Current dividends = 2.50% of 1320 = 33.00

	1	*2*	*3*	*4*	*5*
Expected Dividends =	$35.48	$38.14	$41.00	$44.07	$47.38
Present Value =	$32.52	$32.04	$31.57	$31.11	$30.65

The present value is computed by discounting back the dividends at 9.1%. To estimate the terminal value, we estimate dividends in year 6 on the index:

Expected dividends in year 6 = $47.38 (1.05) = $49.74

$$\text{Terminal value of the index} = \frac{\text{Expected Dividends}_6}{\text{r - g}} = \frac{\$49.74}{0.091 - 0.05} = \$1213$$

$$\text{Present value of Terminal value} = \frac{\$1213}{1.091^5} = \$785$$

The value of the index can now be computed:

Value of index = Present value of dividends during high growth + Present value of terminal value = $32.52+32.04+31.57+$31.11+ $30.65+ $785 = $943

Based upon this, we would have concluded that the index was over valued at 1320.

The Value of Growth

Investors pay a price premium when they acquire companies with high growth potential. This premium takes the form of higher price-earnings or price-book value ratios. While no one will contest the proposition that growth is valuable, it is possible to pay too much for growth. In fact, empirical studies that show low price-earnings ratio stocks earning return premiums over high price-earnings ratio stocks in the long term supports the notion that investors overpay for growth. This section uses the two-stage dividend discount model to examine the value of growth and it provides a benchmark that can be used to compare the actual prices paid for growth.

Estimating the value of growth

The value of the equity in any firm can be written in terms of three components:

$$P_0 = \left[\frac{DPS_0 * (1+g) * \left(1 - \frac{(1+g)^n}{(1+k_{e,hg})^n}\right)}{k_{e,hg} - g} + \frac{DPS_{n+1}}{(k_{e,st} - g_n)(1+k_{e,hg})^n} - \frac{DPS_1}{(k_{e,st} - g_n)} \right]$$

Extraordinary Growth

$$+ \left[\frac{DPS_1}{(k_{e,st} - g_n)} - \frac{DPS_0}{k_{e,st}} \right] + \frac{DPS_0}{k_{e,st}}$$

Stable Growth | Assets in place

where

DPS_t = Expected dividends per share in year t

k_e = Required rate of return

P_n = Price at the end of year n

g = Growth rate during high growth stage

g_n = Growth rate forever after year n

Value of extraordinary growth = Value of the firm with extraordinary growth in first n years - Value of the firm as a stable growth firm[3]

Value of stable growth = Value of the firm as a stable growth firm - Value of firm with no growth

[3] The payout ratio used to calculate the value of the firm as a stable firm can be either the current payout ratio, if it is reasonable, or the new payout ratio calculated using the fundamental growth formula.

Assets in place = Value of firm with no growth

In making these estimates, though, we have to remain consistent. For instance, to value assets in place, you would have to assume that the entire earnings could be paid out in dividends, while the payout ratio used to value stable growth should be a stable period payout ratio.

Illustration 13.6: The Value of Growth: P&G in May 2001

In illustration 13.3, we valued P&G using a 2-stage dividend discount model at $66.99. We first value the assets in place using current earnings ($3.00) and assume that all earnings are paid out as dividends. We also use the stable growth cost of equity as the discount rates.

$$\text{Value of the assets in place} = \frac{\text{Current EPS}}{k_{e,st}} = \frac{\$3}{0.094} = \$31.91$$

To estimate the value of stable growth, we assume that the expected growth rate will be 5% and that the payout ratio is the stable period payout ratio of 66.67%:

$$\text{Value of stable growth} \quad \frac{(\text{Current EPS})(\text{Stable Payout Ratio})(1+g_n)}{k_{e,st}-g_n} - \$31.91$$

$$= \frac{(\$3.00)(0.6667)(1.05)}{0.094-0.05} - \$31.91 = \$15.81$$

Value of extraordinary growth = $66.99 - $31.91 - $15.81 = $19.26

The Determinants of the Value of Growth

1. *Growth rate during extraordinary period:* The higher the growth rate in the extraordinary period, the higher the estimated value of growth will be. If the growth rate in the extraordinary growth period had been raised to 20% for the Procter & Gamble valuation, the value of extraordinary growth would have increased from $19.26 to $39.45. Conversely, the value of high growth companies can drop precipitously if the expected growth rate is reduced, either because of disappointing earnings news from the firm or as a consequence of external events.
2. *Length of the extraordinary growth period:* The longer the extraordinary growth period, the greater the value of growth will be. At an intuitive level, this is fairly simple to illustrate. The value of $19.26 obtained for extraordinary growth is predicated on the assumption that high growth will last for five years. If this is revised to last ten years, the value of extraordinary growth will increase to $43.15.
3. *Profitability of projects*: The profitability of projects determines both the growth rate in the initial phase and the terminal value. As projects become more

profitable, they increase both growth rates and growth period, and the resulting value from extraordinary growth will be greater.

4. *Riskiness of the firm/equity* The riskiness of a firm determines the discount rate at which cashflows in the initial phase are discounted. Since the discount rate increases as risk increases, the present value of the extraordinary growth will decrease.

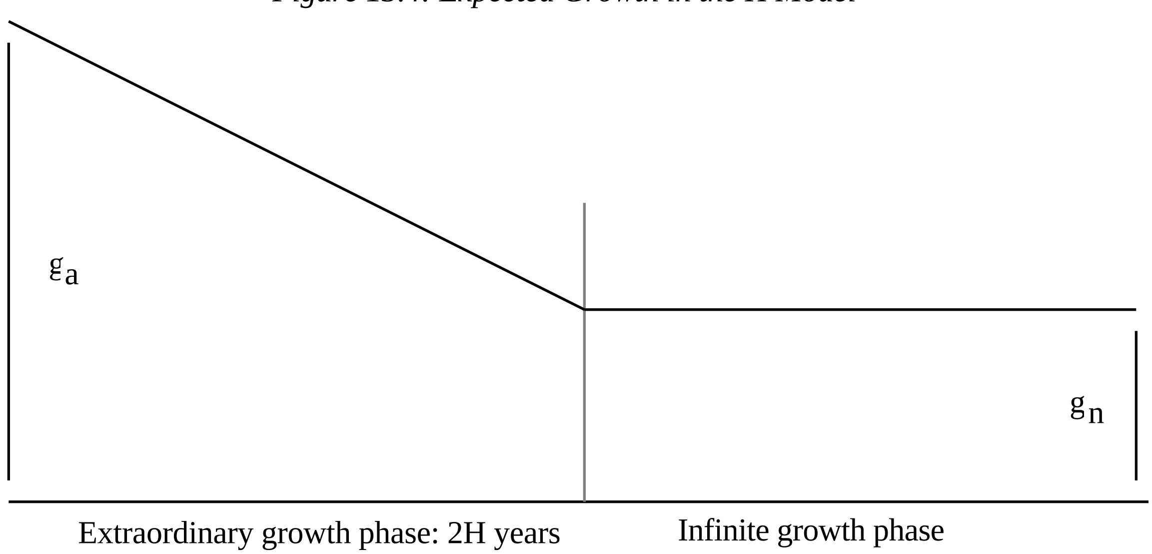
III. The H Model for valuing Growth

The H model is a two-stage model for growth, but unlike the classical two-stage model, the growth rate in the initial growth phase is not constant but declines linearly over time to reach the stable growth rate in steady stage. This model was presented in Fuller and Hsia (1984).

The Model

The model is based upon the assumption that the earnings growth rate starts at a high initial rate (g_a) and declines linearly over the extraordinary growth period (which is assumed to last 2H periods) to a stable growth rate (g_n). It also assumes that the dividend payout and cost of equity are constant over time and are not affected by the shifting growth rates. Figure 13.4 graphs the expected growth over time in the H Model.

Figure 13.4: Expected Growth in the H Model



The value of expected dividends in the H Model can be written as:

$$P_0 = \frac{DPS_0 * (1+g_n)}{(k_e - g_n)} + \frac{DPS_0 * H * (g_a - g_n)}{(k_e - g_n)}$$

Stable growth Extraordinary growth

where,

P_0 = Value of the firm now per share,

DPS_t = DPS in year t

k_e= Cost of equity

g_a = Growth rate initially

g_n = Growth rate at end of 2H years, applies forever afterwards

Limitations

This model avoids the problems associated with the growth rate dropping precipitously from the high growth to the stable growth phase, but it does so at a cost. First, the decline in the growth rate is expected to follow the strict structure laid out in the model -- it drops in linear increments each year based upon the initial growth rate, the stable growth rate and the length of the extraordinary growth period. While small deviations from this assumption do not affect the value significantly, large deviations can cause problems. Second, the assumption that the payout ratio is constant through both phases of growth exposes the analyst to an inconsistency -- as growth rates decline the payout ratio usually increases.

Works best for:

The allowance for a gradual decrease in growth rates over time may make this a useful model for firms which are growing rapidly right now, but where the growth is expected to decline gradually over time as the firms get larger and the differential advantage they have over their competitors declines. The assumption that the payout ratio is constant, however, makes this an inappropriate model to use for any firm that has low or no dividends currently. Thus, the model, by requiring a combination of high growth and high payout, may be quite limited[4] in its applicability.

Illustration 13.7: Valuing with the H model: Alcatel

Alcatel is a French telecommunications firm, paid dividends per share of 0.72 Ffr on earnings per share of 1.25 Ffr in 2000. The firm's earnings per share had grown at 12% over the prior 5 years but the growth rate is expected to decline linearly over the next 10 years to 5%, while the payout ratio remains unchanged. The beta for the stock is 0.8, the riskfree rate is 5.1% and the market risk premium is 4%.

[4] Proponents of the model would argue that using a steady state payout ratio for firms which pay little or no dividends is likely to cause only small errors in the valuation.

Cost of equity = 5.1% + 0.8*4% = 8.30%

The stock can be valued using the H model:

$$\text{Value of stable growth} = \frac{(0.72)(1.05)}{0.083 - 0.05} = \$22.91$$

$$\text{Value of extraordinary growth} = \frac{(0.72)(10/2)(0.12 - 0.05)}{0.083 - 0.05} = 7.64$$

Value of stock = 22.91 + 7.64 = 30.55

The stock was trading at 33.40 Ffr in May 2001.

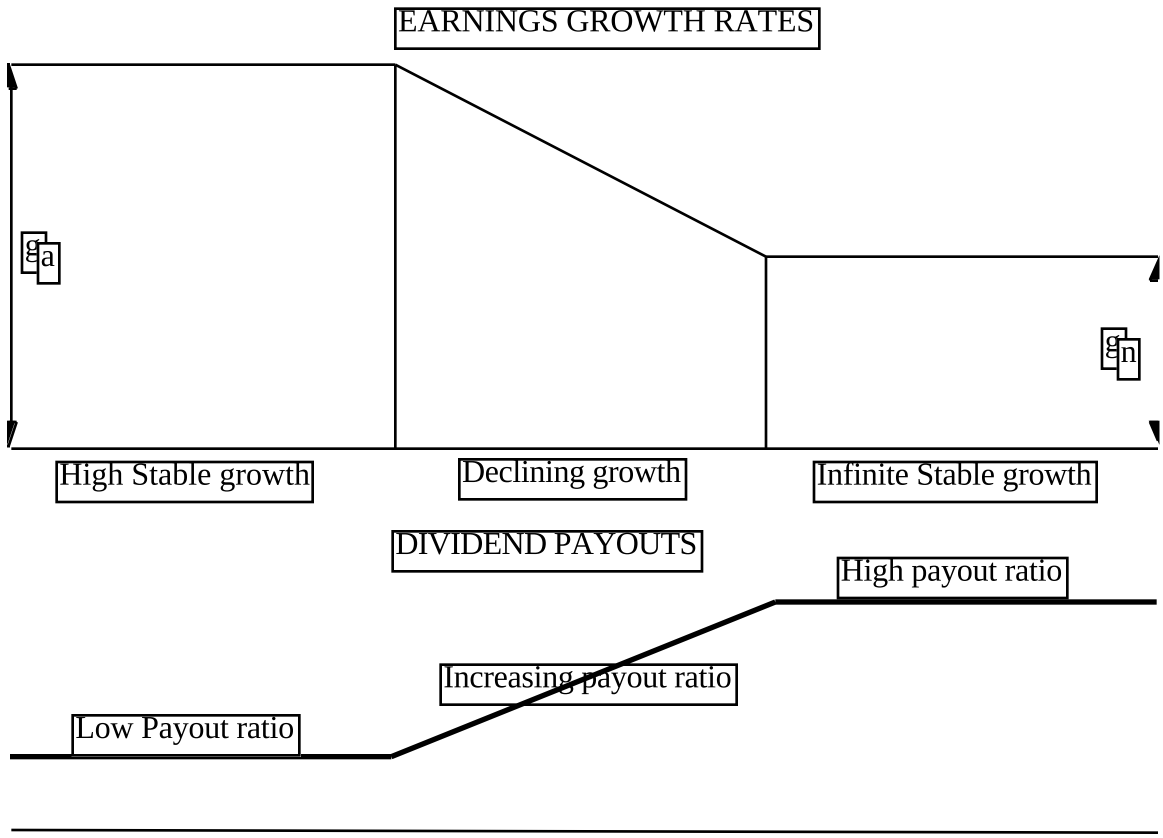
IV. Three-stage Dividend Discount Model

The three-stage dividend discount model combines the features of the two-stage model and the H-model. It allows for an initial period of high growth, a transitional period where growth declines and a final stable growth phase. It is the most general of the models because it does not impose any restrictions on the payout ratio.

The Model

This model assumes an initial period of stable high growth, a second period of declining growth and a third period of stable low growth that lasts forever. Figure 13.5 graphs the expected growth over the three time periods.

Figure 13.5: Expected Growth in the Three-Stage DDM



The value of the stock is then the present value of expected dividends during the high growth and the transitional periods and of the terminal price at the start of the final stable growth phase.

$$P_0 = \sum_{t=1}^{t=n1} \frac{EPS_0 * (1+g_a)^t * \Pi_a}{(1+k_{e,hg})^t} + \sum_{t=n1+1}^{t=n2} \frac{DPS_t}{(1+k_{e,t})^t} + \frac{EPS_{n2} * (1+g_n) * \Pi_n}{(k_{e,st} - g_n)(1+r)^n}$$

High growth phase Transition Stable growth phase

where,

EPS_t = Earnings per share in year t

DPS_t = Dividends per share in year t

g_a = Growth rate in high growth phase (lasts n1 periods)

g_n = Growth rate in stable phase

Π_a = Payout ratio in high growth phase

Π_n = Payout ratio in stable growth phase

k_e= Cost of equity in high growth (hg), transition (t) and stable growth (st)

Assumptions

This model removes many of the constraints imposed by other versions of the dividend discount model. In return, however, it requires a much larger number of inputs - year-specific payout ratios, growth rates and betas. For firms where there is substantial noise in the estimation process, the errors in these inputs can overwhelm any benefits that accrue from the additional flexibility in the model.

Works best for:

This model's flexibility makes it a useful model for any firm, which in addition to changing growth over time is expected to change on other dimensions as well - in particular, payout policies and risk. It is best suited for firms which are growing at an extraordinary rate now and are expected to maintain this rate for an initial period, after which the differential advantage of the firm is expected to deplete leading to gradual declines in the growth rate to a stable growth rate. Practically speaking, this may be the more appropriate model to use for a firm whose earnings are growing at very high rates[5], are expected to continue growing at those rates for an initial period, but are expected to start declining gradually towards a stable rate as the firm become larger and loses its competitive advantages.

Illustration 13.8: Valuing with the Three-stage DDM model: Coca Cola

Coca Cola, the owner of the most valuable brand name in the world according to Interbrand, was able to increase its market value ten-fold in the 1980s and 1990s. While growth has leveled off in the last few years, the firm is still expanding both into other products and other markets.

A Rationale for using the Three-Stage Dividend Discount Model

- *Why three-stage?* Coca Cola is still in high growth, but its size and dominant market share will cause growth to slide in the second phase of the high growth period. The high growth period is expected to last 5 years and the transition period is expected to last an additional 5 years.
- *Why dividends?* The firm has had a track record of paying out large dividends to its stockholders, and these dividends tend to mirror free cash flows to equity.
- The financial leverage is stable.

Background Information

- Current Earnings / Dividends
    - Earnings per share in 2000 = $1.56

[5] The definition of a 'very high' growth rate is largely subjective. As a rule of thumb, growth rates over 25% would qualify as very high when the stable growth rate is 6-8%.

- Dividends per share in 2000 = $0.69
- Payout ratio in 2000 = 44.23%
- Return on Equity = 23.37%

Estimate

a. Cost of Equity

We will begin by estimating the cost of equity during the high growth phase, expected. We use a bottom-up levered beta of 0.80 and a riskfree rate of 5.4%. We use a risk premium of 5.6%, significantly higher than the mature market premium of 4%, which we have used in the valuation so far, to reflect Coca Cola's exposure in Latin America, Eastern Europe and Asia. The cost of equity can then be estimated for the high growth period.

$\text{Cost of equity}_{\text{high growth}} = 5.4\% + 0.8\ (5.6\%) = 9.88\%$

In stable growth, we assume that the beta will remain 0.80, but reduce the risk premium to 5% to reflect the expected maturing of many emerging markets.

$\text{Cost of equity}_{\text{stable growth}} = 5.4\% + 0.8\ (5.0\%) = 9.40\%$

During the transition period, the cost of equity will linearly decline from 9.88% in year 5 to 9.40% in year 10.

b. Expected Growth and Payout Ratios

The expected growth rate during the high growth phase is estimated using the current return on equity of 23.37% and payout ratio of 44.23%.

Expected growth rate = Retention ratio * Return on equity = (1-0.4423)(0.2337) = 13.03%

During the transition phase, the expected growth rate declines linearly from 13.03% to a stable growth rate of 5.5%. To estimate the payout ratio in stable growth, we assume a return on equity of 20% for the firm:

$$\text{Stable period payout ratio} = 1 - \frac{g}{ROE} = 1 - \frac{5.5\%}{20\%} = 72.5\%$$

During the transition phase, the payout ratio adjusts upwards from 44.23% to 72.5% in linear increments.

Estimating the Value

These inputs are used to estimate expected earnings per share, dividends per share and costs of equity for the high growth, transition and stable periods. The present values are also shown in the last column table 13.3.

Table 13.3: Expected EPS, DPS and Present Value: Coca Cola

Year	*Expected Growth*	*EPS*	*Payout ratio*	*DPS*	*Cost of Equity*	*Present Value*

High Growth Stage						
1	13.03%	$1.76	44.23%	$0.78	9.88%	$0.71
2	13.03%	$1.99	44.23%	$0.88	9.88%	$0.73
3	13.03%	$2.25	44.23%	$1.00	9.88%	$0.75
4	13.03%	$2.55	44.23%	$1.13	9.88%	$0.77
5	13.03%	$2.88	44.23%	$1.27	9.88%	$0.79
Transition Stage						
6	11.52%	$3.21	49.88%	$1.60	9.78%	$0.91
7	10.02%	$3.53	55.54%	$1.96	9.69%	$1.02
8	8.51%	$3.83	61.19%	$2.34	9.59%	$1.11
9	7.01%	$4.10	66.85%	$2.74	9.50%	$1.18
10	5.50%	$4.33	72.50%	$3.14	9.40%	$1.24

(Note: Since the costs of equity change each year, the present value has to be calculated using the cumulated cost of equity. Thus, in year 7, the present value of dividends is:

$$\text{PV of year 7 dividend} = \frac{\$1.96}{(1.0988)^5 (1.0978)(1.0969)} = \$1.02$$

The terminal price at the end of year 10 can be calculated based upon the earnings per share in year 11, the stable growth rate of 5%, a cost of equity of 9.40% and the payout ratio of 72.5% -

$$\text{Terminal price} = \frac{\$4.33(1.055)(0.725)}{0.094 - 0.055} = \$84.83$$

The components of value are as follows:

Present Value of dividends in high growth phase:$ 3.76

Present Value of dividends in transition phase:$ 5.46

Present Value of terminal price at end of transition:$ 33.50

Value of Coca Cola Stock :$ 42.72

Coca Cola was trading at $46.29 in May 21, 2001.

.*DDM3st.xlss*: This spreadsheet allows you to value a firm with a period of high growth followed by a transition period where growth declines to a stable growth rate.

What is wrong with this model? (3 stage DDM)	
If this is your problem	*this may*
· If you are getting too low a value from this model,	
- the stable period payout ratio is too low for a stable firm (< 40%)	If using fundame
	If entering direct
- the beta in the stable period is too high for a stable firm	Use a beta close
· If you get an extremely high value,	
- the growth rate in the stable growth period is too high for stable firm	Use a growth rat
- the period of growth (high + transition) is too high	Use shorter high

Issues in using the Dividend Discount Model

The dividend discount model's primary attraction is its simplicity and its intuitive logic. There are many analysts, however, who view its results with suspicion because of limitations that they perceive it to possess. The model, they claim, is not really useful in valuation, except for a limited number of stable, high-dividend paying stocks. This section examines some of the areas where the dividend discount model is perceived to fall short.

(a) Valuing non-dividend paying or low dividend paying stocks

The conventional wisdom is that the dividend discount model cannot be used to value a stock that pays low or no dividends. It is wrong. If the dividend payout ratio is adjusted to reflect changes in the expected growth rate, a reasonable value can be obtained even for non-dividend paying firms. Thus, a high-growth firm, paying no dividends currently, can still be valued based upon dividends that it is expected to pay out when the growth rate declines. If the payout ratio is not adjusted to reflect changes in the growth rate, however, the dividend discount model will underestimate the value of non-dividend paying or low-dividend paying stocks.

(b) Is the model too conservative in estimating value?

A standard critique of the dividend discount model is that it provides too conservative an estimate of value. This criticism is predicated on the notion that the value is determined by more than the present value of expected dividends. For instance, it is argued that the dividend discount model does not reflect the value of 'unutilized assets'. There is no reason, however, that these unutilized assets cannot be valued separately and added on to the value from the dividend discount model. Some of the assets that are supposedly ignored by the dividend discount model, such as the value of brand names, can be dealt with simply within the context of the model.

A more legitimate criticism of the model is that it does not incorporate other ways of returning cash to stockholders (such as stock buybacks). If you use the modified version of the dividend discount model, this criticism can also be countered.

(c) The contrarian nature of the model

The dividend discount model is also considered by many to be a contrarian model. As the market rises, fewer and fewer stocks, they argue, will be found to be undervalued using the dividend discount model. This is not necessarily true. If the market increase is due to an improvement in economic fundamentals, such as higher expected growth in the economy and/or lower interest rates, there is no reason, a priori, to believe that the values from the dividend discount model will not increase by an equivalent amount. If the market increase is not due to fundamentals, the dividend discount model values will not follow suit, but that is more a sign of strength than weakness. The model is signaling that the market is overvalued relative to dividends and cashflows and the cautious investor will pay heed.

Tests of the Dividend Discount Model

The ultimate test of a model lies in how well it works at identifying undervalued and overvalued stocks. The dividend discount model has been tested and the results indicate that it does, in the long term, provide for excess returns. It is unclear, however, whether this is because the model is good at finding undervalued stocks or because it proxies for well-know empirical irregularities in returns relating to price-earnings ratios and dividend yields.

A Simple Test of the Dividend Discount model

A simple study of the dividend discount model was conducted by Sorensen and Williamson, where they valued 150 stocks from the S&P 400 in December 1980, using the dividend discount model. They used the difference between the market price at that time and the model value to form five portfolios based upon the degree of under or over valuation. They made fairly broad assumptions in using the dividend discount model.

(a) The average of the earnings per share between 1976 and 1980 was used as the current earnings per share.

(b) The cost of equity was estimated using the CAPM.

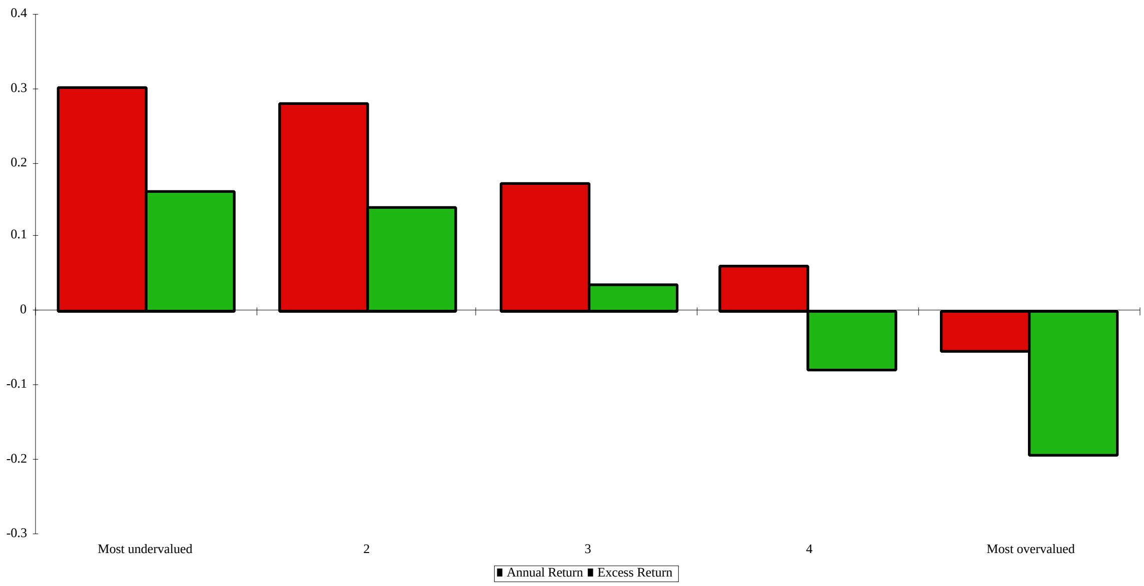
(c) The extraordinary growth period was assumed to be five years for all stocks and the I/B/E/S consensus forecast of earnings growth was used as the growth rate for this period.

(d) The stable growth rate, after the extraordinary growth period, was assumed to be 8% for all stocks.

(e) The payout ratio was assumed to be 45% for all stocks.

The returns on these five portfolios were estimated for the following two years (January 1981-January 1983) and excess returns were estimated relative to the S&P 500 Index using the betas estimated at the first stage and the CAPM. Figure 13.6 illustrates the excess returns earned by the portfolio that was undervalued by the dividend discount model relative to both the market and the overvalued portfolio.

Figure 13.6 Performance of the Dividend Discount Model: 1981-83



The undervalued portfolio had a positive excess return of 16% per annum between 1981 and 1983, while the overvalued portfolio had a negative excess return of 15% per annum during the same time period. Other studies which focus only on the dividend discount model come to similar conclusions. In the long term, undervalued (overvalued) stocks from the dividend discount model outperform (under perform) the market index on a risk adjusted basis.

Caveats on the use of the dividend discount model

The dividend discount model provides impressive results in the long term. There are, however, three considerations in generalizing the findings from these studies.

The dividend discount model does not beat the market every year

The dividend discount model outperforms the market over five-year time periods, but there have been individual years where the model has significantly under performed the market. Haugen reports on the results of a fund that used the dividend discount model to analyze 250 large capitalization firms and to classify them into five quintiles from the first quarter of 1979 to the last quarter of 1991. The betas of these quintiles were roughly equal. The valuation was done by six analysts who estimated an extraordinary growth rate for the initial high growth phase, the length of the high growth phase and a transitional phase for each of the firms. The returns on the five portfolios as well as the returns on all 250 stocks and the S&P 500 from 1979 to 1991 are reported in Table 13.4.

Table 13.4: Returns on Quintiles: Dividend Discount Model

Quintile							
	Under Valued	2	3	4	Over Valued	250 Stocks	S&P 500
1979	35.07%	25.92%	18.49%	17.55%	20.06%	23.21%	18.57%
1980	41.21%	29.19%	27.41%	38.43%	26.44%	31.86%	32.55%
1981	12.12%	10.89%	1.25%	-5.59%	-8.51%	28.41%	24.55%
1982	19.12%	12.81%	26.72%	28.41%	35.54%	24.53%	21.61%
1983	34.18%	21.27%	25.00%	24.55%	14.35%	24.10%	22.54%
1984	15.26%	5.50%	6.03%	-4.20%	-7.84%	3.24%	6.12%
1985	38.91%	32.22%	35.83%	29.29%	23.43%	33.80%	31.59%
1986	14.33%	11.87%	19.49%	12.00%	20.82%	15.78%	18.47%
1987	0.42%	4.34%	8.15%	4.64%	-2.41%	2.71%	5.23%
1988	39.61%	31.31%	17.78%	8.18%	6.76%	20.62%	16.48%
1989	26.36%	23.54%	30.76%	32.60%	35.07%	29.33%	31.49%
1990	-17.32%	-8.12%	-5.81%	2.09%	-2.65%	-6.18%	-3.17%
1991	47.68%	26.34%	33.38%	34.91%	31.64%	34.34%	30.57%
1979-91	*1253%*	*657%*	*772%*	*605%*	*434%*	*722%*	*654%*

The undervalued portfolio earned significantly higher returns than the overvalued portfolio and the S&P 500 for the 1979-91 period, but it under performed the market in five of the twelve years and the overvalued portfolio in four of the twelve years.

Is the model just a proxy for low PE ratios and dividend yields?

The dividend discount model weights expected earnings and dividends in near periods more than earnings and dividends in far periods., It is biased towards finding low price-earnings ratio stocks with high dividend yields to be undervalued and high price-earnings ratio stocks with low or no dividend yields to be overvalued. Studies of market efficiency indicate that low PE ratio stocks have outperformed (in terms of excess returns) high PE ratio stocks over extended time periods. Similar conclusions have been drawn about high-dividend yield stocks relative to low-dividend yield stocks. Thus, the valuation findings of the model are consistent with empirical irregularities observed in the market.

It is unclear how much the model adds in value to investment strategies that use PE ratios or dividend yields to screen stocks. Jacobs and Levy (1988b) indicate that the marginal gain is relatively small.

Attribute *Average Excess Return per Quarter: 1982-87*

Dividend Discount Model	0.06% per quarter
Low P/E Ratio	0.92% per quarter
Book/Price Ratio	0.01% per quarter
Cashflow/Price	0.18% per quarter
Sales/Price	0.96% per quarter
Dividend Yield	-0.51% per quarter

This suggests that using low PE ratios to pick stocks adds 0.92% to your quarterly returns, whereas using the dividend discount model adds only a further 0.06% to quarterly returns. If, in fact, the gain from using the dividend discount model is that small, screening stocks on the basis of observables (such as PE ratio or cashflow measures) may provide a much larger benefit in terms of excess returns.

The tax disadvantages from high dividend stocks

Portfolios created with the dividend discount model are generally characterized by high dividend yield, which can create a tax disadvantage if dividends are taxed at a rate greater than capital gains or if there is a substantial tax timing[6] liability associated with dividends. Since the excess returns uncovered in the studies presented above are pre-tax to the investor, the introduction of personal taxes may significantly reduce or even eliminate these excess returns.

In summary, the dividend discount model's impressive results in studies looking at past data have to be considered with caution. For a tax-exempt investment, with a long time horizon, the dividend discount model is a good tool, though it may not be the only one, to pick stocks. For a taxable investor, the benefits are murkier, since the tax consequences of the strategy have to be considered. For investors with shorter time horizons, the dividend discount model may not deliver on its promised excess returns, because of the year-to-year volatility in its performance.

Conclusion

When you buy stock in a publicly traded firm, the only cash flow you receive directly from this investment are expected dividends. The dividend discount model builds on this simple propositions and argues that the value of a stock then has to be the present value of expected dividends over time. Dividend discount models can range from simple growing perpetuity models such as the Gordon Growth model, where a stock's value is a function of

[6] Investors do not have a choice of when they receive dividends, whereas they have a choice on the timing of capital gains.

its expected dividends next year, the cost of equity and the stable growth rate, to complex three stage models, where payout ratios and growth rates change over time.

While the dividend discount model is often criticized as being of limited value, it has proven to be surprisingly adaptable and useful in a wide range of circumstances. It may be a conservative model that finds fewer and fewer undervalued firms as market prices rise relative to fundamentals (earnings, dividends, etc.) but that can also be viewed as a strength. Tests of the model also seem to indicate its usefulness in gauging value, though much of its effectiveness may be derived from its finding low PE ratio, high dividend yield stocks to be undervalued.

Problems

1. Respond true or false to the following statements relating to the dividend discount model:
   A. The dividend discount model cannot be used to value a high growth company that pays no dividends.
   B. The dividend discount model will undervalue stocks, because it is too conservative.
   C. The dividend discount model will find more undervalued stocks, when the overall stock market is depressed.
   D. Stocks that are undervalued using the dividend discount model have generally made significant positive excess returns over long time periods (five years or more).
   E. Stocks which pay high dividends and have low price-earnings ratios are more likely to come out as undervalued using the dividend discount model.

2. Ameritech Corporation paid dividends per share of $3.56 in 1992 and dividends are expected to grow 5.5% a year forever. The stock has a beta of 0.90 and the treasury bond rate is 6.25%.
   a. What is the value per share, using the Gordon Growth Model?
   b. The stock was trading for $80 per share. What would the growth rate in dividends have to be to justify this price?

3. Church & Dwight, a large producer of sodium bicarbonate, reported earnings per share of $1.50 in 1993 and paid dividends per share of $0.42. In 1993, the firm also reported the following:

Net Income = $30 million
Interest Expense = $0.8 million
Book Value of Debt = $7.6 million
Book Value of Equity = $160 million

The firm faced a corporate tax rate of 38.5%. (The market value debt to equity ratio is 5%.) The treasury bond rate is 7%.

The firm expected to maintain these financial fundamentals from 1994 to 1998, after which it was expected to become a stable firm with an earnings growth rate of 6%. The firm's financial characteristics were expected to approach industry averages after 1998. The industry averages were as follows:

Return on Capital = 12.5%
Debt/Equity Ratio = 25%
Interest Rate on Debt = 7%

Church and Dwight had a beta of 0.85 in 1993 and the unlevered beta was not expected to change over time.

a. What is the expected growth rate in earnings, based upon fundamentals, for the high-growth period (1994 to 1998)?

b. What is the expected payout ratio after 1998?

c. What is the expected beta after 1998?

d. What is the expected price at the end of 1998?

e. What is the value of the stock, using the two-stage dividend discount model?

f. How much of this value can be attributed to extraordinary growth? to stable growth?

4. Oneida Inc, the world's largest producer of stainless steel and silverplated flatware, reported earnings per share of $0.80 in 1993 and paid dividends per share of $0.48 in that year. The firm was expected to report earnings growth of 25% in 1994, after which the growth rate was expected to decline linearly over the following six years to 7% in 1999. The stock was expected to have a beta of 0.85. (The treasury bond rate was 6.25%)

a. Estimate the value of stable growth, using the H Model.

b. Estimate the value of extraordinary growth, using the H Model.

c. What are the assumptions about dividend payout in the H Model?

5. Medtronic Inc., the world's largest manufacturer of implantable biomedical devices, reported earnings per share in 1993 of $3.95 and paid dividends per share of $0.68. Its earnings were expected to grow 16% from 1994 to 1998, but the growth rate was expected to decline each year after that to a stable growth rate of 6% in 2003. The payout ratio was expected to remain unchanged from 1994 to 1998, after which it would increase each year to reach 60% in steady state. The stock was expected to have a beta of 1.25 from 1994 to 1998, after which the beta would decline each year to reach 1.00 by the time the firm becomes stable. (The treasury bond rate was 6.25%)

a. Assuming that the growth rate declines linearly (and the payout ratio increases linearly) from 1999 to 2003, estimate the dividends per share each year from 1994 to 2003.

b. Estimate the expected price at the end of 2003.

c. Estimate the value per share, using the three-stage dividend discount model.

The U.S. Equity Return Premium: Past, Present and Future

J. Bradford DeLong and Konstantin Magin

February 2008

J. Bradford DeLong is Professor of Economics, University of California at Berkeley, Berkeley, California; Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts; and Visiting Scholar, Federal Reserve Bank of San Francisco, San Francisco, California.

Konstantin Magin is visiting professor at the University of California at Berkeley, Berkeley, California.

Their e-mail addresses are <delong@econ.berkeley.edu> and <magin@berkeley.edu>, respectively.

We would like to thank Bob Anderson, Nick Barberis, Roger Craine, Christopher DeLong, Milton Friedman, Jim Hines, Chad Jones, Ulrike Malmendier, Rajnish Mehra, Andrei Shleifer, Timothy Taylor, Robert Waldmann, Marty Weitzman, and especially Jeremy Stein for helpful comments and discussions. Part of this research was supported by the National Science Foundation. The opinions expressed are those of the authors alone, and not of the Federal Reserve Bank of San Francisco or any other institution.

ABSTRACT

For more than a century, diversified long-horizon investors in America's stock market have invariably received much higher returns than investors in bonds: a return gap averaging some six percent per year that Rajnish Mehra and Edward Prescott (1985) labeled the "equity premium puzzle." The existence of this equity return premium has been known for generations: more than eighty years ago financial analyst Edgar L. Smith (1924) publicized the fact that long-horizon investors in diversified equities got a very good deal relative to investors in debt: consistently higher long-run average returns with less risk. As of this writing— October 16, 2007, 11.44 PDT—the annual earnings yield on the value-weighted S&P composite index is 5.53%. This is a wedge of 3.22 % per year when compared to the annual yield on 10-year Treasury inflation-protected bonds of 2.31%. The existence of the equity return premium in the past offered long-horizon investors a chance to make very large returns in return for bearing little risk. It appears likely that the current configuration of market prices offers a similar opportunity to long-horizon investors today.

I. Introduction

For more than a century, diversified long-horizon investors in America's stock market have invariably received much higher returns than investors in bonds: a return gap averaging some six percent per year that Rajnish Mehra and Edward Prescott (1985) labeled the "equity premium puzzle." The existence of this equity return premium has been known for generations: more than eighty years ago financial analyst Edgar L. Smith (1924) publicized the fact that long-horizon investors in diversified equities got a very good deal relative to investors in debt: consistently higher long-run average returns with less risk. It was true, Smith wrote three generations ago, that each individual company's stock was very risky: "subject to the temporary hazard of hard times, and [to the hazard of] a radical change in the arts or of poor corporate management." But these risks could be managed via diversification across stocks: "effectively eliminated through the application of the same principles which make the writing of fire and life insurance policies profitable."

Edgar L. Smith was right.

Common stocks have consistently been extremely attractive as long-term investments.

Over the half century before Smith wrote, the Cowles Commission index of American stock prices deflated by consumer prices shows an average real return on equities of 6.5 percent per year— compared to an average real long-term government bond return of 3.6 percent and an average real bill return of 4.5 percent.[1] Since the start of the twentieth century, the Cowles Commission index linked to the Standard and Poor's Composite shows an average real equity return of 6.0 percent per year, compared to a real bill return of 1.6 percent per year and a real long-term government bond return of 1.8 percent per year. Since World War II equity returns have averaged 6.9 percent per year, bill returns 1.4 percent per year, and bond returns 1.1 percent per year. Similar gaps between stock and bond and bill returns have typically existed in other economies. Mehra (2003)[2] reports an annual equity return premium of 4.6 percent in post-World War II Britain, 3.3 percent in Japan since 1970, and 6.6 percent and 6.3 percent respectively in Germany and Britain since the mid-1970s.

Edgar Smith was right about both his past and our past. It appears likely[3] that Smith is right about our future as well. The arguments that the equity return premium should not be a puzzle in the future appear to imply that the equity return premium should not have existed in the past, yet it did.

The equity return premium has existed in the American stock market since it consisted of

---

[1]In the data set of Robert Shiller (2006): http://www.econ.yale.edu/~shiller/data.htm.

[2]Citing Jeremy Siegel (1998) and John Campbell (2001).

[3]Along with Rajnish Mehra (2006).

a few canal and railroad companies and John Jacob Astor's fur-trading empire. Its existence has been broadly known for 80 years. It is one of the most durable macroeconomic facts in the economy. Thus it appears overwhelmingly likely that the equity return premium has a future as well as a past, and there is little or no apparent reason for us economists to believe that in this case we economists know better than the market.

II. The Arithmetic of the Equity Premium

To pose the equity premium return puzzle, consider a marginal investor with a 20-year horizon—somebody in elementary school receiving a bequest from grandparents, somebody in their 30s with children putting money away to spend on college, somebody age 50 contemplating medical bills or wanting to leave a bequest, a life-insurance company collecting premiums from the middle-aged, or a company offering its workers a defined-benefit pension.

One margin such an investor must consider is the choice between:

(1) investing in a diversified portfolio of equities, reinvesting payouts and rebalancing periodically to maintain diversification;

(2) investing in short-term safe bills, rolling the portfolio over into similar short-term debt instruments as pieces of it mature.

The marginal investor must expect that their marginal dollars would be equally attractively employed in each of these strategies.

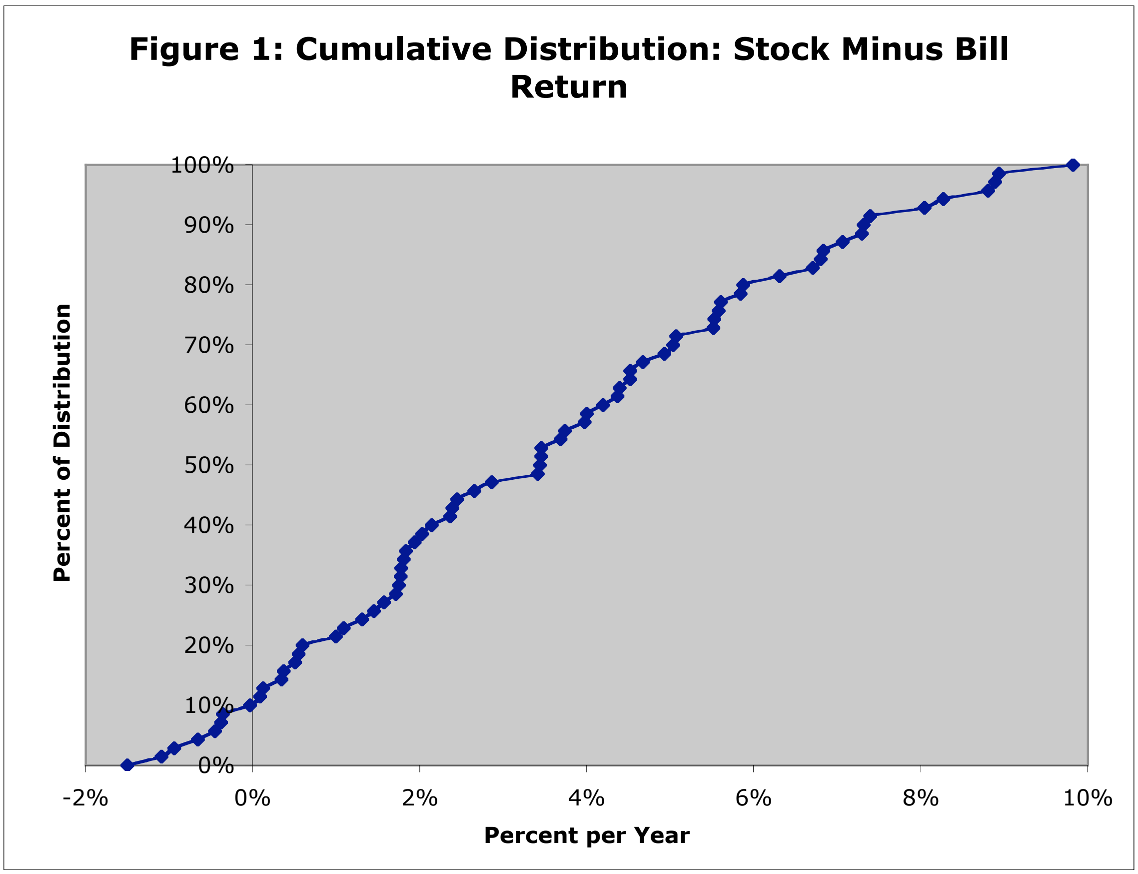
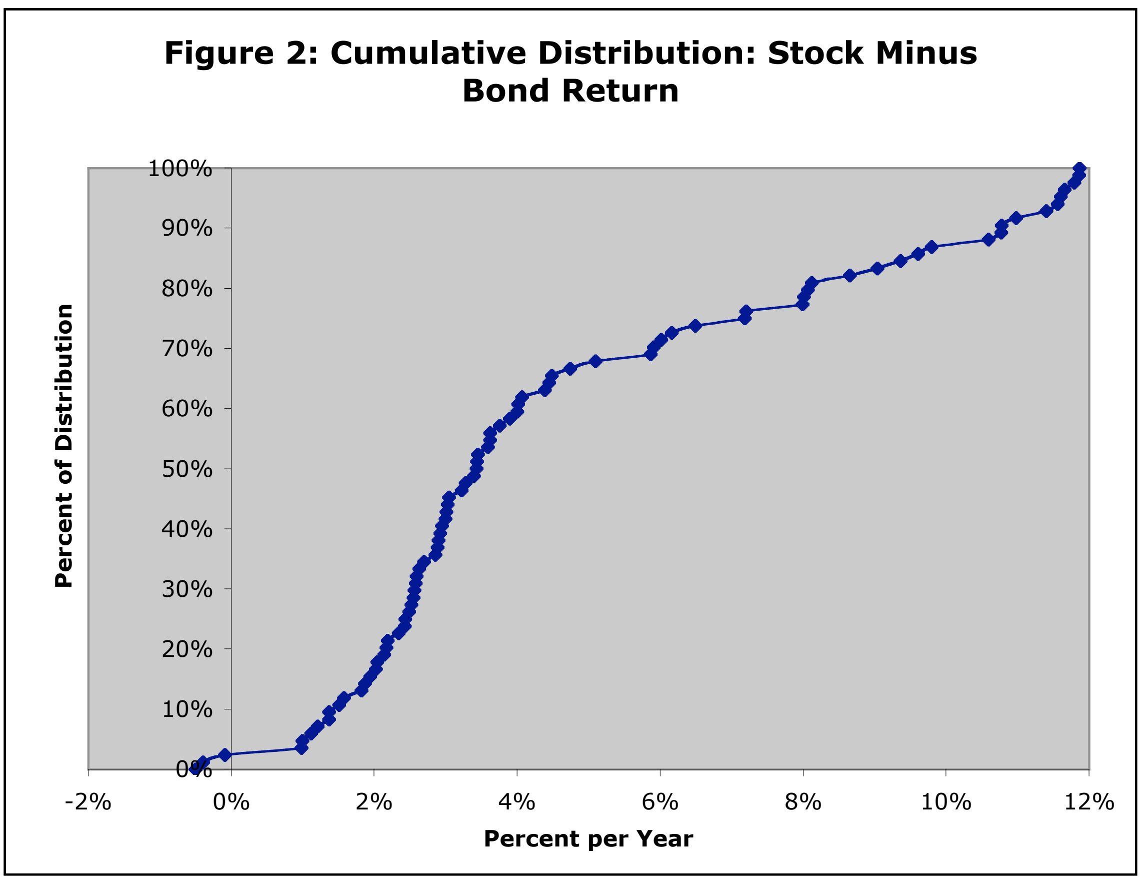


Figure 1 plots the cumulative return distribution for the relative returns for these two twenty-year portfolio strategies starting in each year since the start of the twentieth century. The average geometric return differential since 1901 is some 4.9 percent per year. When the portfolios are cashed in after twenty years, investments in diversified stock portfolios are on average 2.67 times as large as an investment in short-term Treasury bills after twenty years. Stock investors more than double their relative wealth 60 percent of the time, more than quadruple their relative wealth 30 percent of the time, and have a 17 percent chance of a more than seven-fold multiplication of relative wealth. The downside is small: the empirical CDF finds that stocks do worse than bills less than 9 percent of the time. The very worst case observed is the 20 years starting in 1965, when investing in stocks yields a relative cumulative wealth loss of 17 percent compared to investing in bills.



This equity return premium is not a liquidity effect driven by the special ease with which short-term bills can be turned into cash even in emergencies. Figure 2 shows the CDF of relative returns from the twenty-year strategies of investing in a diversified stock portfolio and investing in a long-term Treasury bond portfolio. This time lower tail is even smaller: in only 2 percent of the cases in the twentieth century would investing in bonds for 20 years outperformed investing in stocks. In the worst relative case—1929—the returns to bonds would have been only 8 percent more than stocks when the portfolios were cashed in 1949.

If the actual twentieth-century CDF is a good proxy for the true underlying *ex ante* return distribution, these return patterns have powerful implications for investors' expectations about their relative marginal utility of wealth. If the marginal investor's marginal dollar is no more advantageously employed in stocks than bonds, it must be the case that:

$$\frac{\text{(chance of loss)}\left[\mathit{Average}\left[\text{(amount of loss)} \times \text{(marginal utility of wealth if loss)}\right]\right]}{\text{(chance of gain)}\left[\mathit{Average}\left[\text{(amount of gain)} \times \text{(marginal utility of wealth if gain)}\right]\right]} = 1$$

Over the twentieth century, the chance of relative gain is ten times the chance of loss. The average amount of gain—167%—is seventeen times the average amount of loss. If the marginal utility in gain states is perfectly correlated with the amount of gain and the marginal utility in loss states uncorrelated with the amount of loss, then the average marginal utility of wealth in "stocks lose" states must be 50 times as great as in "stocks gain" states. This is the equity return premium puzzle at its sharpest: how is one to account for this extraordinary divergence?

The equity premium puzzle appears softer if attention is focused on short-horizon investors who invest for one year only. Stocks are very risky in the short run. 1931 sees a return differential of –60%. And bonds have outperformed stocks in some 35% of the past century's years. Twenty-year investors appear to have turned their backs on nearly riskless opportunities for profit. One-year investors did not. For investors with a time horizon of one year, stocks *are* much more risky than bills[4].

Yet even on a year-to-year scale the equity premium return remains. And there are no visible[5] large year-to-year fluctuations in the consumption of investors correlated with

---

[4]One reason that the puzzle is softer at short horizons is that a substantial share of year-to-year variability in the stock market appears to be transitory. Stock prices look as though they are somewhat mean reverting: at the level of the stock market as a whole, past performance is not only not a guarantee of future results, past performance is negatively correlated with future results. The variance of 20-year stock returns is only 45% of what it would be if returns were serially uncorrelated (see, for example, Cochrane, 1994; Cochrane, 2006; Campbell and Shiller, 1989). Thus Samuelson (1969)'s proof that horizon is irrelevant for asset allocation fails to go through. Mean reversion can make long-term equity investments more attractive than short-term investments because investments made at one moment insure against investments made at another.

[5] Barro (2005) and others believe that there is here a small numbers problem: with a long enough sample

stock returns that would create a high marginal utility of wealth in “stocks lose” states and so account for the premium. At the one-year horizon an investor would be indifferent at the margin between stocks and bills only if he or she had a marginal utility of wealth in the gain state 83% of the way up the return distribution that was half that of marginal utility in the loss state 17% of the way up. Such a difference in marginal utilities is very difficult to square with the low variability in aggregate consumption: Rajnish Mehra and Edward Prescott (2003) report an annual standard deviation of consumption growth of only 3.6%, which they believe could support an equity return premium for a representative investor of at most two-tenths of a percentage point per year—not six.

The basic point is Richard Thaler and Matthew Rabin (2001): expected utility theory pushes us economists toward the view that agents should be nearly risk-neutral on all bets that do not involve a substantial fraction of lifetime wealth, for only substantial variations in lifetime wealth and thus in current consumption produce enough variation in marginal utility to justify substantial risk aversion. And annual stock market returns do not covary enough with current consumption and lifetime wealth.

Thus order to solve the equity premium puzzle, an economist must propose an explanation that does at least one of:

---

we would see occasional collapses in consumption and stock values that would account for what we have observed.

- providing a reason for a very large gap in the marginal utility of wealth between states of the world in which stocks do well and states of the world in which stocks do poorly.
- demonstrating that the *ex-post* return distribution seen over the twentieth century is very different from the true *ex*-ante distribution in important ways that make stocks no real bargain.
- explaining why it is that, even though stocks have been an extremely attractive investment relative to bonds and bill, money has not flowed out of bonds and bills and into stocks—pushing equity prices up and equity returns down.

A very large number of economists have done excellent work investigating and assessing different potential explanations. Among the most promising lines of work have been investigations of the implications of risk aversion, non-standard preferences; transactions costs; lower-tail risk; persistent mistakes; investor confusion; and cognitive biases.[6] A full and satisfactory explanation of the equity premium return puzzle continues to elude economists. However, none of what appear to be the live possibilities would lead one to anticipate the disappearance of the premium in the future.

III. A Preferences Explanation?

A first potential explanation is simply that rational investors prefer the portfolios they

[6]Of course, space prevents us from even noting the existence of more than a very small fraction of even the most important contributions to the literature. We can only glance at those we regard as most promising.

hold: investors truly are risk averse enough that the observed configuration of returns does not leave unexploited profit opportunities. The difficulties are twofold: first, the low average return debt securities used as a yardstick in measuring the equity return premium are not really low in risk; second, even taking debt to be risk free the degree of risk aversion needed to keep long-term investors from seeing large gains from further investments in equities must be extremely high.

As the late Fisher Black once put it in conversation, in terms of the coefficient of relative risk aversion—the standard way of measuring tolerance for risk—explaining the configuration of asset returns requires a coefficient of about 50. Consider of an agent offered a choice between (a) their current lifetime wealth and (b) a gamble where with probability p they obtain twice and with probability 1-p half their lifetime wealth. An agent with a coefficient of 2 would reject (b) if p were less than 80%; for a coefficient of 10 the critical value is 99.8%; and for a coefficient of 50 the critical value is 99.99999999995%. Many economists argue that both observed purchases of insurance and our intuitions suggest a coefficient of relative risk aversion parameter not of 50 but more in the range of 1 to 3,[7] which corresponds to Mehra-Prescott's estimate of a warranted equity premium of about 0.2 percentage points per year.

Moreover, as we economists learned from Philippe Weil (1989), a standard time-separable utility function with a high degree of risk aversion also generates both a high

[7]See, for example, Partha Dasgupta (2007).

risk-free rate of return (in economies with the roughly two percent per year consumption growth of our own economy) and smooth consumption paths that do not respond to changes in rates of return. Neither of these is observed

The most promising preference-based line of research—exemplified by papers like Lawrence Epstein and Stanley Zin (1991), George Constantinides (1990), Andrew Abel (1990), and John Campbell and John Cochrane (1995)—considers non-standard preferences, making utility dependent not just on consumption but on consumption relative to the consumption of others or to one's own past consumption and separating preferences for risk from preferences for income growth over time. These approaches account for the coexistence of a high degree of effective risk aversion and a low risk-free interest rate: the features of the utility function that make investors extremely averse to stock-market losses have no bearing on the connection between economic growth and the safe real interest rate. But these approaches still require something to generate very high effective risk aversion.

Narayana Kocherlakota (1996) summed up the results from this line of research:

> The risk-free rate puzzle can be resolved as long as the link between individual attitudes toward risk and growth contained in the standard preferences is broken…. [T]he equity premium puzzle is much more robust: individuals must either be highly averse to their own consumption

risk or to per capita consumption risk…

The modern finance literature on the equity premium puzzle is now more than two decades old. The historical investment literature looking back into observers' pasts and noting the existence of a very large equity return premium is now more than eight decades old. Yet to date no critical mass of long-term investors has taken large-enough long-enough-run positions to try to profit from the equity return premium to substantially arbitrage it away.

It is premature to say that these lines of research will never be able to satisfactorily account for the equity premium that has been observed in the past. But they do not to date appear to have done so. It is not clear how they might do so. If, however, they turn out to be correct, they do imply a future equity return premium likely to be about the six percent or so a year observed in the past.

An alternative is offered by behavioral finance economists, for example Benartzi and Thaler (1995), see investors—even professional and highly-compensated investors in it for the long run—as institutionally and psychologically incapable of framing their portfolio-choice problem in a way that allows them to appropriately discount and thus ignore the high short-term risks of equities. If investors could focus instead on the long-term returns of stocks they would realize that there is very little long-term risk in stocks relative to bonds. But they cannot. Rabin and Thaler (2001) argue that expected utility maximization cannot account for most behavior economists label “risk averse,” and should be replaced by “loss aversion” as a model of investor behavior—individuals simply feel the pain of a loss more acutely than the pleasure of an equal-sized gain. Hong and Stein point to “disagreement models” that motivate high trading volumes as a potential explanation for other asset pricing anomalies like the equity premium. Glamor stocks exhibit greater than average turnover rates, high trading volumes, tend to be overpriced and exhibit low rates of return; value stocks exhibit lower than average turnover rates, low trading volumes, tend to be underpriced, and exhibit high rates of return: perhaps this could be built into an explanation of the equity return premium.

It is not clear whether these are explanations of the puzzle or reframings of it. Humans know that they have psychological biases, and build social and economic institutions to compensate for them and to guide them into framing problems in a way that is in their long-term interest. Humans have built mechanisms like automatic payroll deductions, like inducing caution by valuing assets at the lower of cost and market, like entails and trusts. A bias-based psychological explanation must account not just for the bias but for the failure of investors to figure out ex ante how to bind themselves to the mast like Ulysses did with the Sirens.

IV. Transaction Costs and Investor Heterogeneity

Another line of research has attempted to explain the equity premium as due to transaction costs and investor heterogeneity.[8] Gregory Mankiw and Stephen Zeldes (1991) were among the very first to point out that two-thirds of Americans have next to no stock market investments—presumably because of some form of transaction cost that keeps them from being able to recognize and act on the fact that equity investments have a substantial place in every optimal portfolio. Transactions costs keeping a substantial share of the population at a zero position lock up what representative-agent models see as society's risk-bearing capacity, which then cannot be tapped and mobilized to bear equity risk.

Mankiw and Zeldes found that stockholders' consumption does not vary nearly enough to account for the equity premium. If standard representative agent models suggest that the warranted equity return premium should be on the order of 0.2 percentage points per year, a transactions-cost model in which only one-third of agents hold stocks suggests a warranted equity premium on the order of three times as large. This line of research could diminish the magnitude of the equity premium puzzle,[9] but appears to still leave an order of magnitude gap to be accounted for.

---

[8]These go together: if investors are effectively identical they do not trade and transactions costs are irrelevant; if there are no transactions costs than investor heterogeneity does not reduce the net risk-bearing capacity of the economy.

[9] See Vissing-Jorgenson (2002).

This line of research also leaves unanswered the question of just what these transaction costs are. Even back in the nineteenth century "bucket shops"—most of them honest—allowed people with very small amounts of money to "invest:" as little as one dollar could "buy" or "sell" a fractional share at the last ticker price. A bucket shop was not a brokerage. It did not invest its clients' money in the market: it paid today's withdrawals out of yesterday's deposits and relied on commissions and the law of large numbers to make it profitable.[10] And even if there were large transaction costs to buying and selling stocks, could this account for the equity premium puzzle? High costs of buying and selling are amortized over decades when investors follow multi-decade buy-and-hold strategies, and the most vivid advantages of stock investments produced by the equity return premium accrue to those who follow such strategies.

More recently, Constantinides, Donaldson, and Mehra (2002) suggest that the equity premium may be due to transaction costs in the form of borrowing constraints. The relatively young with the option of declaring bankruptcy have difficulty borrowing on a large scale. Because of such borrowing constraints, investors find it optimal to build up stocks of liquid wealth (see, for example, Mark Huggett, 1993; John Heaton and Deborah Lucas, 1995). This argument takes us economists far toward explaining why the risk-free rate of return might be low: people's unwillingness to have even temporarily negative net

[10]Nineteenth-century speculator Daniel Drew found when young that he did better at bucket shops than on Wall Street. His actual purchases and sales generated price pressure against himself, while his notional bucket shop transactions did not.

worth increases saving, increases the capital stock, and so pushes down the rate of interest and profit. But could such borrowing constraints bear much of the weight of accounting for the equity premium? Built-up stocks of wealth could be invested in either stocks or bonds, and stocks offer higher returns with little extra long-horizon risk.

The transaction costs approach that in our view comes closest to accounting for the equity premium puzzle is that of George Constantinides and Darryl Duffie (1996). They propose that investors are subject to uninsurable idiosyncratic income shocks correlated with returns on equities. Thus investors bear a large amount of equity risk embedded in their human capital, and are uninterested in further leveraging their total implicit portfolios. Advancing this explanation would require identifying groups of people whose labor income is subject to shocks correlated with equity returns and demonstrating that those investors' portfolios drive the lack of investment in equities. This has not yet been accomplished.

V. Lower-Tail Risk?

The equity premium return puzzle might be resolved by breaking the assumption that the ex post return distribution over the twentieth century is an adequate proxy for the ex ante return distribution. A high equity premium might be observed in the sample that is our past if that sample does not contain low-probability but large-magnitude economic catastrophe. A small chance of winding up truly far out in the lower tail of a return distribution can have a significant effect on ex ante and—if unobserved in sample—an even more significant effect on ex post return premia. Proposed solutions along these lines have been put forward by authors like Thomas Rietz (1988); Stephen Brown, William Goetzmann, and Stephen Ross (1995); and Robert Barro (2005). If correct, this family of solutions would imply that we economists will continue to observe a large equity premium in-sample for a while—until The Day when the long run arrives while some of us at least are still alive, the economic catastrophe occurs, and investors find their stocks nearly worthless.

This explanation must pass a camel through the eye of a needle. The unobserved-in-sample low-probability catastrophe must occur with a probability small enough that it is plausible that it has not observed. Yet the chance and magnitude of the catastrophe must be large enough to have substantial effects on prices and returns. And the catastrophe must diminish the value of stocks but not of bonds or bills—for a catastrophe that hits stocks and bonds equally has no effect on the equity premium return.[11]

This theory has considerable attractiveness. But it has one principal difficulty: it is not

[11]There is a fourth requirement, for too great a risk of a collapse in the stock market and in consumption will not only produce a high equity premium but a negative real interest rate. The size of collapse must be on a knife-edge in these models: large but not too large—large enough to create the observed equity premium, but small enough to leave a positive safe real interest rate.

obvious what the low-probability economic catastrophes with powerful negative impacts on real equity returns and little effect on bond returns are. Investors and economists can envision a great many potential political and economic catastrophes: defeat in a major war; a populist unraveling of government finances generating hyperinflation; an exhaustion of technological possibilities for innovation; or a banking-sector collapse or other financial crisis that generates a steep but transitory collapse in profits. However these catastrophes are likely to affect both stock and bond values. A permanent decline in the rate of total factor productivity and consumption growth ought to affect stock and bond returns proportionately. War defeat or populist-crisis crashes of government finance are highly likely to produce rapid inflation, which is poison to real debt returns. A transitory collapse in corporate profitability has little effect on far-sighted valuations of equities—unless it is accompanied by a collapse in consumption as well, in which case the reduced tax base is likely to lead to substantial money printing and inflation.

A large deflationary episode like the Great Depression itself could serve as a source of risk to stocks but not bonds. Few, however, believe that any future central bank would allow such a steep and persistent deflation as the Federal Reserve allowed in the 1930s. And the Great Depression is already in our sample. It is hard to argue that its absence from our sample is the cause of the observed equity return premium puzzle.

This difficulty applies also to the "survivorship" argument that looking across countries the U.S. is a large positive outlier in stock returns. It is a large positive outlier in bond returns as well.

There is one possible source that can be envisioned of a collapse in real equity values that would not much affect the real values of government bonds. If the U.S. government were to decide to put extraordinarily heavy taxes on corporate profits or to impose extraordinarily heavy regulatory burdens on corporations, those policies could redirect a substantial amount of cash flow away from shareholders without affecting bond values. Yet is the rational fear of future tax increases or regulatory burdens narrowly targeted on corporate profits large enough to support anything like the observed equity premium? But perhaps we overestimate the competence of our government, and underestimate the strength of a populism that really does believe that when the government taxes corporations no individual pays. Moreover, as public finance economists like James Hines (2005) point out, in a world of mobile capital tax competition restrains governments from pursuing tax policies very different from those of other nations. A radical failure of such tax competition would have to be required as well.

An analogous argument to Rietz (1988) and Barro (2005) is made by Martin Weitzman (2006). Weitzman argues not that lower tail risk is large, but that investors do not and cannot know what the lower tail risk truly is: Knightian uncertainty rather than von Neumann-Morgenstern risk. Once again, the principal difficulty is to identify the potential the events that investors believe might generate a long fat lower tail of equity returns and yet leave real government debt returns unaffected.

A final unresolved difficulty with the unobserved lower-tail hypothesis is that, as Barro (2005) points out, this explanation carries the implication that the greater the chance of a collapse the higher are equity prices. In this theory, 2000 is a year in which investors expected a high, and 1982 a year in which investors expected a low, probability of macroeconomic disaster.[12]

If the arguments for heretofore unobserved lower-tail risk hold true, then the appearance of an equity premium puzzle will not persist forever. At some point the risks that underpin the asset price configuration would manifest themselves, at which point it will become very clear that the equity premium puzzle never really existed at all.

VI. Learning About the Return Distribution

Yet another path assumes that economic agents are not extraordinarily risk averse, that economic agents are not limited in their risk-bearing capacity by transactions costs and heterogeneity, that the in-sample return distribution is a good proxy to the ex-ante return distribution, but that investors early in the twentieth century mistook the parameters of

[12]This is a somewhat disturbing artifact of the Lucas (1978) model that underpins papers like Rietz (1988), Barro (2005), Weitzmann (2006), and Mehra and Prescott (1985).

the fundamental return distribution, and that it has taken them a very long time indeed to learn what the true parameters of the fundamental return distribution are. Thus misperceptions created the equity premium. And the process of correcting these misperceptions has given a boost to stock prices that has further driven up the in-sample equity premium. This argument carries a corollary: the equity premium has a solid past, but it will not have as much of a future: investors have learned and will continue learn from experience over time, and if there is an equity return premium still in existence today it is likely to shrink relatively rapidly.

McGrattan and Prescott (2003) develop this argument by pointing to changing institutions as a source of the equity premium in the past that is not present today. Regulatory restrictions imposed by legislatures and courts that had too great a fear of the riskiness of equities used to encourage over-investment in debt by pension funds. Until the passage of ERISA in the mid-1970s it was unclear what a pension fund trustee could and could not do without risking legal liability. But it was clear that a trustee who invested in investment-grade bonds was in a safe harbor with respect to any possible legal liability for maladministration. And it was clear that a trustee who invested in stocks was not in a safe harbor. As time passed and as even government officials learned that the riskiness of stocks had been overstated, these regulatory restrictions fell. Thus changing expectations working through the channel of the creation of better financial institutions greatly contributed to this fall in the market risk premium on stocks.

Yet another exploration of this alternative is Olivier Blanchard (1993), who sees two major macroeconomic events driving the movements of the equity premium from 1927 until the early 1990s. He sees high equity premiums as a reaction to the shock of the Great Crash of 1929-1933, and a subsequent decline as the memory and thus the perceived likelihood of a repetition of that extraordinary event has dimmed. He also sees, as do others like Modigliani and Cohn (1979), Campbell and Vuolteenaho (20040), and Randolph Cohen, Chris Polk, and Tuomo Vuolteenaho (2005), a strong correlation of the equity premium and inflation in the 1970s and the 1980s. John Campbell and Tuomo Vuolteenaho (2004) call this effect of inflation on the equity premium a "mispricing" attributed to expectations implicit in market prices "deviating from the rational forecast." They point to Wall Street traders' use of the 'Fed model' to value stocks—believing that the *nominal* coupon yield on debt ought to be in some equilibrium relationship with the *real* earnings yield on equity—as a conceptual error that generates inflation illusion.[13]

These factors led Blanchard back in 1993 to predict that the future equity premium would "remain small," because inflation was likely to remain low and because the memory of the Great Depression was dim and would continue to erode. But Blanchard's regressions were reduced forms, and changing economic institutions and structures would lead one to fear that reduced forms might not track their future very well, and indeed this did not.

[13]It is not clear whether Campbell and Vuoleenaho view this as a misperception to be corrected by learning or as the result of psychological biases that cause confusion between real and nominal magnitudes that will persist.

Over the fourteen years from 1993 to 2007 the real return on Treasury bills has been 2.1 percent while the real return on stocks has been 7.6 percent, for an equity premium of 5.5 percent per year. Perhaps post-1993 estimates of the equity premium are high because of the stock market boom of the late 1990s, but the data since the early 1990s provides little evidence that the equity premium faded away with the vanishing of the memory of the Great Depression and the inflation of the 1970s. An 18 year-old runner from the floor of the New York Stock Exchange in 1929 would have turned 96 in 2007.

What appears as the most powerful attempt to flesh out this alternative is Fama and French (2002). Over the medium run, they argue, the risk premium on stocks has fallen as a result of the correction of misapprehensions about riskiness. Such a fall in the risk premium shows up as a jump in stock prices. Thus learning that the ex-ante equity premium should be lower than in the past produces an in-sample past equity premium even higher than its misperceived ex-ante value.

Fama and French thus argue that one should not estimate the post-World War II ex ante equity premium by looking at ex-post returns—that is, adding dividend yields to the rate of growth of stock prices. That procedure is biased because it includes this unanticipated windfall from learning about the world. One should, instead, estimate expected stock returns via the Gordon Equation:

$$r = D/P + g$$

where D/P is the dividend yield and g is the expected rate of capital gain. The dividend yield is directly observable. The expected capital gain is not, and must be estimated.

VII. The Future of the Equity Premium

What are the implications of taking Fama and French's advice, and estimating the future equity premium via this Gordon-equation approach? A natural way to estimate expected capital gains is to look at their value over the past. But estimating the expected capital gain by averaging past capital gains will be biased upward when—as Fama and French argue—the past contains learning about reduced risks that lowered required rates of return. On the other hand, estimating the expected capital gain by averaging past rates of dividend growth will be biased downward when—as has happened over the past two generations—firms have substituted stock buybacks for dividends as a wa y of pushing money out of the firm. Estimating the expected capital gain in the Gordon mode from the average of past rates of earnings growth avoids much but not all of this last bias: today's higher rate of retained earnings should fuel somewhat faster earnings growth than was generated by lower rates of retained earnings in the past.

Estimating future stock returns via the Gordon model from today's dividend yield and using the post-WWII average rate of earnings growth to forecast expected capital gains produces an expected equity premium of 4.3% per year.

But, as Fama and French further observe, we economists have had good macroeconomic news over the past century: earnings growth since 1950 has probably exceeded what would have been rational expectations formed in the shadow of the Great Depression. Thus Fama and French assess the likely equity premium going forward as likely to be less than this 4.3% per year.

The Gordon equation approach, however, faces a Modigliani-Miller problem. Optimizing firms have chosen their dividend yields for a reason. If dividend yields are currently low it might be because opportunities to invest retained earnings are especially high—in which case properly anticipated likely capital gains in the future will be higher than past historical averages. If dividend yields are currently high it might be because opportunities to invest retained earnings are especially poor—in which case properly anticipated likely capital gains in the future will be lower than past historical averages. An alternative favored by Siegel (2007) is to attempt to estimate equity returns by looking at earnings yields.

The wedge between accounting earnings yields and bond rates is not necessarily the expected equity premium. Do accounting earnings overstate or understate the true Haig-Simons earnings of the corporation, and by how much? By how much do stock options granted but not yet exercised dilute ownership, and so reduce earnings per share? What proportion of the current earnings yield is a cyclical phenomenon? To what extent do retained earnings reinvested inside of firms earn higher rates of return than outside investments subject to information and incentive problems? To what extent to retained earnings reinvested inside of firms earn lower rates of return than outside investments because of corporate control issues? Are there expectations of changes in expected rates of return which thus induce expected capital gains and losses that drive a further wedge between accounting profitability and expected real returns?

Cutting through this Gordian knot of issues, if expected rates of return are constant, accounting earnings equal Haig-Simons earnings, stock options do not much dilute ownership, earnings are not much boosted or depressed by the business cycle, and retained earnings yield the same return as outside investments, then the accounting earnings yield is the expected rate of return. As of this writing— October 16, 2007, 11.44 PDT—the annual earnings yield on the value-weighted S&P composite index is 5.530%. This is a wedge of 3.220 percent per year when compared to the annual yield on 10-year Treasury inflation-protected bonds of 2.310%.

Thus both Gordon and earnings-based approaches confirm the research-surveying judgment in Rajnish Mehra (2003) that the equity premium is likely to persist into the future, but at a level somewhat but not enormously smaller than the original estimated Mehra and Prescott (1985) 6 percent per year. As Mehra (2003) wrote—based not on his commitment to a particular model of the equity return premium but rather on agnostic uncertainty about the sources of the equity return:

> The data used to document the equity premium over the past 100 years are as good an economic data set as analysts have, and 100 years is long series when it comes to economic data. Before the equity premium is dismissed, not only do researchers need to understand the observed phenomena, but they also need a plausible explanation as to why the future is likely to be any different from the past. In the absence of this explanation, and on the basis of what is currently known, I make the following claim: Over the long term, the equity premium is likely to be similar to what it has been in the past and returns to investment in equity will continue to substantially dominate returns to investment in T-bills for investors with a long planning horizon.

Many Wall Street observers appear to agree that there remains a substantial equity premium. Ivo Welch (2000) surveyed 226 financial economists, asking them to provide their estimates of the future equity premium. Their consensus was that stocks will outperform bills by 6-7% per year for the next ten to thirty years. Gram and Harvey (2007) surveyed nonfinancial corporations' Chief Financial Officers (CFOs). Their 7,316 responses produce an expected annual equity premium of 3.2% per year. There appears to be no compelling reason why CFOs' expectations should be biased in one direction or another.

The modern finance literature on the equity premium puzzle is now more than two decades old. The historical investment literature looking back into observers' pasts and noting the existence of a very large equity return premium is now more than eight decades old. Yet to date no critical mass of long-term investors has taken large-enough long-enough-run positions to try to profit from the equity return premium to substantially arbitrage it away.

Keynes (1936) proposed an explanation. He believed that the finance practitioner professon selects for financial practitioners who are especially vulnerable to these behavioral-finance biases. He wrote that the craft of managing investments is "intolerably boring and over-exacting to any one who is entirely exempt from the gambling instinct." Thus those who would be able to ignore the short-run risks of equities do not stay in the profession. And for those who do have "the gambling instinct"? "He who has it must pay to this propensity the appropriate toll."

From Keynes's proto-behavioral-finance perspective, our collective failure to date to build institutions that will curb psychological propensities for long-run investors to overweight the short-run risks of equity investments is not a thing of the past that the finance practitioners can learn was a mistake and adjust for, but rather a sign that the equity premium return is here for a long run to come.

It would, however, be surprising if the equity premium were as large today as it has been over the past century. The memory of the Great Depression *has* faded. Institutional changes like ERISA *have* removed constraints on investing in equities. Private equity *does* lock investors' money away and so rescues it from the propensity to churn. Individual investors who control their own retirement planning through defined-contribution pension plans *do* find it easier to invest in equities, and the rise in mutual funds *has* in theory made it easier to achieve the benefits of diversification—even if a look at the spread of mutual fund returns shows that the typical mutual fund carries an astonishing amount of idiosyncratic risk.

It would be astonishing if these institutional developments had no effect on the equity return premium.

Yet if the market can be trusted, the equity premium persists today at a level difficult to account for as compensation for the long-term risks of equity investment. There are powerful expected utility-theoretic arguments that the economy has the risk-bearing capacity to make an appropriate equity return premium for visible long-run risks equal to no more than tenths of a percent per year. The existence of the equity return premium in the past offered long-horizon investors a chance to make very large returns in return for bearing little risk. It appears likely that the current configuration of market prices offers a similar opportunity to long-horizon investors today.

How damaging to the economy is this market failure to mobilize its risk-bearing capacity and drive the equity premium down by orders of magnitude? If the failure makes the cost of capital higher because capital ownership involves risk, then the throwing-away of the economy's risk-bearing capacity implies that the economy's capital-output ratio is likely to be much too low. Institutional changes that mobilized some of this absent risk-bearing capacity would then promise enormous dividends. But there is another possibility: perhaps we economists have not an equity return premium but instead a debt return discount puzzle. Firms must then overpay for equity only to the extent that investors overpay for debt. In this case the distortions created are more subtle ones of organizational form—a disfavoring of equity and a favoring of debt-heavy modes—and are presumably smaller in magnitude.

A great many agents and institutions in the economy should have a strong interest in profiting from the extremely high value of the equity return premium. There are lots of long-horizon investors who know that they will not need the money they are investing now until twenty or thirty years in the future. Think of parents of newborns looking forward to their children's college, the middle-aged looking at rapidly-escalating health-care costs, the elderly looking forward to bequeathing some of their wealth, workers with defined-contribution pensions, businesses with defined-benefit pensions, life insurance companies, governments facing an aging population, the rapidly-growing exchange reserve accounts of the world's central banks. On the other side of the market, there are companies that appear underleveraged: replacing high-priced equity capital with low-priced debt capital would seem to be as profitable a strategy for a long-lived company as investing in high-return equity rather than low-return debt is for a long-term investor.

It is understandable that some of these groups chose the aggregate debt-heavy portfolios that they must have done in order to generate the equity return premiums observed over the past century. We economists can build models about principal-agent problems in financial institutions that make portfolio managers seek trades that have high payoffs in a small fraction of a career rather than a large fraction of a lifetime. We economists can speculate about how imperatives of organizational survival lead managers to be strongly averse to putting themselves in a position where they could be bankrupted by unlikely risks that are unknown to them. And we economists can point to institutions and portfolio managers that do borrow long-term to invest in equities: many leveraged buyouts, junk bonds, private equity partnerships, Warren Buffett's career at Berkshire-Hathaway spent buying up insurance companies and putting their reserves to work buying equities. But does this add up to an explanation?

These considerations suggest a strong case for revisiting issues of financial institution design, in order to give the market a push toward being more willing to invest in equities. Economists need to think about institutions that would make long-run buy-and-hold bets on equities easier and more widespread. Mandatory personal retirement or savings accounts with default investments in equity index funds? Automatical investment of tax refunds into diversified equity funds via personal savings accounts? Investing the Social Security trust fund balance in equities as well?

References

Abel, Andrew. 1990. "Asset Prices under Habit Formation and Catching Up with the Joneses." *American Economic Review* 80:2, pp. 38-42.

Alvagari, S. Rao and Gertler, Mark. 1991. "Asset Returns with Transactions Costs and Uninsured Individual Risk." *Journal of Monetary Economics* 27:3, pp. 311-31.

Barro, Robert. 2005. "Rare Events and the Equity Premium." Cambridge: NBER Working Paper 11310.

Barsky, Robert and DeLong J. Bradford. 1993. "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*.

Benartzi, Shlomo and Thaler Richard H. 1995. "Mypic Loss Aversion and the Equity Premium Puzzle."

Blanchard, Olivier J. 1993. "Movements in the Equity Premium." *Brookings Papers on Economic Activity* 1993:2, pp. 75-138

Brown, Stephen, William Goetzmann, and Stephen Ross. 1995. "Survival." *Journal of Finance* 50:3, pp.853-73

Bush, Georgia. 2007. "Volatility Makes a Comeback." New York: Citi Global Wealth Management Investment Research.

Campbell, John. 2003. "Consumption Based Asset Pricing." In Constantinides, George, Mark Harris, and Rene Stulz, eds. *Handbook of the Economics of Finance*. Amsterdam: North Holland.

Campbell, John and Cochrane, John. 1995. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." Cambridge: NBER Working Paper 4995.

Campbell, John and Shiller, Robert. 1989. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Journal of Financial Studies.* 1, pp. 195-228.

Campbell, John and Tuomo Vuolteenaho. 2004. "Inflation Illusion and Stock Prices." Cambridge: NBER working paper 10263.

Cochrane, John. 1994. "Permanent and Transitory Components of GDP and Stock Prices." *Quarterly Journal of Economics* 109, pp. 241-65.

Cochrane, John. 2006. "The Dog that Did Not Bark: A Defense of Return Predictability."

Cambridge: NBER working paper 12026.

Cohen, Randolph, Polk Christopher, and Tuomo Vuolteenaho, 2005. "Money Illusion in the Stock Market: The Modigliani-Cohn Hypothesis." Cambridge: NBER working paper 11018.

Constantinides, George. 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy* 98:3, pp. 519-543.

Constantinides, George and Duffie, Darrel. 1996. "Asset Pricing with Heterogeneous Consumers." Bonn: University of Bonn Discussion Paper Series A 1992.

Constantinides, George, Donaldson, John, and Mehra, Rajnish. 2002. "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle." *Quarterly Journal of Economics,* 117:1, pp. 269-96.

Dasgupta, Partha. 2007. "Discounting Climate Change." Cambridge: University of Cambridge working paper.

Dimson, Elroy, Marsh, Paul and Staunton, Mike (2006) "The Worldwide Equity Premium: A Smaller Puzzle." Working Paper, London School of Economics

Epstein, Larry and Stanley Zin. 1990. "'First-Order' Risk Aversion and the Equity Premium Puzzle." *Journal of Monetary Economics* 26:3, pp. 387-407.

Fama, Eugene F. and French, Kenneth R. (1988), Permanent and Temporary Components of Stock Prices, Journal of Political Economy, April 1988; 96(2): 246-73

Friedman, Milton and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1857-1960*. Princeton: Princeton University Press.

Graham, John R. and Campbell Harvey. 2007. "The Equity Risk Premium in January 2007: Evidence from the Global CFO Outlook Survey." Durham: Duke University <http://ssrn.com/abstract=959703>

Heaton, John and Deborah Lucas. 1995. "The Importance of Investor Heterogeneity and Financial Market Imperfections for the Behavior of Asset Prices." *Carnegie-Rochester Conference Series on Public Policy*. 42, pp. 1-32.

Hines, James R. 2005. "Corporate Taxation and International Competition." Ann Arbor: University of Michigan.

Huggett, Mark. 1993. "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies." *Journal of Economic Dynamics and Control.* 17:5, pp. 953-69.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. London: Macmillan.

Kocherlakota, Narayana. 1996, “The Equity Premium: It's Still a Puzzle.” *Journal of Economic Literature* 34:1, pp. 42-71.

Lucas, Robert. 1978. “Asset Prices in an Exchange Economy.” *Econometrica* 46:6.

Macurdy, Thomas and John Shoven. 1992. “Stocks, Bonds, and Pension Wealth.” In Wise, David, ed. *Topics in the Economics of Aging.* Chicago: University of Chicago Press, pp. 61-75.

Mankiw, N. Gregory and Stephen Zeldes. 1991. “The Consumption of Stockholders and Nonstockholders.” *Journal of Financial Economics* 29:1, pp. 97-112.

McGrattan, Ellen R. and Edward C. Prescott. 2003. “Average Debt and Equity Returns: Puzzling?” Minneapolis: Federal Reserve Bank of Minneapolis, Research Department Staff Report 313.

Mehra, Rajinish. 2003. “The Equity Premium: Why Is It a Puzzle?” Cambridge: NBER Working Paper 9512.

Mehra, Rajinish and Edward C. Prescott. 1985. "The Equity Premium: a Puzzle." *Journal of Monetary Economics* 15:2, pp. 145-162.

Mehra, Rajinish and Edward C. Prescott. 2003. "The Equity Premium in Retrospect." Cambridge: NBER Working Paper 9525.

Modigliani, Franco and Richard A. Cohn. 1979. "Money Illusion in the Stock Market." *Financial Analysts Journal* pp. 3-23.

Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2004. "New Forecasts of the Equity Premium." Cambridge: NBER working paper 10406.

Rietz, Thomas. 1998. "The Equity Risk Premium: A Solution." *Journal of Monetary Economics.*

Samuelson, Paul. 1969. "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Review of Economics and Statistics* 51(3), pp. 239-46.

Shiller, Robert. 2006. Updated Data Set from Chapter 26 of Irrational Exuberance <http://www.econ.yale.edu/~shiller/data.htm>.

Siegel, Jeremy. 2007. "The Future for Investors" Yahoo Finance.

Siegel, Jeremy. 2002. Stocks for the Long Run, Second ed. New York: McGraw Hill.

Smith, Edgar L. 1924. Common Stocks as Long-Term Investments. New York: Macmillan.

Summers, Lawrence H. 1983. “The Nonadjustment of Nominal Interest Rates: A Study of the Fisher Effect.” In James Tobin, ed., *A Symposium in Honor of Arthur Okun*. Washington: Brookings Institution, pp. 201-241.

Vissing-Jorgenson, Annette. 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* 110:41, pp. 825-837.

Weil, Philippe. 1989. “The Equity Premium Puzzle and the Risk-Free Rate Puzzle.” *Journal of Monetary Economics* 24:3, pp.401-21.

Weitzman, Martin L. 2005. “A Unified Bayesian Theory of Equity ‘Puzzles’,”

Welch, Ivo. 2000. “Views of Financial Economists on the Equity Premium and on Professional Controversies.” *Journal of Business* 73:4, pp. 501-537.

Title: Equity Risk Premium: Expectations Great and Small

Authors: Richard A. Derrig and Elisha D. Orr

Contact: Richard A. Derrig
Senior Vice President
Automobile Insurers Bureau of Massachusetts
101 Arch Street
Boston, MA 02110
617-439-4542 x718
617-439-6789 (fax)
richard@aib.org

Version 3.0
August 28, 2003

Disclaimer: The views expressed in this paper are solely the responsibility of the authors, and do not necessarily represent the views of the AIB or of any member of AIB.

Equity Risk Premium: Expectations Great and Small

What I actually think is that our prey, called the equity risk premium, is extremely elusive.

Stephen A. Ross 2001

Abstract:
The Equity Risk Premium (ERP) is an essential building block of the market value of risk. In theory, the collective action of all investors results in an equilibrium expectation for the return on the market portfolio excess of the risk-free return, the equity risk premium. The ability of the valuation actuary to choose a sensible value for the ERP, whether as a required input to CAPM valuation, or any of its descendants, is as important as choosing risk-free rates and risk relatives (betas) to the ERP for the asset at hand. The historical realized ERP for the stock market appears to be at odds with pricing theory parameters for risk aversion. Since 1985, there has been a constant stream of research, each of which reviews theories of estimating market returns, examines historical data periods, or both. Those ERP value estimates vary widely from about minus one percent to about nine percent, based on a geometric or arithmetic averaging, short or long horizons, short or long-run expectations, unconditional or conditional distributions, domestic or international data, data periods, and real or nominal returns. This paper will examine the principal strains of the recent research on the ERP and catalogue the empirical values of the ERP implied by that research. In addition, the paper will supply several time series analyses of the standard Ibbotson Associates 1926-2002 ERP data using short Treasuries for the risk-free rate. Recommendations for ERP values to use in common actuarial valuation problems will be offered.

Acknowledgments
The authors acknowledge the helpful comments from participants in the 2003 Bowles Symposium, Louise Francis, Francis Analytics & Actuarial Data Mining Inc., and four anonymous referees. The authors would also like to thank Jack Wilson for supplying his data series from Wilson & Jones (2002).

Keywords: Equity Risk Premium, Risk Premium Puzzle, Market Return Models, CAPM, Dividend Growth Models, Actuarial Valuations.

Introduction

The Equity Risk Premium (ERP) is an essential building block of the market value of risk. In theory, the collective action of all investors results in an equilibrium expectation for the return on the market portfolio excess of the risk-free return, the equity risk premium. The ability of the valuation actuary to choose a sensible value for the ERP, whether as a required input to CAPM valuation, or any of its descendants[1], is as important as choosing risk-free rates and risk relatives (betas) to the ERP for the asset at hand. Risky discount rates, asset allocation models, and project costs of capital are common actuarial uses of ERP as a benchmark rate.

The equity risk premium should be of particular interest to actuaries. For pensions and annuities backed by bonds and stocks, the actuary needs to have an understanding of the ERP and its variability compared to fixed horizon bonds. Variable products, including Guaranteed Minimum Death Benefits, require accurate projections of returns to ensure adequate future assets. With the latest research producing a relatively low equity risk premium, the rationale for including equities in insurers' asset holdings is being tested. In describing individual investment account guarantees, LaChance and Mitchell (2003) point out an underlying assumption of pension asset investing that, based only on the historical record, future equity returns will continue to outperform bonds; they clarify that those higher expected equity returns come with the additional higher risk of equity returns. Ralfe et al. (2003) support the risky equity view and discuss their pension experience with an all bond portfolio. Recent projections in some literature of a zero or negative equity risk premium challenge the assumptions underlying these views. By reviewing some of the most recent and relevant work on the issue of the equity risk premium, actuaries will have a better understanding of how these values were estimated, critical assumptions that allowed for such a low EPR, and the time period for the projection. Actuaries can then make informed decisions for expected investment results going forward.[2]

In 1985, Mehra and Prescott published their work on the so-called Equity Risk Premium Puzzle: The fact that the historical realized ERP for the stock market 1889-1978 appeared to be at odds with and, relative to Treasury bills, far in excess of asset pricing theory values based on investors with reasonable risk aversion parameters. Since then, there has been a constant stream of research, each of which reviews theories of estimating market returns, examines historical data periods, or both.[3] Those ERP value estimates vary <u>widely</u> from about minus one percent to about nine percent, based on geometric or arithmetic averaging, short or long horizons, short or long-run means, unconditional or conditional expectations, using domestic or international data, differing data periods, and real or nominal returns. Brealey and Myers, in the sixth edition of their standard corporate finance textbook, believe a range of 6% to 8.5% for the US ERP is reasonable for practical project valuation. Is that a fair estimate?

[1] The multifactor arbitrage pricing theory (APT) of Ross (1976), the three-factor model of Fama and French (1992) and the recent Mamaysky (2002) five-factor model for stocks and bonds are all examples of enhanced CAPM models.
[2] See Appendix D
[3] For example, see Cochrane (1997), Cornell (1999), or Leibowitz (2001).

Current research on the equity risk premium is plentiful (Leibowitz, 2001). This paper covers a selection of mainstream articles and books that describe different approaches to estimating the ex ante equity risk premium. We select examples of the research that cover the most important approaches to the ERP. We begin by describing the methodology of using historical returns to predict future estimates. We identify the many varieties of ERPs in order to alert the reader to the fact that numerical estimates of the ERP that appear different may instead be about the same under a common definition. We examine the well-known Ibbotson Associates 1926-2002 data series for stationarity, i.e. time invariance of the mean ERP. We show by several statistical tests that stationarity cannot be rejected and the best estimate going forward, ceteris paribus, is the realized mean. This paper will examine the principal strains of the recent research on the ERP and catalogue the empirical values of the ERP implied by that research.[4]

We first discuss how the Social Security Administration derives estimates of the equity risk premium. Then, we survey the puzzle research, that is, the literature written in response to the Equity Premium Puzzle suggested by Mehra and Prescott (1985). We cover five major approaches from the literature. Next, we report from two surveys of "experts" on the equity risk premium. Finally, after we describe the main strains of research, we explore some of the implications for practicing actuaries.

We do not discuss the important companion problem of estimating the risk relationship of an individual company, line of insurance, or project with the overall market. Within a CAPM or Fama-French framework, the problem is estimating a market beta.[5] Actuaries should be aware, however, that simple 60-month regression betas are biased low where size or non-synchronous trading is a substantial factor (Kaplan and Peterson (1998), Pratt (1998), p86). Adjustments are made to historical betas in order to remove the bias and derive more accurate estimates. Elton and Gruber (1995) explain that by testing the relationship of beta estimates over time, empirical studies have shown that an adjustment toward the mean should be made to project future betas.[6]

The Equity Risk Premium

Based on the definition in Brealey and Myers, *Principles of Corporate Finance* textbook, the equity risk premium (ERP) is the "expected additional return for making a risky investment rather than a safe one". In other words, the ERP is the difference between the market return and a risk-free return. Market returns include both dividends and capital gains. Because both the historical ERP and the prospective ERP have been referred to simply as the equity risk premium, the terms *ex post* and *ex ante* are used to differentiate between them but are often omitted. Table 1 shows the historical annual

[4] The research catalogued appears as Appendix B.

[5] According to CAPM, investors are compensated only for non-diversifiable, or market, risk. The market beta becomes the measurement of the extent to which returns on an individual security covary with the market. The market beta times the ERP represents the non-diversifiable expected return from an individual security.

[6] Elton and Gruber (1995), p148.

average returns from 1926-2002 for large company equities (S&P 500), Treasury Bills and Bonds, and their arithmetic differences using the Ibbotson data (Ibbotson Associates, 2003).[7]

US Equity Risk Premia 1926-2002 Annual Equity Returns and Premia versus Treasury Bills, Intermediate, and Long Term Bonds			
Horizon	**Equity Returns**	**Risk-Free Return**	**ERP**
Short	12.20%	3.83%	8.37%
Intermediate	12.20%	4.81%	7.40%
Long	12.20%	5.23%	6.97%
Source: Ibbotson Yearbook (2003)			

Table 1

In 1985, Mehra and Prescott introduced the idea of the equity risk premium puzzle. The puzzling result is that the historical realized ERP for the stock market using 1889-1978 data appeared to be at odds with and, relative to Treasury bills, far in excess of asset pricing theory values based on normal parametrizations of risk aversion. When using standard frictionless return models and historical growth rates in consumption, the real risk-free rate, and the equity risk premium, the resulting relative risk aversion parameter appears too high. By choosing a maximum relative risk aversion parameter to be 10 and using the growth in consumption, Mehra and Prescott's model produces an ERP much lower than the historical.[8] Their result inspired a stream of finance literature that attempts to solve the puzzle. Two different research threads have emerged. One thread, including behavioral finance, attempts to explain the historical returns with new models and different assumptions about investors.[9] A second thread is from a group that provides estimates of the ERP that are derived from historical data and/or standard economic models. Some in this latter group argue that historical returns may have been higher than those that should be required in the future. In a curiously asymmetric way, there are no serious studies yet concluding that the historical results are too low to serve as ex ante estimates. Although both groups have made substantial and provocative contributions, the behavioral models do not give any ex ante ERP estimates other than explaining and supporting the historical returns. We presume, until results show otherwise, the behavioralists support the historical average as the ex ante unconditional long-run expectation. Therefore, we focus on the latter to catalogue equity risk premium estimates other than the historical approach, but we will discuss both as important strains for puzzle research.

Equity Risk Premium Types

Many different types of equity risk premium estimates can be given even though they are labeled by the same general term. These estimates vary widely; currently the estimates range from about nine percent to a small negative. When ERP estimates are

[7] Ibbotson's 1926-2002 series from the 2003 Yearbook, Valuation Edition. The entire series is shown in Appendix A.

[8] Campbell, Lo, and MacKinlay (1997) perform a similar analysis as Mehra and Prescott and find a risk-aversion coefficient of 19, larger than the reasonable level suggested in Mehra and Prescott's paper, pp307-308.

[9] See, for example, Benartzi and Thaler (1995) and Mehra (2002).

given, one should determine the type before comparing to other estimates. We point out seven important types to look for when given an ERP estimate. They include:

- Geometric vs. arithmetic averaging
- Short vs. long investment horizon
- Short vs. long-run expectation
- Unconditional vs. conditional on some related variable
- Domestic US vs. international market data
- Data sources and periods
- Real vs. nominal returns

The average market return and ERP can be stated as a geometric or arithmetic mean return. An arithmetic mean return is a simple average of a series of returns. The geometric mean return is the compound rate of return; it is a measure of the actual average performance of a portfolio over a given time period. Arithmetic returns are the same or higher than geometric returns, so it is not appropriate to make a direct comparison between an arithmetic estimate and a geometric estimate. However, those two returns can be transformed one to the other. For example, arithmetic returns can be approximated from geometric returns by the formula.[10]

$$AR = GR + \frac{s^2}{2}, s^2 \text{ the variance of the (arithmetic) return process}$$

Arithmetic averages of periodic returns are to be preferred when estimating next period returns since they, not geometric averages, reproduce the proper probabilities and means of expected returns.[11] ERPs can be generated by arithmetic differences (Equity – Risk Free) or by geometric differences ([(1 + Equity)/(1 + Risk Free)]-1). Usually, the arithmetic and geometric differences produce similar estimates.[12]

A second important difference in ERP estimate types is the horizon. The horizon indicates the total investment or planning period under consideration. For estimation purposes, the horizon relates to the term or maturity of the risk-free instrument that is used to determine the ERP.[13] The Ibbotson Yearbook (2003) provides definitions for three different horizons.[14] The short-horizon expected ERP[15] is defined as "the large company stock total returns minus U.S. Treasury bill *total* returns". Note, the income return and total return are the same for U.S. Treasury bills. The intermediate-horizon expected ERP is "the large company stock total returns minus intermediate-term government bond *income* returns". Finally, the long-horizon expected ERP is "the large company stock total returns minus long-term government bond *income* returns". For the Ibbotson data, Treasury bills have a maturity of approximately one month; intermediate-term government bonds have a maturity around five years; long-term government bonds

[10] See Welch (2000), Dimson et al. (2002), Ibbotson and Chen (2003).
[11] For example, see Ibbotson Yearbook, Valuation Edition (2003), pp71-73 for a complete discussion of the arithmetic/geometric choice. See also Dimson et al. (2000), p35 and Brennan and Schwartz (1985).
[12] The arithmetic difference is the geometric difference multiplied by 1 + Risk Free.
[13] See Table 1.
[14] See Ibbotson 2003 Yearbook, p177.
[15] Table 1 displays the short horizon ERP calculation for the 1926-2002 Ibbotson Data.

have a maturity of about 20 years. Although the Ibbotson definitions may not apply to other research, we will classify equity risk premium estimates based on these guidelines to establish some consistency among the current research. The reader should note that Ibbotson Associates recommends the income return (or the yield) when using a bond as the risk free rate rather than the total return.[16]

A third type is the length of time of the equity risk premium forecast. We distinguish between short-run and long-run expectations. Short-run expectations refer to the current equity risk premium, or for this paper, a prediction of up to ten years. In contrast, the long-run expectation is a forecast over ten years to as much as seventy-five years for social security purposes. Ten years appears an appropriate breaking point based on the current literature surveyed.

The next difference is whether the equity risk premium estimate is unconditional or conditioned on one or more related variables. In defining this type, we refer to an admonition by Constantinides (2002, p1568) of the differences in these estimates:

> "First, I draw a sharp distinction between *conditional, short-term forecasts* of the mean equity return and premium and *estimates of the unconditional mean*. I argue that the currently low conditional short-term forecasts of the return and premium do not lessen the burden on economic theory to explain the large unconditional mean equity return and premium, as measured by their sample average over the past one hundred and thirty years."

Many of the estimates we catalogue below will be conditional ones, conditional on dividend yield, expected earnings, capital gains, or other assumptions about the future.

ERP estimates can also exhibit a US versus international market type depending upon the data used for estimation purposes and the ERP being estimated. Dimson, et al. (2002) notes that at the start of 2000, the US equity market, while dominant, was slightly less than one-half (46.1%) of the total international market for equities, capitalized at 52.7 trillion dollars. Data from the non-US equity markets are clearly different from US markets and, hence, will produce different estimates for returns and ERP.[17] Results for the entire world equity market will, of course, be a weighted average of the US and non-US estimates.

[16] The reason for this is two-fold. First, when issued, the yield is the expected market return for the entire horizon of the bond. No net capital gains are expected for the market return for the entire horizon of the bond. No capital gains are expected at the default-free maturity. Second, historical annual capital gains on long-term Government Bonds average near zero (0.4%) over the 1926-2002 period (Ibbotson Yearbook, 2003, Table 6-7).

[17] One qualitative difference can arise from the collapse of equity markets during war time.

Worldwide Equity Risk Premia, 1900-2000 Annual Equity Risk Premium Relative to Treasury Bills		
Country	**Geometric Mean**	**Arithmetic Mean**
United States	5.8%	7.7%
World	4.9%	6.2%
Source: Dimson, et al. (2002), pages 166-167		

Table 2

The next type is the data source and period used for the market and ERP estimates. Whether given an historical average of the equity risk premium or an estimate from a model using various historical data, the ERP estimate will be influenced by the length, timing, and source of the underlying data used. The time series compilations are primarily annual or monthly returns. Occasionally, daily returns are analyzed, but not for the purpose of estimating an ERP. Some researchers use as much as 200 years of history; the Ibbotson data currently uses S&P 500 returns from 1926 to the present.[18] As an example, Siegel (2002) examines a series of real US returns beginning in 1802.[19] Siegel uses three sources to obtain the data. For the first period, 1802 to 1870, characterized by stocks of financial organizations involved in banking and insurance, he cites Schwert (1990). The second period, 1871-1925, incorporates Cowles stock indexes compiled in Shiller (1989). The last period, beginning in 1926, uses CRSP data; these are the same data underlying Ibbotson Associates calculations.

Goetzmann et al. (2001) construct a NYSE data series for 1815 to 1925 to add to the 1926-1999 Ibbotson series. They conclude that the pre-1926 and post-1926 data periods show differences in both risk and reward characteristics. They highlight the fact that inclusion of pre-1926 data will generally produce lower estimates of ERPs than relying exclusively on the Ibbotson post-1926 data, similar to that shown in Appendix A. Several studies that rely on pre-1926 data, catalogued in Appendix B, show the magnitudes of these lower estimates.[20] Table 3 displays Siegel's ERPs for three subperiods. He notes that subperiod III, 1926-2001, shows a larger ERP (4.7%), or a smaller real risk-free mean (2.2%), than the prior subperiods[21].

---

[18] For the Ibbotson analysis of the small stock premium, the NYSE/AMEX/NASDAQ combined data are used with the S&P 500 data falling within deciles 1 and 3 (Ibbotson 2002 Yearbook, pp122-136.)

[19] A more recent alternative is Wilson and Jones (2002) as cited by Dimson et al. (2002), p39.

[20] Using Wilson and Jones' 1871-2002 data series, time series analyses show no significant ERP difference between the 1871-1925 period and the 1926-2002 period; one cannot distinguish the old from the new. The overall average is lower with the additional 1871-1925 data, but on a statistical basis, they are not significantly different. Assuming the equivalency of the two data series for 1871 to 1925 (series of Goetzmann et al. and Wilson & Jones), the risk difference found by Goetzmann et al. must be determined by a significantly different ERP in the pre-1871 data. The 1871-1913 return is prior to personal income tax and appears to be about 35% lower than the 1926-2002 period average of 11.8%, might reflect a zero valuation for income taxes in the pre-1914 returns. Adjusting the pre-1914 data for taxes would most likely make the ERP for the entire period (1871-2002) approximately equal to 7.5%, the 1926-2002 average.

[21] The low risk-free return is indicative of the "risk-free rate puzzle", the twin of the ERP puzzle. For details see Weil (1989).

Short-Horizon Equity Risk Premium by Subperiods			
	Subperiod I	**Subperiod II**	**Subperiod III**
	1802-1870	**1871-1925**	**1926-2001**
Real Geometric Stock Returns	7.0%	6.6%	6.9%
Real Geometric Long Term Governments	4.8%	3.7%	2.2%
Equity Risk Premium	2.2%	2.9%	4.7%
Source: Siegel (2002), pages 13 and 15.			

Table 3

Smaller subperiods will show much larger variations in equity, bill and ERP returns. Table 4 displays the Ibbotson returns and short horizon risk premia for subperiods as small as 5 years. The scatter of results is indicative of the underlying large variation (20% sd) in annual data.

Average Short-Horizon Risk Premium over Various Time Period				
		Common Stocks	U. S. Treasury Bills	Short-Horizon
Year		Total Annual Returns	Total Annual Returns	Risk Premium
All Data	1926-2002	12.20%	3.83%	8.37%
50 Year	1953-2002	12.50%	5.33%	7.17%
40 Year	1963-2002	11.80%	6.11%	5.68%
30 Year	1943-1972	14.55%	2.54%	12.02%
	1973-2002	12.21%	6.61%	5.60%
15 Year	1928-1942	5.84%	0.95%	4.89%
	1943-1957	17.14%	1.20%	15.94%
	1958-1972	11.96%	3.87%	8.09%
	1973-1987	11.42%	8.20%	3.22%
	1988-2002	13.00%	5.03%	7.97%
10 Year	1933-1942	12.88%	0.15%	12.73%
	1943-1952	17.81%	0.81%	17.00%
	1953-1962	15.29%	2.19%	13.11%
	1963-1972	10.55%	4.61%	5.94%
	1973-1982	8.67%	8.50%	0.17%
	1983-1992	16.80%	6.96%	9.84%
	1993-2002	11.17%	4.38%	6.79%
5 Year	1928-1932	- 8.25%	2.55%	-10.80%
	1933-1937	19.82%	0.22%	19.60%
	1938-1942	5.94%	0.07%	5.87%
	1943-1947	15.95%	0.37%	15.57%
	1948-1952	19.68%	1.25%	18.43%
	1953-1957	15.79%	1.97%	13.82%
	1958-1962	14.79%	2.40%	12.39%
	1963-1967	13.13%	3.91%	9.22%
	1968-1972	7.97%	5.31%	2.66%
	1973-1977	2.55%	6.19%	- 3.64%
	1978-1982	14.78%	10.81%	3.97%
	1983-1987	16.93%	7.60%	9.33%
	1988-1992	16.67%	6.33%	10.34%
	1993-1997	21.03%	4.57%	16.46%
	1998-2002	1.31%	4.18%	- 2.88%

Table 4

In calculating an expected market risk premium by averaging historical data, projecting historical data using growth models, or even conducting a survey, one must determine a proxy for the “market”. Common proxies for the US market include the S&P 500, the NYSE index, and the NYSE, AMEX, and NASDAQ index.[22] For the purpose of this paper, we use the S&P 500 and its antecedents as the market. However, in the various research surveyed, many different market proxies are assumed. We have already discussed using international versus domestic data when describing different MRP types. With international data, different proxies for other country, region, or world markets are used.[23] For domestic data, different proxies have been used over time as stock market exchanges have expanded.[24] Fortunately, as shown in the Ibbotson Valuation yearbook, the issue of a US market proxy does not have a large effect on the MRP estimate because the various indices are highly correlated. For example, the S&P 500 and the NYSE have a correlation of 0.95, the S&P 500 and NYSE/AMEX/NASDAQ 0.97, and the NYSE and NYSE/AMEX/NASDAQ 0.90.[25] Therefore, the market proxy selected is one reason for slight differences in the estimates of the market risk premium.

As a final note, stock returns and risk-free rates can be stated in nominal or real terms. Nominal includes inflation; real removes inflation. The equity risk premium should not be affected by inflation because either the stock return and risk-free rate both include the effects of inflation (both stated in nominal terms) or neither have inflation (both stated in real terms). If both returns are nominal, the difference in the returns is generally assumed to remove inflation. Otherwise, both terms are real, so inflation is removed prior to finding the equity risk premium. While numerical differences in the real and nominal approaches may exist, their magnitudes are expected to be small.

Equity Risk Premia 1926-2002

As an example of the importance of knowing the types of equity risk premium estimates under consideration, Table 5 displays ERP returns that each use the same historical data, but are based on arithmetic or geometric returns and the type of horizon. The ERP estimates are quite different.[26]

[22] 2003 Ibbotson Valuation Yearbook, p92.
[23] For example, Dimson (2002) and Claus and Thomas (2001) use international market data.
[24] For a data series that is a mixture of the NYSE exchange, NYSE, AMEX, and NASDAQ stock exchange, and the Wilshire 5000, see Dimson (2002), p306.
[25] 2003 Ibbotson Valuation Yearbook, p93; using data from October 1997 to September 2002.
[26] The nominal and real ERPs are identical in Table 5 because the ERPs are calculated as arithmetic differences, and the same value of inflation will reduce the market return and the risk-free return equally. Geometric differences would produce minimally different estimates for the same types.

ERP using same historical data (1926-2002)		
RFR Description	**ERP Description**	**ERP Historical Return**
Short nominal	Arithmetic Short-horizon	8.4%
Short nominal	Geometric Short-horizon	6.4%
Short real	Arithmetic Short-horizon	8.4%
Short real	Geometric Short-horizon	6.4%
Intermediate nominal	Arithmetic Inter-horizon	7.4%
Intermediate nominal	Geometric Inter-horizon	5.4%
Intermediate real	Arithmetic Inter-horizon	7.4%
Intermediate real	Geometric Inter-horizon	5.4%
Long nominal	Arithmetic Long-horizon	7.0%
Long nominal	Geometric Long-horizon	5.0%
Long real	Arithmetic Long-horizon	7.0%
Long real	Geometric Long-horizon	5.0%

Table 5

Historical Methods

The historical methodology uses averages of past returns to forecast future returns. Different time periods may be selected, but the two most common periods arise from data provided by either Ibbotson or Siegel. The Ibbotson series begins in 1926 and is updated each year. The Siegel series begins in 1802 with the most recent compilation using returns through 2001. Appendix A provides equity risk premium estimates using Ibbotson data for the 1926-2002 period that we use in this paper for most illustrations. We begin with a look at the ERP history through a time series analysis of the Ibbotson data.

Time Series Analysis

Much of the analysis addressing the equity risk premium puzzle relies on the annual time series of market, risk-free and risk premium returns. Two opposite views can be taken of these data. One view would have the 1926-2002 Ibbotson data, or the 1802-2001 Siegel data, represent one data point; i.e., we have observed one path for the ERP through time from the many possible 77 or 200 year paths. This view rests upon the existence or assumption of a stochastic process with (possibly) inter-temporal correlations. While mathematically sophisticated, this model is particularly unhelpful without some testable hint at the details of the generating stochastic process. The practical view is that the observed returns are random samples from annual distributions that are iid, independent and identically distributed about the mean. The obvious advantage is that we have at hand 77 or 200 observations on the iid process to analyze. We adopt the latter view.

Some analyses adopt the assumption of stationarity of ERP, i.e., the true mean does not change with time. Figure 1 displays the Ibbotson ERP data and highlights two subperiods, 1926-1959 and 1960-2002.[27] While the mean ERP for the two subperiods appear quite different (11.82% vs. 5.27%), the large variance of the process (std dev 20.24%) should make them indistinguishable statistically speaking.

[27] The ERP shown here are the geometric differences (calculated) rather than the simple arithmetic differences in Table 1; i.e. ERP = $[(1+r_m)/(1+r_f)] - 1$. The test results are qualitatively the same for the arithmetic differences.

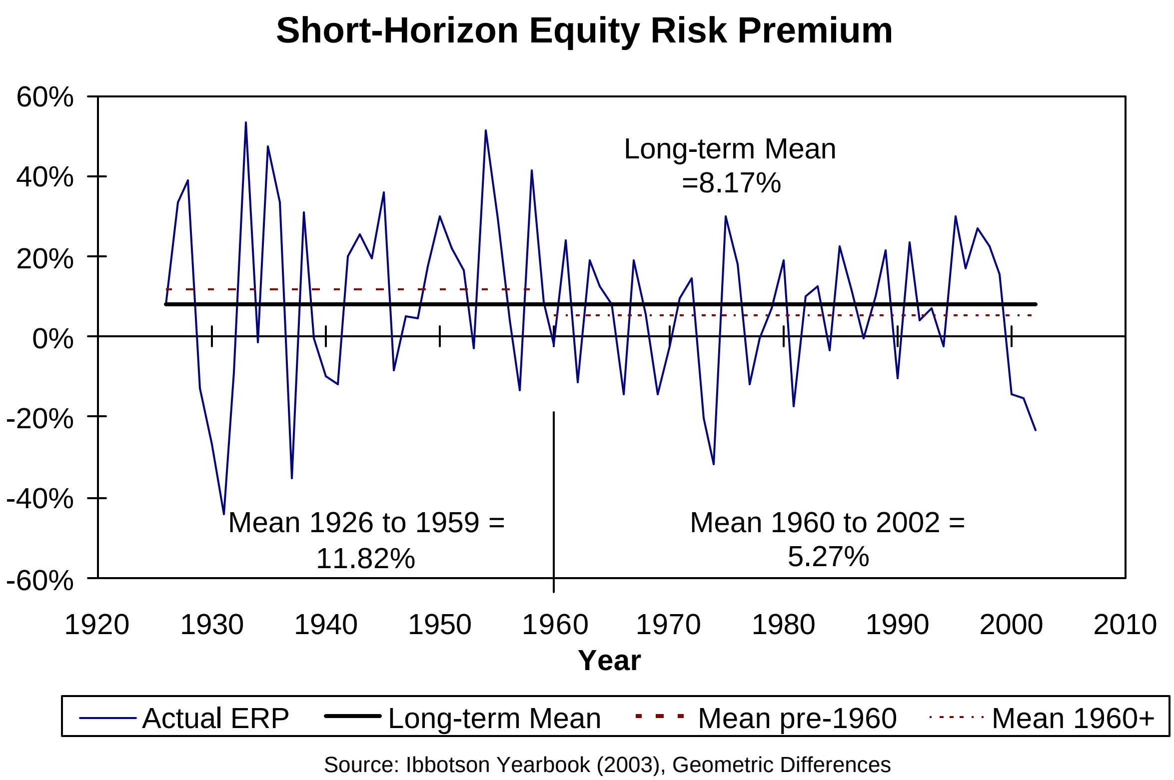


Figure 1

T-Tests

The standard T-test can be used for the null hypothesis H_o : mean 1960-2002 = 8.17%, the 77 year mean.[28] The outcome of the test is shown in Table 6; the null hypothesis cannot be rejected.

T-Test Under the Null Hypothesis that ERP (1960-2002) = ERP (1926-2002) = 8.17%	
Sample mean 1960-2002	5.27%
Sample s.d. 1960-2002	15.83%
T value (DF=42)	-1.20
PR > |T|	0.2374
Confidence Interval 95%	(0.0040, 0.1014)
Confidence Interval 90%	(0.0121, 0.0933)

Table 6

Another T-Test can be used to test whether the subperiod means are different in the presence of unequal variances.[29] The result is similar to Table 6 and the difference of subperiod means equal to zero cannot be rejected.[30]

[28] Standard statistical procedures in SAS 8.1 have been used for all tests.
[29] Equality of variances is rejected at the one percent level by an F test (F=2.39, DF=33,42)
[30] t-value 1.35, PR> |T| = 0.1850 with the Cochran method.

Time Trends

The supposition of stationarity of the ERP series can be supported by ANOVA regressions. The results of regressing the ERP series on time is shown in Table 7.

ERP ANOVA Regressions on Time		
Period	**Time Coefficient**	**P-Value**
1926-1959	0.004	0.355
1960-2002	0.001	0.749
1926-2002	-0.001	0.443

Table 7

There are no significant time trends in the Ibbotson ERP data.[31]

ARIMA Model

Time series analysis using the well established Box-Jenkins approach can be used to predict future series values through the lag correlation structure.[32] The SAS ARIMA procedure applied to the full 77 time series data shows:

(1) No significant autocorrelation lags.
(2) An identification of the series as white noise.
(3) ARIMA projection of year 78+ ERP is 8.17%, the 77 year average.

All of the above single time series tests point to the reasonability of the stationarity assumption for (at least) the Ibbotson ERP 77 year series.[33]

Social Security Administration

In the current debate on whether to allow private accounts that may invest in equities, the Office of the Chief Actuary of the Social Security Administration has selected certain assumptions to assess various proposals (Goss, 2001). The relevant selection is to use 7 percent as the real (geometric) annual rate of return for equities.[34] This assumption is based on the historical return of the 20th century. SSA received further support that showed the historical return for the last 200 years is consistent with this estimate, along with the Ibbotson series beginning in 1926. For SSA, the calculation of the equity risk premium uses a long-run real yield on Treasury bonds as the risk-free rate. From the assumptions in the 1995 Trustees Report, the long-run real yield on Treasury bonds that the Advisory Council proposals use is 2.3%. Using a future Treasury securities real yield of 2.3% produces a geometric equity risk premium of 4.7% over long-term Treasury securities. More recently, the Treasury securities assumption has increased to 3%[35], yielding a 4% geometric ERP over long-term Treasury securities.

---

[31] The result is confirmed by a separate Chow test on the two subperiods.
[32] See Harvey (1990), p30.
[33] The same tests applied to the Wilson and Jones 1871-2002 data series show similar results: Neither the 1871-1925 period nor the 1926-2002 period is different from the overall 1871-2002 period. The overall period and subperiods also show no trends over time.
[34] Compare Table 3, subperiod III.
[35] 1999 Social Security Trustees Report.

At the request of the Office of the Chief Actuary of the Social Security Administration (OCACT), John Campbell, Peter Diamond, and John Shoven were engaged to give their expert opinions on the assumptions Social Security mode. Each economist begins with the Social Security assumptions and then explains any difference he feels would be more appropriate.

In John Campbell's response, he considers valuation ratios as a comparison to the returns from the historical approach (Campbell 2001). The current valuation ratios are at unusual levels, with a low dividend-price ratio and high price-earnings ratio. He reasons that the prices are what have dramatically changed these ratios. Campbell presents two views as to the effect of valuation ratios in their current state. One view is that valuations will remain at the current level, suggesting much lower expected returns. The second view is a correction to the ratios, resulting in less favorable returns until the ratios readjust. He decides to give some weight to both possibilities, so he lowers the geometric equity return estimate to 5-5.5% from 7%. For the risk-free rate, he uses the yield on the long-term inflation-indexed bonds[36] of 3.5% or the OCACT assumption of 3%. Therefore, his geometric equity premium estimate is around 1.5 to 2.5%.

Peter Diamond uses the Gordon growth formula to calculate an estimate of the equity return (Diamond 2001). The classic Gordon Dividend Growth model is[37]:

$$K = (D_1 / P_0) + g$$

K = Expected Return or Discount Rate $\quad P_0$ = Price this period
D_1 = Expected Dividend next period $\quad$ g = Expected growth in dividends in perpetuity

Based on his analysis, he feels that the equity return assumption of 7% for the next 75 years is not consistent with a reasonable level of stock value compared to GDP. Even when increasing the GDP growth assumption, he still does not feel that the equity return is plausible. By reasoning that the next decade of returns will be lower than normal, only then is the equity return beyond that time frame consistent with the historical return. By considering the next 75 years together, he would lower the overall projected equity return to 6-6.5%. He argues that the stock market is overvalued, and a correction is required before the long-run historical return is a reasonable projection for the future. By using the OCACT assumption of 3.0% for the long-term real yield on Treasury bonds, Diamond estimates a geometric equity risk premium of about 3-3.5%.

John Shoven begins by explaining why the traditional Gordon growth model is not appropriate, and he suggests a modernized Gordon model that allows share repurchases to be included instead of only using the dividend yield and growth rate (Shoven 2001). By assuming a long-term price-earnings ratio between its current and historical value, he comes up with an estimate for the long-term real equity return of 6.125%. Using his general estimate of 6-6.5% for the equity return and the OCACT assumptions for the long-term bond yield, he projects a long-term equity risk premium of approximately 3-3.5%. All the SSA experts begin by accepting the long-run historical

[36] See discussion of current yields on TIPS below.
[37] Brealey and Myers (2000), p67.

ERP analyses and then modifying that by changes in the risk-free rate or by decreases in the long-term ERP based on their own personal assessments. We now turn to the major strains in ERP puzzle research.

ERP Puzzle Research

Campbell and Shiller (2001) begin with the assumption of mean reversion of dividend/price and price/earnings ratios. Next, they explain the result of prior research which finds that the dividend-price ratio predicts future prices, and historically, the price corrects the ratio when it diverts from the mean.[38] Based on this result, they then use regressions of the dividend-price ratio and the price-smoothed-earnings[39] ratio to predict future stock prices out ten years. Both regressions predict large losses in stock prices for the ten year horizon. Although Campbell and Shiller do not rerun the regression on the dividend-price ratio to incorporate share repurchases, they point out that the dividend-price ratio should be upwardly adjusted, but the adjustment only moves the ratio to the lower range of the historical fluctuations (as opposed to the mean). They conclude that the valuation ratios indicate a bear market in the near future[40]. They predict for the next ten year period negative real stock returns. They caution that because valuation ratios have changed so much from their normal level, they may not completely revert to the historical mean, but this does not change their pessimism about the next decade of stock market returns.

Arnott and Ryan (2001) take the perspective of fiduciaries, such as pension fund managers, with an investment portfolio. They begin by breaking down the historical stock returns (past 74 years since December 1925) by analyzing dividend yields and real dividend growth. They point out that the historical dividend yield is much higher than the current dividend yield of about 1.2%. They argue that the changes from stock repurchases, reinvestment, and mergers and acquisitions, which affect the lower dividend yield, can be represented by a higher dividend growth rate. However, they cap real dividend or earnings growth at the level of real economic growth. They add the dividend yield and the growth in real dividends to come up with an estimate for the future equity return; the current dividend yield of 1.2% and the economic growth rate of 2.0% add to the 3.2% estimated real stock return. This method corresponds to the dividend growth model or earnings growth model and does not take into account changing valuation levels. They cite a TIPS yield of 4.1% for the real risk-free rate return.[41] These two estimates yield a negative geometric long-horizon conditional equity risk premium.

Arnott and Bernstein (2002) begin by arguing that in 1926 investors were not expecting the realized, historical compensation that they later received from stocks. They cite bonds' reaction to inflation, increasing valuations, survivorship bias[42], and changes in

---

[38] Campbell and Shiller (1989).
[39] Earnings are "smoothed" by using ten year averages.
[40] The stock market correction from year-end 1999 to year-end 2002 is a decrease of 37.6% or 14.6% per year. Presumably, the "next ten years" refers to 2000 to 2010.
[41] See the current TIPS yield discussion near end of paper.
[42] See Brown et al. (1992, 1995) for details on potential survivorship bias.

regulation as positive events that helped investors during this period. They only use the dividend growth model to predict a future expected return for investors. They do not agree that the earnings growth model is better than the dividend growth model both because earnings are reported using accounting methods and earnings data before 1870 are inaccurate. Even if the earnings growth model is chosen instead, they find that the earnings growth rate from 1870 only grows 0.3% faster than dividends, so their results would not change much. Because of the Modigliani-Miller theorem[43], a change in dividend policy should not change the value of the firm. They conclude that managers benefited in the "era of 'robber baron' capitalism" instead of the conclusion reached by others that the dividend growth model under-represents the value of the firm.

By holding valuations constant and using the dividend yield and real growth of dividends, Arnott and Bernstein calculate the equity return that an investor might have expected during the historical time period starting in 1802. They use an expected dividend yield of 5.0%, close to the historical average of 1810 to 2001. For the real growth of dividends, they choose the real per capita GDP growth less a reduction for entrepreneurial activity in the economy plus stock repurchases. They conclude that the net adjustment is negative, so the real GDP growth is reduced from 2.5-3% to only 1%. A fair expectation of the stock return for the historical period is close to 6.1% by adding 5.0% for the dividend yield and a net real GDP per capita growth of 1.1%. They use a TIPS yield of 3.7% for the real risk-free rate, which yields a geometric intermediate-horizon equity risk premium of 2.4% as a fair expectation for investors in the past. They consider this a "normal" equity risk premium estimate. They also opine that the current ERP is zero; i.e. they expect stocks and (risk-free) bonds to return the same amounts.

Fama and French (2002) use both the dividend growth model and the earnings growth model to investigate three periods of historical returns: 1872 to 2000, 1872 to 1950, and 1951 to 2000. Their ultimate aim is to find an unconditional equity risk premium. They cite that by assuming the dividend-price ratio and the earnings-price ratio follow a mean reversion process, the result follows that the dividend growth model or earnings growth model produce approximations of the unconditional equity return. Fama and French's analysis of the earlier period of 1872 to 1950 shows that the historical average equity return and the estimate from the dividend growth model are about the same. In contrast, they find that the 1951 to 2000 period has different estimates for returns when comparing the historical average and the growth models' estimates. The difference in the historical average and the model estimates for 1951 to 2000 is interpreted to be "unexpected capital gains" over this period. They find that the unadjusted growth model estimates of the ERP, 2.55% from the dividend model and 4.32% from the earnings model, fall short of the realized average excess return for 1951-2000. Fama and French prefer estimates from growth models instead of the historical method because of the lower standard error using the dividend growth model. Fama and French provide 3.83% as the unconditional expected equity risk premium return (referred to as the annual bias-adjusted ERP estimate) using the dividend growth model with underlying data from 1951 to 2000. They give 4.78% as the unconditional expected equity risk

[43] Brealey and Myers (2000), p447. See also discussion in Ibbotson and Chen (2003).

premium return using the earnings growth model with data from 1951 to 2000. Note that using a one-month Treasury bill instead of commercial paper for the risk-free rate would increase the ERP by about 1% to nearly 6% for the 1951-2000 period.

Ibbotson and Chen (2003) examine the historical real geometric long-run market and long risk-free returns using their "building block" methodology.[44] They use the full 1926-2000 Ibbotson Associates data and consider as building blocks all of the fundamental variables of the prior researchers. Those blocks include (not all simultaneously):

- Inflation
- Real risk-free rates (long)
- Real capital gains
- Growth of real earnings per share
- Growth of real dividends
- Growth in payout ratio (dividend/earnings)
- Growth in book value
- Growth in ROE
- Growth in price/earnings ratio
- Growth in real GDP/population
- Growth in equities excess of GDP/POP
- Reinvestment

Their calculations show that a forecast real geometric long run return of 9.4% is a reasonable extrapolation of the historical data underlying a realized 1926-2000 return of 10.7%, yielding a long horizon arithmetic ERP of 6%, or a short horizon arithmetic ERP of about 7.5%.

The authors construct six building block methods; i.e., they use combinations of historic estimates to produce an expected geometric equity return. They highlight the importance of using both dividends and capital gains by invoking the Modigliani-Miller theorem. The methods, and their component building blocks are:

- Method 1: Inflation, real risk free rate, realized ERP
- Method 2: Inflation, income, capital gains and reinvestment
- Method 3: Inflation, income, growth in price/earnings, growth in real earnings per share and reinvestment.
- Method 4: Inflation, growth rate of price/earnings, growth rate of real dividends, growth rate of payout ratio dividend yield and reinvestment
- Method 5: Inflation, income growth rate of price/earnings, growth of real book value, ROE growth and reinvestment
- Method 6: Inflation, income, growth in real GDP/POP, growth in equities excess GDP/POP and reinvestment.

[44] See Appendix D for a summary of their building block estimates. See also Pratt (1998) for a discussion of the Building Block, or Build-Up Model, cost of capital estimation method.

All six methods reproduce the historical long horizon geometric mean of 10.70% as shown in Appendix D. Since the source of most other researchers' lower ERP is the dividend yield, the authors recast the historical results in terms of ex ante forecasts for the next 75 years. Their estimate of 9.37% using supply side methods 3 and 4 is approximately 130 basis points lower than the historical result. Within their methods, they also show how the substantially lower expectation of 5.44% for the long mean geometric return is calculated by omitting one or more relevant variables. Underlying these ex ante methods are the assumptions of stationarity of the mean ERP return and market efficiency, the absence of the assumption that the market has mispriced equities. All of their methods are aimed at producing an unconditioned estimate of the ex ante ERP.

As opposed to short-run, conditional estimates from Campbell and Shiller and others, Constantinides (2002) seeks to estimate the unconditional equity risk premium, more in line with the goal of Fama and French (2002) and Ibbotson and Chen (2003). He begins with the premise that the unconditional ERP can be estimated from the historical average using the assumption that the ERP follows a stationary path. He suggests most of the other research produces conditional estimates, conditioned upon beliefs about the future paths of fundamentals such as dividend growth, price-earnings ratio and the like. While interesting in themselves, they add little to the estimation of the unconditional mean ERP.

Constantinides uses the historical return and adjusts downward by the growth in the price-earnings ratio to calculate the unconditional equity risk premium. He removes the growth in the price-earnings ratio because he is assuming no change in valuations in the unconditional state. He gives estimates using three periods. For 1872-2000, he uses the historical equity risk premium which is 6.9%, and after amortizing the growth in the price-dividend ratio or price-earnings ratio over a period as long as 129 years, the effect of the potential reduction is no change. Therefore, he finds an unconditional arithmetic, short-horizon equity risk premium of 6.9% using the 1872-2000 underlying data. For 1951-2000, he again starts with the historical equity risk premium which is 8.7% and lowers this estimate by the growth in the price-earnings ratio of 2.7% to find an unconditional arithmetic, short-horizon equity risk premium of 6.0%. For 1926-2000, he uses the historical equity risk premium which is 9.3% and reduces this estimate by the growth in the price-earnings ratio of 1.3% to find an unconditional arithmetic, short-horizon equity risk premium of 8.0%. He appeals to behavioral finance to offer explanations for such high unconditional equity risk premium estimates.

From the perspective of giving practical investor advice, Malkiel (1999) discusses "the age of the millennium" to give some indication of what investors might expect for the future. He specifically estimates a reasonable expectation for the first few decades of the twenty-first century. He estimates the future bond returns by giving estimates if bonds are held to maturity with corporate bonds of 6.5-7%, long-term zero-coupon Treasury bonds of about 5.25%, and TIPS with a 3.75% return. Depending on the desired level of risk, Malkiel indicates bondholders should be more favorably compensated in the future compared to the historical returns from 1926 to 1998. Malkiel uses the earnings growth model to predict future equity returns. He uses the current dividend yield of 1.5% and an earnings growth estimate of 6.5%, yielding an 8% equity return estimate compared with an 11% historical return. Malkiel's estimated range of the equity risk premium is from 1% to 4.25%, depending on the risk-free instrument selected. Although his equity risk premium is lower than the historical return, his selection of a relatively high earnings growth rate is similar to Ibbotson and Chen's forecasted models. In contrast with Ibbotson and Chen, Malkiel allows for a changing equity risk premium and advises investors to not rely solely on the past "age of exuberance" as a guide for the future. Malkiel points out the impact of changes in valuation ratios, but he does not attempt to predict future valuation levels.

Finally, Mehra (2002) summarizes the results of the research since the ERP puzzle was posed. The essence of the puzzle is the inconsistency of the ERPs produced by descriptive and prescriptive economic models of asset pricing on the one hand and the historical ERPs realized in the US market on the other. Mehra and Prescott (1985) speculated that the inconsistency could arise from the inadequacy of standard models to incorporate market imperfections and transaction costs. Failure of the models to reflect reality rather than failure of the market to follow the theory seems to be Mehra's conclusion as of 2002. Mehra points to two promising threads of model-modifying research. Campbell and Cochrane (1999) incorporate economic cycles and changing risk aversion while Constantinides et al. (2002) propose a life cycle investing modification, replacing the representative agent by segmenting investors into young, middle aged, and older cohorts. Mehra sums up by offering:

> "Before we dismiss the premium, we not only need to have an understanding of the observed phenomena but also why the future is likely to be different. In the absence of this, we can make the following claim based on what we know. Over the long horizon the equity premium is likely to be similar to what it has been in the past and the returns to investment in equity will continue to substantially dominate those in bonds for investors with a long planning horizon."

Financial Analyst Estimates

Claus and Thomas (2001) and Harris and Marston (2001) both provide equity premium estimates using financial analysts' forecasts. However, their results are rather different. Claus and Thomas use an abnormal earnings model with data from 1985 to 1998 to calculate an equity risk premium as opposed to using the more common dividend growth model. Financial analysts project five year estimates of future earnings growth rates. When using this five year growth rate for the dividend growth rate in perpetuity in the Gordon growth model, Claus and Thomas explain that there is a potential upward bias in estimates for the equity risk premium. Therefore, they choose to use the abnormal earnings model instead and only let earnings grow at the level of inflation after five years. The abnormal earnings model replaces dividends with "abnormal earnings" and discounts each flow separately instead of using a perpetuity. The average estimate that they find is 3.39% for the equity risk premium. Although it is generally recognized that financial analysts' estimates have an upward bias, Claus and Thomas propose that in the current literature, financial analysts' forecasts have underestimated short-term earnings in order for management to achieve earnings estimates in the slower economy. Claus and Thomas conclude that their findings of the ERP using data from the past fifteen years are not in line with historical values.

Harris and Marston use the dividend growth model with data from 1982 to 1998. They assume that the dividend growth rate should correspond to investor expectations. By using financial analysts' longest estimates (five years) of earnings growth in the model, they attempt to estimate these expectations. They argue that if investors are in accord with the optimism shown in analysts' estimates, even biased estimates do not pose a drawback because these market sentiments will be reflected in actual returns. Harris and Marston find an equity risk premium estimate of 7.14%. They find fluctuations in the equity risk premium over time. Because their estimates are close to historical returns, they contend that investors continue to require a high equity risk premium.

Survey Methods

One method to estimate the ex ante equity risk premium is to find the consensus view of experts. John Graham and Campbell Harvey perform a survey of Chief Financial Officers to determine the average cost of capital used by firms. Ivo Welch surveys financial economists to determine the equity risk premium that academic experts in this area would estimate.

Graham and Harvey administer surveys from the second quarter of 2000 to the third quarter of 2002 (Graham and Harvey, 2002). For their survey format, they show the current ten year bond yield and then ask CFOs to provide their estimate of the S&P 500 return for the next year and over the next ten years. CFOs are actively involved in setting a company's individual hurdle[45] rate and are therefore considered knowledgeable about investors' expectations.[46] When comparing the survey responses of the one and ten year returns, the one year returns have so much volatility that they conclude that the ten-year equity risk premium is the more important and appropriate return of the two when making financial decisions such as hurdle rates and estimating cost of capital. The average ten-year equity risk premium estimate varies from 3% to 4.7%.

The most current Welch survey compiles the consensus view of about five hundred financial economists (Welch 2001). The average arithmetic estimate for the 30-year equity risk premium relative to Treasury bills is 5.5%; the one-year arithmetic equity risk premium consensus is 3.4%. Welch deduces from the average 30-year geometric

[45] A "hurdle" rate is a benchmark cost of capital used to evaluate projects to accept (expected returns greater than hurdle rate) or reject (expected returns less than hurdle rate).
[46] Graham and Harvey claim three-fourths of the CFOs use CAPM to estimate hurdle rates.

equity return estimate of 9.1% that the arithmetic equity return forecast is approximately 10%.[47]

Welch's survey question allows the participants to self select into different categories based upon their knowledge of ERP. The results indicate that the responses of the less ERP knowledgeable participants showed more pessimism than those of the self reported experts. The experts gave 30-year estimates that are 30 to 150 basis points above the estimates of the non-expert group.

Differences in Forecasts across Expertise Level				
Relative Expertise	**Statistic**	**Stock Market**	**Equity Premium**	
		30-Year Geometric	30-Year Arithmetic	30-Year Geometric
188 Less Involved	Mean	8.5%	4.9%	4.4%
	Median	8%	5%	4%
	IQ Range	6%-10%	3%-6%	2%-5.5%
235 Average	Mean	9.2%	5.8%	4.8%
	Median	9%	5%	4%
	IQ Range	7.5%-10%	3.5%-7%	3%-6%
72 Experts	Mean	10.1%	6.2%	5.4%
	Median	9%	5.4%	5%
	IQ Range	8%-11%	4%-7.5%	3.4%-6%
Data Source: Welch (2001), Table 5				

Table 8

Table 8 shows that there may be a "lemming" effect, especially among economists who are not directly involved in the ERP question. Stated differently, all the academic and popular press, together with the prior Welch survey[48] could condition the non-expert, the "less involved", that the expected ERP was lower than historic levels.

The Behavioral Approach

Benartzi and Thaler (1995) analyze the equity risk premium puzzle from the point of view of prospect theory (Kahneman and Tversky; 1979). Prospect theory[49] has "loss aversion", the fact that individuals are more sensitive to potential loss than gain, as one of its central tenets. Once an asymmetry in risk aversion is introduced into the model of the rational representative investor or agent, the unusual risk aversion problem raised initially by Mehra and Prescott (1985) can be "explained" within this behavioral model of decision-making under uncertainty. Stated differently, given the historical ERP series, there exists a model of investor behavior that can produce those or similar results. Benartzi and Thaler combine loss aversion with "mental accounting", the behavioral process people use to evaluate their status relative to gains and losses compared to expectations, utility and wealth, to get "myopic loss aversion". In particular, mental

[47] For the Ibbotson 1926-2002 data, the arithmetic return is about 190 basis points higher than the geometric return rather than the inferred 90 basis points. This suggests the participant's beliefs may not be internally consistent.
[48] The prior Welch survey in 1998 had a consensus ERP of about 7%.
[49] A current survey of the applications of prospect theory to finance can be found in Benartzi et al. (2001).

accounting for a portfolio needs to take place infrequently because of loss aversion, in order to reduce the chances of observing loss versus gain. The authors concede that there is a puzzle with the standard expected utility-maximizing paradigm but that the myopic loss aversion view may resolve the puzzle. The authors' views are not free of controversy; any progress along those lines is sure to match the advance of behavioral economics in the large.

The adoption of other behavioral aspects of investing may also provide support for the historical patterns of ERPs we see from 1802-2002. For example, as the true nature of risk and rewards has been uncovered by the virtual army of 20th century researchers, and as institutional investors held sway in the latter fifty years of the century, the demand for higher rewards seen in the later historical data may be a natural and rational response to the new and expanded information set. Dimson et al. (2002, Figure 4-6) displays increasing real US equity returns of 6.7, 7.4, 8.2 and 10.2 for periods of 101, 75, 50 and 25 years ending in 2001 consistent with this "risk-learning" view.

Next Ten Years

The "next ten years" is an issue that experts reviewing Social Security assumptions and Campbell and Shiller address either explicitly or implicitly. Experts evaluating Social Security's proposals predicted that the "next ten years", indicating a period beginning around 2000, of returns were likely to be below the historical return. However, a historical return was recommended as appropriate for the remaining 65 of the 75 years to be projected. For Campbell and Shiller (2001), the period they discuss is approximately 2000-2010. Based on the current state of valuation ratios, they predict lower stock market returns over "the next ten years". These expert predictions, and other pessimistic low estimates, have already come to fruition as market results 2000 through 2002.[50] The US equities market has decreased 37.6% since 1999, or an annual decrease of 14.6%. Although these forecasts have proved to be accurate in the short term, for future long-run projections, the market is not at the same valuation today as it was when these conditional estimates were originally given. Therefore, actuaries should be wary of using the low long-run estimates made prior to the large market correction of 2000-2002.

Treasury Inflation Protection Securities (TIPS)

Several of the ERP researchers refer to TIPS when considering the real risk free rates. Historically, they adjust Treasury yields downward to a real rate by an estimate of inflation, presumably for the term of the Treasury security. As Table 3 shows using the Siegel data, the modern era data show a low real long-term risk-free rate of return (2.2%). This contrasts with the initial[51] TIPS issue yields of 3.375%. Some researchers use those TIPS yields as (market) forecasts of real risk-free returns for intermediate and long-horizon, together with reduced (real) equity returns to produce low estimates of ex ante ERPs. None consider the volatility of TIPS as indicative of the accuracy of their ERP estimate.

[50] The Social Security Advisory Board will revisit the seventy five year rate of return assumption during 2003, Social Security Advisory Board (2002).

[51] TIPS were introduced by the Treasury in 1996 with the first issue in January, 1997.

Table 9 shows a recent market valuation of ten and thirty year TIPS issued in 1998-2002.

Inflation-Indexed Treasury Securities		
Maturity	**Coupon Issue Rate**	**Yield to Maturity**
1/11	3.500	1.763
1/12	3.375	1.831
7/12	3.000	1.878
4/28	3.625	2.498
4/29	3.875	2.490
4/32	3.375	2.408
Source: WSJ 1 2/24/2003		

Table 9

Note the large 90-180 basis point decrease in the current "real" yields from the issue yields as recent as ten months ago. While there can be several explanations for the change (revaluation of the inflation option, flight to Treasury quality, paucity of 30 year Treasuries), the use of these current "real" risk free yields, with fixed expected returns, would raise ERPs by at least one percent.

Conclusion

This paper has sought to bring the essence of recent research on the equity risk premium to practicing actuaries. The researchers covered here face the same ubiquitous problems that actuaries face daily: Do I rely on past data to forecast the future (costs, premiums, investments) or do I analyze the past and apply informed judgment as to future differences, if any, to arrive at actuarially fair forecasts? Most of the ERP estimates lower than the unconditional historical estimate have an undue reliance on recent lower dividend yields (without a recognition of capital gains[52]) and/or on data prior to 1926.

Despite a spate of research suggesting ex ante ERPs lower than recent realized ERPs, actuaries should be aware of the range of estimates covered here (Appendix B); be aware of the underlying assumptions, data and terminology; and be aware that their independent analysis is required before adopting an estimate other than the historical average. We believe that the Ibbotson-Chen (2003) layout, reproduced here as Appendix D, offers the actuary both an understanding of the fundamental components of the historical ERP and the opportunity to change the estimates based upon good judgment and supportable beliefs. We believe that reliance solely on "expert" survey averages, whether of financial analysts, academic economists, or CFOs, is fraught with risks of statistical bias to fair estimates of the forward ERP.

[52] Under the current US tax code, capital gains are tax-advantaged relative to dividend income for the vast majority of equity holders (households and mutual funds are 55% of the total equity holders, Federal Flow of Funds, 2002 Q3, Table L-213). Curiously, the reverse is true for property-liability insurers because of the 70% stock dividend exclusion afforded insurers.

It is dangerous for actuaries to engage in simplistic analyses of historical ERPs to generate ex ante forecasts that differ from the realized mean.[53] The research we have catalogued in Appendix B, the common level ERPs estimated in Appendix C, and the building block (historical) approach of Ibboston and Chen in Appendix D all discuss important concepts related to both ex post and ex ante ERPs and cannot be ignored in reaching an informed estimate. For example, Richard Wendt, writing in a 2002 issue of Risks and Rewards, a newsletter of the Society of Actuaries, concludes that a linear relationship is a better predictor of future returns than a "constant" ERP based on the average historical return. He arrives at this conclusion by estimating a regression equation[54] relating long bond yields with 15-year geometric mean market returns starting monthly in 1960. First, there is no significant relationship between short, intermediate or long-term income returns over 1926-2002 (or 1960-2002) and ERPs, as evidenced by simple regressions using Ibbotson data.[55] Second, if the linear structural equation indeed held, there would be no need for an ERP since the (15-year) return could be predicted within small error bars. Third, there is always a negative bias introduced when geometric averages are used as dependent variables (Brennan and Schwartz, 1985). Finally, the results are likely to be spurious due to the high autocorrelations of the target and independent variables; an autocorrelation correction would eliminate any significant relationship of long-yields to the ERP.

Actuaries should also be aware of the variability of both the ERP and risk-free rate estimates discussed in this paper (see Tables 4 and 9). All too often, return estimates are made without noting the error bars and that can lead to unexpected "surprises". As one example, recent research by Francis Longstaff (2002), proposes that a 1991-2001 "flight to quality" has created a valuation premium (and lowered yields) in the entire yield curve of Treasuries. He finds a 10 to 16 basis point liquidity premium throughout the zero coupon Treasury yield curve. He translates that into a 10% to 15% pricing difference at the long end. This would imply a simple CAPM market estimate for the long horizon might be biased low.

Finally, actuaries should know that the research catalogued in Appendix B is not definitive. No simple model of ERP estimation has been universally accepted. Undoubtedly, there will be still more empirical and theoretical research into this data rich financial topic. We await the potential advances in understanding the return process that the behavioral view may uncover.

Post Script: Appendices A-D

We provide four appendices that catalogue the ERP approaches and estimates discussed in the paper. Actuaries, in particular, should find the numerical values, and descriptions of assumptions underlying those values, helpful for valuation work that

[53] ERPs are derived from historical or expected after corporate tax returns. Pre-tax returns depend uniquely on the tax schedule for the differing sources of income.
[54] 15-year mean returns = 2.032 (Long Government Bond Yield) – 0.0242, R^2 = 0.882.
[55] The p-values on the yield-variables in an ERP/Yield regression using 1926-2002 annual data are 0.1324, 0.2246, and 0.3604 for short, intermediate and long term yields respectively with adjusted r square virtually zero.

adjusts for risk. Appendix A provides the annual Ibbotson data from 1926 through 2002 from Ibbotson Associates referred to throughout this paper. The equity risk-premium shown is a simple difference of the arithmetic stock returns and the arithmetic U.S. Treasury Bills total returns. Appendix B is a compilation of articles and books related to the equity risk premium. The puzzle research section contains the articles and books that were most related to addressing the equity risk premium puzzle. Page 1 of Appendix B gives each source, along with risk-free rate and equity risk premium estimates. Then, each source's estimate is classified by type (indicated with an X for the appropriate type). Page 2 of Appendix B shows further details collected from each source. This page adds the data period used, if applicable, and the projection period. We also list the general methodology used in the reference. The final three pages of Appendix B provide the footnotes which give additional details on the sources' intent.

Appendix C adjusts all the equity risk premium estimates to a short-horizon, arithmetic, unconditional ERP estimate. We begin with the authors' estimates for a stock return (the risk-free rate plus the ERP estimate). Next, we make adjustments if the ERP "type" given by the author(s) is not given in this format. For example, to adjust from a geometric to an arithmetic ERP estimate, we adjust upwards by the 1926-2002 historical difference in the arithmetic large company stocks' total return and the geometric large company stocks' total return of 2%. Next, if the estimate is given in real instead of nominal terms, we adjust the stock return estimate upwards by 3.1%, the 1926-2002 historical return for inflation.

We make an approximate adjustment to move the estimate from a conditional to unconditional estimate based on Fama and French (2002). Using the results for the 1951-2000 period shown in Table 4 of their paper and the standard deviations provided in Table 1, we have four adjustments based on their data. For the 1951-2000 period, Fama and French use an adjustment of 1.28% for the dividend growth model and 0.46% for the earnings growth model. Following a similar calculation, the 1872-2000 period would require a 0.82% adjustment using a dividend growth model; the 1872-1950 period would require a 0.54% adjustment using a dividend growth model. Earnings growth models were used by Fama and French only for the 1951-2000 data period. Therefore, we selected the lowest adjustment (0.46%) as a minimum adjustment from a conditional estimate to an unconditional estimate. Finally, we subtract the 1926-2002 historical U.S. Treasury Bills' total return to arrive at an adjusted equity risk premium.

These adjustments are only approximations because the various sources rely on different underlying data, but the changes in the ERP estimate should reflect the underlying concept that different "types" of ERPs cannot be directly compared and require some attempt to normalize the various estimates.

Page 1 of Appendix D is a table from Ibbotson and Chen which breaks down historical returns using various methods that correspond to their 2003 paper (reprinted with permission of Ibbotson Associates). The bottom portion provides forward-looking estimates. Page 2 of Appendix D is provided to show the formulas that Ibbotson and Chen develop within their paper.

References

Arnott, Robert D., and Peter L. Bernstein, 2002, What Risk Premium Is "Normal"?. *Financial Analysts Journal,* 58, 64-85.

Arnott, Robert D., and Ronald J. Ryan, 2001, The Death of the Risk Premium: Consequences of the 1990s, *Journal of Portfolio Management,* 27, 61-74.

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect Theory and Asset Prices, *Quarterly Journal of Economics,* 116, 1-53.

Benartzi, Shlomo, and Richard H. Thaler, 1995, Myopic Loss Aversion and the Equity Premium Puzzle, *Quarterly Journal of Economics,* 110, 73-92.

Brealey, Richard A., and Stewart C. Myers, 2000, *Principles of Corporate Finance*, 6th ed. (Boston: Irwin McGraw-Hill).

Brennan, Michael J., and Eduardo S. Schwartz, 1985, On the Geometric Mean Index: A note, *Journal of Financial and Quantitative Analysis*, 20,119-122.

Brown, Stephen J., William Goetzmann, Roger G. Ibboston and Stephen A. Ross 1992, Survivorship Bias in Performance Studies, *Review of Financial Studies,* 5, 553-580.

Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance,* 50, 853-873.

Campbell, John Y., 2001, Forecasting U.S. Equity Returns in the 21st Century, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board. http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Campbell, John Y., and John H. Cochrane, 1999, By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economics* 107: 205-251.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton: Princeton University Press).

Campbell, John Y., and Robert J. Shiller, 1989, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1, 195-228.

Campbell, John Y., and Robert J. Shiller, 2001, Valuation Ratios and the Long-Run Stock Market Outlook: An Update, Working paper #8221, NBER, forthcoming in

Advances in Behavioral Finance, Volume II, Nicholas Barberis and Richard Thaler eds., Russell Sage Foundation, 2003.

Claus, James, and Jacob Thomas, 2001, Equity Premia as Low as Three Percent?, *Journal of Finance,* 56, 1629-1666.

Cochrane, John H., 1997, Where is the Market Going? Uncertain Facts and Novel Theories, *Economic Perspectives,* 21, 3-37.

Constantinides, George M., John B. Donaldson, and Rajnish Mehra, 2002, Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle, *Quarterly Journal of Economics* 117: 269-296.

Constantinides, George M., 2002, Rational Asset Prices, *Journal of Finance,* 57, 1567-1591.

Cornell, Bradford, 1999, *The Equity Risk Premium: The Long-Run Future of the Stock Market* (New York: John Wiley & Sons).

Diamond, Peter A., 1999, What Stock Market Returns to Expect for the Future?, *Center for Retirement at Boston College number 2.*

Diamond, Peter A., 2001, What Stock Market Returns to Expect for the Future: An Update, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board. http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2002, *Triumph of the Optimists: 101 Years of Global Investment Returns* (Princeton: Princeton University Press).

Elton, Edwin J., and Martin J. Gruber, 1995, *Modern Portfolio Theory and Investment Analysis*, 5th ed. (New York: John Wiley & Sons).

Fama, Eugene F., and Kenneth R. French, 2002, The Equity Premium, *Journal of Finance,* 57, 637-659.

Fama, Eugene F., and Kenneth R French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance,* 47, 427-465.

Goetzmann, William N., Roger G. Ibbotson, and Liang Peng, 2001, A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability, *Journal of Financial Markets,* 4, 1-32.

Goss, Stephen C., 2001, Equity Yield Assumptions Used by the Office of the Chief Actuary, Social Security Administration, to Develop Estimates for Proposals with Trust

Fund and/or Individual Account Investments, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board. http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Graham, John R., and Campbell R. Harvey, 2002, Expectations of Equity Risk Premia, Volatility and Asymmetry, Working paper, Duke University.

Grinold, Richard, and Kenneth Kroner, 2002, The Equity Risk Premium, *Barclays Global Investors' Investment Insights* 5.

Harris, Robert S., and Felicia C. Marston, 2001, The Market Risk Premium: Expectational Estimates Using Analysts' Forecasts, *Journal of Applied Finance,* 11, 6-16.

Harvey, Andrew, 1990, *The Econometric Analysis of Time Series*, 2nd ed. (Cambridge, MA: MIT Press).

Ibbotson Associates, 2003, *Stocks, Bonds, Bills, and Inflation Yearbook* (Ibbotson Associates, Chicago).

Ibbotson Associates, 2003, *Stocks, Bonds, Bills, and Inflation Yearbook*, Valuation Edition (Ibbotson Associates, Chicago).

Ibbotson, Roger G., and Peng Chen, 2003, Long-Run Stock Returns: Participating in the Real Economy, *Financial Analysts Journal,* 59, 88-98.

Kaplan, Paul D., and James D. Peterson, 1998, Full-Information Industry Betas, *Financial Management,* 27, 85-93.

LaChance, Marie-Eve, and Olivia S. Mitchell, 2003, Understanding Individual Account Guarantees, University of Michigan Retirement Research Center working paper WP 2003-035.

Leibowitz, Martin L., 2001, Equity Risk Premium Forum, TIAA-CREF, New York City, November. http://www.aimrpubs.org/ap/issues/v2002n1/toc.html

Longstaff, Francis A., 2002, The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices, Working paper #9312, NBER, forthcoming in *Journal of Business*.

Mamaysky, Harry, 2002, A Model for Pricing Stocks and Bonds with Default Risk, Yale ICF Working Paper No. 02-13.

Malkiel, Burton G., 1999, *A Random Walk Down Wall Street* (New York: W. W. Norton & Company).

Mehra, Rajnish, 2002, Equity Premium Puzzle, in *Mastering Investments*, Ed. James Pickford (London: Financial Times-Prentice Hall).

Mehra, Rajnish, and Edward C. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics,* 15, 145-161.

Pratt, Shannon, P., 1998, *Cost of Capital Estimation and Applications*, John Wiley & Sons, New York.

Ralfe, John, Cliff Speed, and Jon Palin, 2003, Pensions and Capital Structure: Why hold equities in the pension fund?, Society of Actuaries Symposium, Vancouver, June.

Ross, Stephen A., 1976, The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory,* 13, 341-360.

Schwert, G. William, 1990, Indexes of United States Stock Prices from 1802 to 1987, *Journal of Business,* 63, 399-426.

Shiller, Robert, 1989, *Market Volatility* (Cambridge, MA: MIT Press).

Shoven, John B., 2001, What Are Reasonable Long-Run Rates of Return to Expect on Equities?, *Estimating the Real Rate of Return on Stocks Over the Long Term*, manuscript presented to the Social Security Advisory Board.
http://www.ssab.gov/estimated%20rate%20of%20return.pdf

Siegel, Jeremy J., 1999, The Shrinking Equity Premium, *Journal of Portfolio Management,* 26, 10-17.

Siegel, Jeremy J., 2002, *Stocks for the Long Run: The Definitive Guide to Financial Market Returns and Long-Term Investment Strategies*, 3rd ed. (New York: McGraw-Hill).

Social Security Advisory Board, 2002, Fiscal Year Annual Report
http://www.ssab.gov/annualreport2002.pdf

Tversky, Amos, and Daniel Kahneman, 1991, Loss Aversion and Riskless Choice: A Reference Dependent Model, *Quarterly Journal of Economics,* 152, 1039-1061.

Weil, Philippe, 1989, The Equity Premium Puzzle and the Risk-free Rate Puzzle, *Journal of Monetary Economics,* 24, 401-421.

Welch, Ivo, 2000, Views of Financial Economists on the Equity Premium and on Professional Controversies, *Journal of Business,* 73, 501-537.

Welch, Ivo, 2001, The Equity Premium Consensus Forecast Revisited, Cowles Foundation discussion paper, No. 1325.

Wendt, Richard, 2002, Understanding Equity Risk Premium, *Risk and Rewards*, No. 38, Society of Actuaries, February.

Wilson, Jack W., and Charles P. Jones, 2002, An Analysis of the S&P 500 Index and Cowles's Extensions: Price Indexes and Stock Returns, 1870-1999, *Journal of Business*, 75:3, 505-533.

Appendix A Ibbotson Market Data 1926-2002*			
	Common Stocks	**U. S. Treasury Bills**	**Arithmetic Short-Horizon**
Year	**Total Annual Returns**	**Total Annual Returns**	**Equity Risk Premia**
1926	11.62%	3.27%	8.35%
1927	37.49%	3.12%	34.37%
1928	43.61%	3.56%	40.05%
1929	- 8.42%	4.75%	-13.17%
1930	-24.90%	2.41%	-27.31%
1931	-43.34%	1.07%	-44.41%
1932	- 8.19%	0.96%	- 9.15%
1933	53.99%	0.30%	53.69%
1934	- 1.44%	0.16%	- 1.60%
1935	47.67%	0.17%	47.50%
1936	33.92%	0.18%	33.74%
1937	-35.03%	0.31%	-35.34%
1938	31.12%	- 0.02%	31.14%
1939	- 0.41%	0.02%	- 0.43%
1940	- 9.78%	0.00%	- 9.78%
1941	-11.59%	0.06%	-11.65%
1942	20.34%	0.27%	20.07%
1943	25.90%	0.35%	25.55%
1944	19.75%	0.33%	19.42%
1945	36.44%	0.33%	36.11%
1946	- 8.07%	0.35%	- 8.42%
1947	5.71%	0.50%	5.21%
1948	5.50%	0.81%	4.69%
1949	18.79%	1.10%	17.69%
1950	31.71%	1.20%	30.51%
1951	24.02%	1.49%	22.53%
1952	18.37%	1.66%	16.71%
1953	- 0.99%	1.82%	- 2.81%
1954	52.62%	0.86%	51.76%
1955	31.56%	1.57%	29.99%
1956	6.56%	2.46%	4.10%

Appendix A
Ibbotson Market Data 1926-2002*

	Common Stocks	U. S. Treasury Bills	Arithmetic Short-Horizon
Year	Total Annual Returns	Total Annual Returns	Equity Risk Premia
1957	-10.78%	3.14%	-13.92%
1958	43.36%	1.54%	41.82%
1959	11.96%	2.95%	9.01%
1960	0.47%	2.66%	- 2.19%
1961	26.89%	2.13%	24.76%
1962	- 8.73%	2.73%	-11.46%
1963	22.80%	3.12%	19.68%
1964	16.48%	3.54%	12.94%
1965	12.45%	3.93%	8.52%
1966	-10.06%	4.76%	-14.82%
1967	23.98%	4.21%	19.77%
1968	11.06%	5.21%	5.85%
1969	- 8.50%	6.58%	-15.08%
1970	4.01%	6.52%	- 2.51%
1971	14.31%	4.39%	9.92%
1972	18.98%	3.84%	15.14%
1973	-14.66%	6.93%	-21.59%
1974	-26.47%	8.00%	-34.47%
1975	37.20%	5.80%	31.40%
1976	23.84%	5.08%	18.76%
1977	- 7.18%	5.12%	-12.30%
1978	6.56%	7.18%	- 0.62%
1979	18.44%	10.38%	8.06%
1980	32.42%	11.24%	21.18%
1981	- 4.91%	14.71%	-19.62%
1982	21.41%	10.54%	10.87%
1983	22.51%	8.80%	13.71%
1984	6.27%	9.85%	- 3.58%
1985	32.16%	7.72%	24.44%
1986	18.47%	6.16%	12.31%
1987	5.23%	5.47%	- 0.24%
1988	16.81%	6.35%	10.46%
1989	31.49%	8.37%	23.12%

Appendix A
Ibbotson Market Data 1926-2002*

Year	Common Stocks Total Annual Returns	U. S. Treasury Bills Total Annual Returns	Arithmetic Short-Horizon Equity Risk Premia
1990	- 3.17%	7.81%	-10.98%
1991	30.55%	5.60%	24.95%
1992	7.67%	3.51%	4.16%
1993	9.99%	2.90%	7.09%
1994	1.31%	3.90%	- 2.59%
1995	37.43%	5.60%	31.83%
1996	23.07%	5.21%	17.86%
1997	33.36%	5.26%	28.10%
1998	28.58%	4.86%	23.72%
1999	21.04%	4.68%	16.36%
2000	- 9.11%	5.89%	-15.00%
2001	-11.88%	3.83%	-15.71%
2002	-22.10%	1.65%	-23.75%
mean=	**12.20%**	**3.83%**	**8.37%**
Standard Dev=	**20.49%**	**3.15%**	**20.78%**

* *2003 SBBI Yearbook pages 38 and 39*

Appendix B

Source	Risk-free-Rate	ERP Estimate	Real risk-free rate	Nominal risk-free rate	Geometric	Arithmetic	Long-horizon	Short-horizon	Short-run expectation	Long-run expectation	Conditional	Unconditional
Historical												
Ibbotson Associates	3.8% [7]	8.4% [31]		X		X		X		X		X
Social Security												
Office of the Chief Actuary [1]	2.3%,3.0% [8]	4.7%,4.0% [32]	X		X		X			X		X
John Campbell [2]	3% to 3.5% [9]	1.5-2.5%, 3-4% [33]	X		X	X	X	X		X	X	
Peter Diamond	2.2% [10]	<4.8% [34]	X		X		X			X	X	
Peter Diamond [3]	3.0% [11]	3.0% to 3.5% [35]	X		X		X			X	X	
John Shoven [4]	3.0%,3.5% [12]	3.0% to 3.5% [36]	X		X		X			X	X	
Puzzle Research												
Robert Arnott and Peter Bernstein	3.7% [13]	2.4% [37]	X		X		X			X	X	
Robert Arnott and Ronald Ryan	4.1% [14]	-0.9% [38]	X		X		X			X	X	
John Campbell and Robert Shiller	N/A	Negative [39]	X		?		?		X		X	
James Claus and Jacob Thomas	7.64% [15]	3.39% or less [40]		X		X	X			X	X	
George Constantinides	2.0% [16]	6.9% [41]	X			X		X		X		X
Bradford Cornell	5.6%, 3.8% [17]	3.5-5.5%, 5-7% [42]		X		X	X	X		X	X	
Dimson, Marsh, & Staunton	1.0% [18]	5.4% [43]	X			X		X		X	X	
Eugene Fama and Kenneth French	3.24% [19]	3.83% & 4.78% [44]	X			X		X		X		X
Robert Harris and Felicia Marston	8.53% [20]	7.14% [45]		X		X	X		X		X	
Roger Ibbotson and Peng Chen	2.05% [21]	4% and 6% [46]	X		X	X	X			X		X
Jeremy Siegel	4.0% [22]	-0.9% to -0.3% [47]	X		X		X			X	X	
Jeremy Siegel	3.5% [23]	2-3% [48]	X		X		X			?	X	
Surveys												
John Graham and Campbell Harvey	? by survey [24]	3-4.7% [49]		X		?	X		X		X	
Ivo Welch	N/A [25]	7% [50]		X		X		X		X	X	
Ivo Welch [5]	5% [26]	5.0% to 5.5% [51]		X		X		X		X	X	
Misc.												
Barclays Global Investors	5% [27]	2.5%, 3.25% [52]		X	X		X		X		X	
Richard Brealey and Stewart Myers	N/A [28]	6 to 8.5% [53]		X		X		X		X		X
Burton Malkiel	5.25% [29]	2.75% [54]		X	X		X			X	X	
Richard Wendt [6]	5.5% [30]	3.3% [55]		X		X	X			X	X	

Long-run expectation considered to be a forecast of more than 10 years.
Short-run expectation considered to be a forecast of 10 years or less.

Source	Risk-free Rate	ERP Estimate	Data Period	Methodology
Historical				
Ibbotson Associates	3.8% [7]	8.4% [31]	1926-2002	Historical
Social Security				
Office of the Chief Actuary [1]	2.3%, 3.0% [8]	4.7%, 4.0% [32]	1900-1995, Projecting out 75 years	Historical
John Campbell [2]	3% to 3.5% [9]	1.5-2.5%, 3-4% [33]	Projecting out 75 years	Historical & Ratios (Div/Price & Earn Gr)
Peter Diamond	2.2% [10]	<4.8% [34]	Last 200 yrs for eq/ 75 for bonds, Proj 75 yrs	Fundamentals: Div Yld, GDP Gr
Peter Diamond [3]	3.0% [11]	3.0% to 3.5% [35]	Projecting out 75 years	Fundamentals: Div/Price
John Shoven [4]	3.0%, 3.5% [12]	3.0% to 3.5% [36]	Projecting out 75 years	Fundamentals: P/E, GDP Gr
Puzzle Research				
Robert Arnott and Peter Bernstein	3.7% [13]	2.4% [37]	1802 to 2001, normal	Fundamentals: Div Yld & Gr
Robert Arnott and Ronald Ryan	4.1% [14]	-0.9% [38]	Past 74 years, 74 year projection [56]	Fundamentals: Div Yld & Gr
John Campbell and Robert Shiller	N/A	Negative [39]	1871 to 2000, ten-year projection	Ratios: P/E and Div/Price
James Claus and Jacob Thomas	7.64% [15]	3.39% or less [40]	1985-1998, long-term	Abnormal Earnings model
George Constantinides	2.0% [16]	6.9% [41]	1872 to 2000, long-term	Hist. and Fund.: Price/Div & P/E
Bradford Cornell	5.6%, 3.8% [17]	3.5-5.5%, 5-7% [42]	1926-1997, long run forward-looking	Weighing theoretical and empirical evid
Dimson, Marsh, & Staunton	1.0% [18]	5.4% [43]	1900-2000, prospective	Adj hist ret, Var of Gordon gr model
Eugene Fama and Kenneth French	3.24% [19]	3.83% & 4.78% [44]	Estimate for 1951-2000, long-term	Fundamentals: Dividends and Earnings
Robert Harris and Felicia Marston	8.53% [20]	7.14% [45]	1982-1998, expectational	Fin analysts' est, div gr model
Roger Ibbotson and Peng Chen	2.05% [21]	4% and 6% [46]	1926-2000, long-term	Historical and supply side approaches
Jeremy Siegel	4.0% [22]	-0.9% to -0.3% [47]	1871 to 1998, forward-looking	Fundamentals: P/E, Div Yld, Div Gr
Jeremy Siegel	3.5% [23]	2-3% [48]	1802-2001, forward-looking	Earnings yield
Surveys				
John Graham and Campbell Harvey	? by survey [24]	3-4.7% [49]	2Q 2000 thru 3Q 2002, 1 & 10 year proj	Survey of CFO's
Ivo Welch	N/A [25]	7% [50]	30-Year forecast, surveys in 97/98 & 99	Survey of financial economists
Ivo Welch [5]	5% [26]	5.0% to 5.5% [51]	30-Year forecast, survey around August 2001	Survey of financial economists
Misc.				
Barclays Global Investors	5% [27]	2.5%, 3.25% [52]	Long-run (10-year) expected return	Fundamentals: Inc, Earn Gr, & Repricing
Richard Brealey and Stewart Myers	N/A [28]	6 to 8.5% [53]	1926-1997	Predominantly Historical
Burton Malkiel	5.25% [29]	2.75% [54]	1926 to 1997, estimate millennium [57]	Fundamentals: Div Yld, Earn Gr
Richard Wendt [6]	5.5% [30]	3.3% [55]	1960-2000, estimate for 2001-2015 period	Linear regression model

Footnotes:

[1] Social Security Administration.

[2] Presented to the Social Security Advisory Board.

[3] Presented to the Social Security Advisory Board. Update of 1999 article.

[4] Presented to the Social Security Advisory Board.

[5] Update to Welch 2000.

[6] Newsletter of the Investment Section of the Society of Actuaries.

[7] Arithmetic mean of U.S. Treasury bills annual total returns from 1926-2002.

[8] 2.3% Long-run real yield on Treasury bonds; used for Advisory Council proposals. 3.0% Long-term real yield on Treasury bonds; used in 1999 Social Security Trustees Report.

[9] Estimate for safe real interest rates in the future based on yield of long-term inflation-indexed Treasury securities of 3.5% and short-term real interest rates recently averaging about 3%.

[10] Real long-term bond yield using 75 year historical average.

[11] Real yield on long-term Treasuries (assumption by OCACT).

[12] 3.0% is the OCACT assumption. 3.5% is the real return on long-run (30-year) inflation-indexed Treasury securities.

[13] Long-term expected real geometric bond return (10 year-horizon).

[14] The yield on US government inflation-indexed bonds (starting bond real yield in Jan 2000).

[15] Average 10-year Government T-bond yield between 1985 and 1998 (yield of 11.43% in 1985 to 5.64% in 1998. The mean 30-year risk-free rate for each year of the U.S. sample period is 31 basis points higher than the mean 10-year risk-free rate.

[16] Rolled-over real arithmetic return of three-month Treasury bills and certificates.

[17] Historical 20-year Treasury bond return of 5.6%. Yield on 20-year Treasury bonds in 1998 was approximately 6%. Historical 1-month Treasury bill return of 3.8%. Yield on 1-month Treasury bills in 1998 was approximately 4%.

[18] United States historical arithmetic real Treasury bill return over 1900-2000 period. 0.9% geometric Treasury bill return.

[19] Average real return on six-month commercial paper (proxy for risk-free interest rate). Substituting the one-month Treasury bill rate for the six-month commercial paper rate causes estimates of the annual equity premium for 1951-2000 to rise by about 1.00%.

[20] Average yield to maturity on long-term U.S. government bonds, 1982-1998.

[21] Real, geometric risk-free rate. Geometric risk-free rate with inflation (nominal) 5.13%.
Nominal yield equivalent to historical geometric long-term government bond income return for 1926-2000.

[22] The ten- and thirty-year TIPS bond yielded 4.0% in August 1999.

[23] Return on inflation-indexed securities.

[24] Current 10-year Treasury bond yield. Survey administered from June 6, 2000 to June 4, 2002. The rate on the 10-year Treasury bond changes in each survey. For example, in the Dec. 1, 2000 survey, the current annual yield on the 10-year Treasury bond was 5.5%. For the June 6, 2001 survey, the current annual yield on the 10-year Treasury bond was 5.3%.

[25] Arithmetic per-annum average return on rolled-over 30-day T-bills.

[26] Average forecast of arithmetic risk-free rate of about 5% by deducting ERP from market return.

[27] Current nominal 10-year bond yield.

[28] Return on Treasury bills. Treasury bills yield of about 5 percent in mid-1998. Average historical return on Treasury bills 3.8 percent.

[29] Good quality corporate bonds will earn approximately 6.5% to 7%. Long-term zero-coupon Treasury bonds will earn about 5.25%. Long-term TIPS will earn a real return of 3.75%.

[30] 1/1/01 Long T-Bond yield; uses initial bond yields in predictive model.

[31] Arithmetic short-horizon expected equity risk premium. Arithmetic intermediate-horizon expected equity risk premium 7.4%. Arithmetic long-horizon expected equity risk premium 7.0%. Geometric short-horizon expected equity risk premium 6.4%.

[32] Geometric equity premium over long-term Treasury securities. OCACT assumes a constant geometric real 7.0% stock return.

[33] Long-run average equity premium of 1.5% to 2.5% in geometric terms and 3% to 4% in arithmetic terms.

[34] Lower return over the next decade, followed by a geometric, real 7.0% stock return for remaining 65 years or lower rate of return for entire 75-year period (obscures pattern of returns).

[35] Most likely poor return over the next decade followed by a return to historic yields. Working from OCACT stock return assumption, he gives a single rate of return on equities for projection purposes of 6.0 to 6.5% (geometric, real).

[36] Geometric real stock return over the geometric real return on long-term government bonds.

[37] Expected geometric return over long-term government bonds. Their current risk premium is approximately zero, and their recommended expectation for the future real return for both stocks and bonds is 2-4 percent. The "normal" level of the risk premium is modest (2.4 percent or quite possibly less).

[38] Geometric real returns on stocks are likely to be in the 3%-4% range for the foreseeable future (10-20 years).

[39] Substantial declines in real stock prices, and real stock returns below zero, over the next ten years (2001-2010).

[40] The equity premium for each year between 1985 and 1998 in the United States. Similar results for five other markets.

[41] Unconditional, arithmetic mean aggregate equity premium over the 1872-2000 period. Over the period 1951 to 2000, the adjusted estimate of the unconditional mean premium is 6.0%. The corresponding estimate over the 1926 to 2000 period is 8.0%. Sharp distinction between conditional, short-term forecasts of the mean equity return and premium and estimates of the unconditional mean.

[42] Long run arithmetic future ERP of 3.5% to 5.5% over Treasury bonds and 5% to 7% over Treasury bills. Compares estimates to historical returns of 7.4% for bond premium and 9.2% for bill premium.

[43] 5.4% United States arithmetic expected future ERP relative to bills. 4.0% World (16 countries) arithmetic expected future ERP relative to bills. 4.1% United States geometric expected future ERP relative to bills. 3.0% World (16 countries) geometric expected future ERP relative to bills.

[44] 3.83% unconditional expected annual simple equity premium return (referred to as the annual-bias adjusted estimate of the annual equity premium) using dividend growth model. 4.78% unconditional expected annual simple equity premium return (referred to as the annual-bias adjusted estimate of the annual equity premium) using earnings growth model. Compares these results against historical real equity risk premium of 7.43% for 1951-2000.

[45] Average expectational risk premium. Because of the possible bias of analysts' optimism, the estimates are interpreted as "upper bounds" for the market premium. The average expectational risk premium is approximately equal to the arithmetic (7.5%) long-term differential between returns on stocks and long-term government bonds.

[46] 4% geometric (real) and 6% arithmetic (real). Forward looking long-horizon sustainable equity risk premium.

[47] Using the dividend discount model, the forward-looking real long-term geometric return on equity is 3.3%. Based on the earnings yield, the forward-looking real long-term geometric return on equity is between 3.1% and 3.7%.

[48] Future geometric equity premium. Future real return on equities of about 6%.

[49] The 10-year premium. The one-year risk premium averages between 0.4 and 5.2% depending on the quarter surveyed.

[50] Arithmetic 30-year forecast relative to short-term bills; 10-year same estimate. Second survey 6.8% for 30 and 10-year estimate. 1-year horizon between 0.5% and 1.5% lower. Geometric 30-year forecast around 5.2% (50% responded to this question).

[51] Arithmetic 30-year equity premium (relative to short-term T-bills). Geometric about 50 basis points below arithmetic. Arithmetic 1-year equity premium 3 to 3.5%.

[52] 2.5% current (conditional) geometric equity risk premium. 3.25% long-run, geometric normal or equilibrium equity risk premium.

[53] Extra arithmetic return versus Treasury bills. "Brealey and Myers have no official position on the exact market risk premium, but we believe a range of 6 to 8.5 percent is reasonable for the United States. We are most comfortable with figures towards the upper end of the range."

[54] The projected geometric (nominal) total return for the S&P 500 is 8 percent per year.

[55] Arithmetic mean 15 year horizon.

[56] 74 years since Dec 1925 and 74 years starting Jan 2000.

[57] Estimate the early decades of the twenty-first century.

Appendix C
Estimating a Short-Horizon Arithmetic Unconditional Equity Risk Premium

Source	Risk-free Rate	ERP Estimate	Stock Return Estimate	Geometric to arithmetic	Real to nominal	Conditional to unconditional [60]	Fixed short-horizon RFR	Short-horizon arithmetic unconditional ERP estimate
	I	II	III	IV	V	VI	VII	VIII
Historical								
Ibbotson Associates	3.8% [7]	8.4% [31]	12.2%	0.0%	0.0%	0.00%	3.8%	8.4%
Social Security								
Office of the Chief Actuary [1]	2.3%,3.0% [8]	4.7%,4.0% [32]	7.0%	2.0%	3.1%	0.00%	3.8%	8.3%
John Campbell [2]	3% to 3.5% [9]	1.5-2.5%, 3-4% [33]	6.0%-7.5%	0.0%	3.1%	0.46%	3.8%	5.8%-7.3%
Peter Diamond	2.2% [10]	<4.8% [34]	<7.0%	2.0%	3.1%	0.46%	3.8%	<8.8%
Peter Diamond [3]	3.0% [11]	3.0% to 3.5% [35]	6.0%-6.5%	2.0%	3.1%	0.46%	3.8%	7.8%-8.3%
John Shoven [4]	3.0%,3.5% [12]	3.0% to 3.5% [36]	6.0%-7.0%	2.0%	3.1%	0.46%	3.8%	7.8%-8.8%
Puzzle Research								
Robert Arnott and Peter Bernstein	3.7% [13]	2.4% [37]	6.1%	2.0%	3.1%	0.46%	3.8%	7.9%
Robert Arnott and Ronald Ryan	4.1% [14]	-0.9% [38]	3.2%	2.0%	3.1%	0.46%	3.8%	5.0%
John Campbell and Robert Shiller	N/A	Negative [39]	Negative	N/A	N/A	N/A	N/A	N/A
James Claus and Jacob Thomas	7.64% [15]	3.39% or less [40]	11.03%	0.0%	0.0%	0.46%	3.8%	7.69%
George Constantinides	2.0% [16]	6.9% [41]	8.9%	0.0%	3.1%	0.00%	3.8%	8.2%
Bradford Cornell	5.6%, 3.8% [17]	3.5-5.5%, 5-7% [42]	8.8%-10.8%	0.0%	0.0%	0.46%	3.8%	5.5%-7.5%
Dimson, Marsh, & Staunton	1.0% [18]	5.4% [43]	6.4% [58]	0.0%	3.1%	0.46%	3.8%	6.2% [61]
Eugene Fama and Kenneth French	3.24% [19]	3.83% & 4.78% [44]	7.07%-8.02%	0.0%	3.1%	0.00%	3.8%	6.37%-7.32%
Robert Harris and Felicia Marston	8.53% [20]	7.14% [45]	12.34% [59]	0.0%	0.0%	0.46%	3.8%	9.00%
Roger Ibbotson and Peng Chen	2.05% [21]	4% and 6% [46]	8.05%	0.0%	3.1%	0.00%	3.8%	7.35%
Jeremy Siegel	4.0% [22]	-0.9% to -0.3% [47]	3.1%-3.7%	2.0%	3.1%	0.46%	3.8%	4.9%-5.5%
Jeremy Siegel	3.5% [23]	2-3% [48]	5.5%-6.5%	2.0%	3.1%	0.46%	3.8%	7.3%-8.3%
Surveys								
John Graham and Campbell Harvey	? by survey [24]	3-4.7% [49]	8.3%-10.2%	N/A	0.0%	0.46%	3.8%	5.0%-6.9%
Ivo Welch	N/A [25]	7% [50]	N/A	0.0%	0.0%	0.46%	0.0%	7.5%
Ivo Welch [5]	5% [26]	5.0% to 5.5% [51]	10.0%-10.5%	0.0%	0.0%	0.46%	3.8%	6.7%-7.2%
Misc.								
Barclays Global Investors	5% [27]	2.5%, 3.25% [52]	7.5%,8.25%	2.0%	0.0%	0.46%	3.8%	6.16%-6.91%
Richard Brealey and Stewart Myers	N/A [28]	6 to 8.5% [53]	N/A	0.0%	0.0%	0.00%	0.0%	6.0%-8.5%
Burton Malkiel	5.25% [29]	2.75% [54]	8.0%	2.0%	0.0%	0.46%	3.8%	6.7%
Richard Wendt [6]	5.5% [30]	3.3% [55]	8.8%	0.0%	0.0%	0.46%	3.8%	5.5%

Column formulas:
III = I + II
VIII = III + IV + V + VI –VII

Source for adjustments:
2003 Ibbotson Yearbook Table 2-1 page 33
Fama French 2002 (see footnote 60)

Footnotes (1-57 from Appendix B):

[58] World estimate of 5.0%.

[59] Long risk-free of 5.2% plus 7.14%.

[60] For the 1951-2000 period, Fama and French (2002) adjust the conditional dividend growth model estimate upwards by 1.28% for an unconditional estimate, and they make a 0.46% upwards adjustment to the earnings growth model. We select the smaller of the two as an approximate minimum adjustment. For the longer period of 1872-2000, a comparable adjustment would be 0.82% for the dividend growth model and 0.54% for the 1872-1950 period using a dividend growth model. Earnings growth rates are shown by Fama and French only for the 1951-2000 period.

[61] World estimate of 4.8%.

Appendix D

Historical and Forecasted Equity Returns- All Ibbotson and Chen Models (Percent).

Method/ Model	Sum	Inflation	Real Risk-Free Rate	Equity Risk Premium	Real Capital Gain	g(Real EPS)	g(Real Div)	- g (Pay out Ratio)	g (BV)	g (ROE)	g P/E)	g(Real GDP/ POP)	g(FS-GDP/ POP)	Income Return	Re-Investment + Interaction	Additional Growth	Forecast Earnings Growth
Column #	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
Historical																	
Method 1	10.70	3.08	2.05	5.24											0.33		
Method 2	10.70	3.08			3.02									4.28	0.32		
Method 3	10.70	3.08				1.75					1.25			4.28	0.34		
Method 4	10.70	3.08					1.23	0.51			1.25			4.28	0.35		
Method 5	10.70	3.08							1.46	0.31	1.25			4.28	0.31		
Method 6	10.70	3.08										2.04	0.96	4.28	0.32		
Forecast with Historical Dividend Yield																	
Model 3F	9.37	3.08				1.75								4.28	0.26		
Model 3F (ERP)	9.37	3.08	2.05	3.97											0.27		
Forecast with Current Dividend Yield																	
Model 4F	5.44	3.08					1.23							1.10 [a]	0.03		
Model 4F (ERP)	5.44	3.08	2.05	0.24											0.07		
Model $4F_2$	9.37	3.08					1.23	0.51						2.05 [b]	0.21	2.28	
Model $4F_2$ (FG)	9.37	3.08												1.10 [a]	0.21		4.98

Source: The data and format was made available by Ibbotson/Chen and is reprinted with permission by Ibbotson Associates. Corresponds to Ibbotson/Chen Table 2 Exhibit; column numbers have been added.

[a] 2000 dividend yield

[b] Assuming the historical average dividend-payout ratio, the 2000 dividend yield is adjusted up 0.95 pps.

	Formula*	Description of Method
Historical		
Method 1	I=(1+II)*(1+III)*(1+IV)-1	Building Blocks Method: inflation, real risk-free rate, and equity risk premium.
Method 2	I=[(1+II)*(1+V)-1]+XIV+XV	Capital Gain and Income Method: inflation, real capital gain, and income return.
Method 3	I=[(1+II)*(1+VI)*(1+XI)-1]+XIV+XV	Earnings Model: inflation, growth in earnings per share, growth in price to earnings ratio, and income return.
Method 4	I=[(1+II)*(1+XI)*(1+VII)/(1-VIII)-1]+XIV+XV	Dividends Model: inflation, growth rate of price earnings ratio, growth rate of the dollar amount of dividends after inflation, growth rate of payout ratio, and dividend yield (income return).
Method 5	I=[(1+II)*(1+XI)*(1+IX)*(1+X)-1]+XIV+XV	Return on Book Equity Model: inflation, growth rate of price earnings ratio, growth rate of book value, growth rate of ROE, and income return.
Method 6	I=[(1+II)*(1+XII)*(1+XIII)-1]+XIV+XV	GDP Per Capita Model: inflation, real growth rate of the overall economic productivity (GDP per capita), increase of the equity market relative to the overall economic productivity, and income return.
Forecast with Historical Dividend Yield		
Model 3F	I=[(1+II)*(1+VI)-1]+XIV+XV	Forward-looking Earnings Model: inflation, growth in real earnings per share, and income return.
Model 3F (ERP)	IV=(1+I)/[(1+II)*(1+III)]-1	Using Model 3F result to calculate ERP.
Forecast with Current Dividend Yield		
Model 4F	I=[(1+II)*(1+VII)-1]+XIV+XV	Forward-looking Dividends Model: inflation, growth in real dividend, and dividend yield (income return); also referred to as Gordon model.
Model 4F (ERP)	IV=(1+I)/[(1+II)*(1+III)]-1	Using Model 4F result to calculate ERP.
Model $4F_2$	I=[(1+II)*(1+VII)*(1+VIII)-1]+XIV+XV+XVI	Attempt to reconcile Model 4F and Model 3F.
Model $4F_2$ (FG)	XVII=[(1+I)/(1+II)-1]-XIV-XV	Using Method $4F_2$ result to calculate forecasted earnings.

Explanation of Ibbotson/Chen Table 2 Exhibit; using column numbers to represent formula.

High stock prices, together with projected slow economic growth, are not consistent with the 7.0 percent return that the Office of the Chief Actuary has generally used when evaluating proposals with stock investments. Routes out of the inconsistency include assuming higher GDP growth, a lower long-run stock return, or a lower short-run stock return with a 7.0 percent return on a lower base thereafter. In short, either the stock market is overvalued and requires a correction to justify a 7.0 percent return thereafter, or it is correctly valued and the long-run return is substantially lower than 7.0 percent (or some combination of the two). This article argues that the former view is more convincing, since accepting the "correctly valued" hypothesis implies an implausibly small equity premium.

This article originally appeared as an Issue in Brief of the Center for Retirement Research at Boston College (No. 2, September 1999). The research reported herein was performed pursuant to a grant from the Social Security Administration (SSA) funded as part of the Retirement Research Consortium. The opinions and conclusions expressed are solely those of the author and should not be construed as representing the opinions or policy of SSA, any agency of the federal government, or the Center for Retirement Research at Boston College.

*The author is a professor at the Massachusetts Institute of Technology. *(Manuscript received November 1999; submitted for external review December 1999; revise and resubmit recommended April 2000; revision received June 2000; paper accepted July 2000.)*

What Stock Market Returns to Expect for the Future?

*by Peter A. Diamond**

Summary

In evaluating proposals for reforming Social Security that involve stock investments, the Office of the Chief Actuary (OCACT) has generally used a 7.0 percent real return for stocks. The 1994-96 Advisory Council specified that OCACT should use that return in making its 75-year projections of investment-based reform proposals. The assumed ultimate real return on Treasury bonds of 3.0 percent implies a long-run equity premium of 4.0 percent. There are two equity-premium concepts: the *realized* equity premium, which is measured by the actual rates of return; and the *required* equity premium, which investors expect to receive for being willing to hold available stocks and bonds. Over the past two centuries, the realized premium was 3.5 percent on average, but 5.2 percent for 1926 to 1998.

Some critics argue that the 7.0 percent projected stock returns are too high. They base their arguments on recent developments in the capital market, the current high value of the stock market, and the expectation of slower economic growth.

Increased use of mutual funds and the decline in their costs suggest a lower required premium, as does the rising fraction of the American public investing in stocks. The size of the decrease is limited, however, because the largest cost savings do not apply to the very wealthy and to large institutional investors, who hold a much larger share of the stock market's total value than do new investors. These trends suggest a lower equity premium for projections than the 5.2 percent of the past 75 years. Also, a declining required premium is likely to imply a temporary increase in the realized premium because a rising willingness to hold stocks tends to increase their price. Therefore, it would be a mistake during a transition period to extrapolate what may be a temporarily high realized return. In the standard (Solow) economic growth model, an assumption of slower long-run growth lowers the marginal product of capital if the savings rate is constant. But lower savings as growth slows should partially or fully offset that effect.

The present high stock prices, together with projected slow economic growth, are not consistent with a 7.0 percent return. With a plausible level of adjusted dividends (dividends plus net share repurchases), the ratio of stock value to gross domestic product (GDP) would rise more than 20-fold over 75 years. Similarly, the steady-state Gordon formula—that stock returns equal the adjusted dividend yield plus the growth rate of stock prices (equal to that of GDP)—suggests a return of roughly 4.0 percent to 4.5 percent. Moreover, when relative stock values have been high, returns over the following decade have tended to be low.

To eliminate the inconsistency posed by the assumed 7.0 percent return, one could assume higher GDP growth, a lower long-run stock return, or a lower short-run stock return with a 7.0 percent return on a lower base thereafter. For example, with an adjusted dividend yield of 2.5 percent to 3.0 percent,

the market would have to decline about 35 percent to 45 percent in real terms over the next decade to reach steady state.

In short, either the stock market is overvalued and requires a correction to justify a 7.0 percent return thereafter, or it is correctly valued and the long-run return is substantially lower than 7.0 percent (or some combination). This article argues that the "overvalued" view is more convincing, since the "correctly valued" hypothesis implies an implausibly small equity premium. Although OCACT could adopt a lower rate for the entire 75-year period, a better approach would be to assume lower returns over the next decade and a 7.0 percent return thereafter.

Introduction

All three proposals of the 1994-96 Advisory Council on Social Security (1997) included investment in equities. For assessing the financial effects of those proposals, the Council members agreed to specify a 7.0 percent long-run real (inflation-adjusted) yield from stocks.[1] They devoted little attention to different short-run returns from stocks.[2] The Social Security Administration's Office of the Chief Actuary (OCACT) used this 7.0 percent return, along with a 2.3 percent long-run real yield on Treasury bonds, to project the impact of the Advisory Council's proposals.

Since then, OCACT has generally used 7.0 percent when assessing other proposals that include equities.[3] In the 1999 Social Security Trustees Report, OCACT used a higher long-term real rate on Treasury bonds of 3.0 percent.[4] In the first 10 years of its projection period, OCACT makes separate assumptions about bond rates for each year and assumes slightly lower real rates in the short run.[5] Since the assumed bond rate has risen, the assumed equity premium, defined as the difference between yields on equities and Treasuries, has declined to 4.0 percent in the long run.[6] Some critics have argued that the assumed return on stocks and the resulting equity premium are still too high.[7]

This article examines the critics' arguments and, rather than settling on a single recommendation, considers a range of assumptions that seem reasonable.[8] The article:

- Reviews the historical record on rates of return,
- Assesses the critics' reasons why future returns may be different from those in the historical record and examines the theory about how those rates are determined, and
- Considers two additional issues: the difference between gross and net returns, and investment risk.

Readers should note that in this discussion, a decline in the equity premium need not be associated with a decline in the return on stocks, since the return on bonds could increase. Similarly, a decline in the return on stocks need not be associated with a decline in the equity premium, since the return on bonds could also decline. Both rates of return and the equity premium are relevant to choices about Social Security reform.

Historical Record

Realized rates of return on various financial instruments have been much studied and are presented in Table 1.[9] Over the past 200 years, stocks have produced a real return of 7.0 percent per year. Even though annual returns fluctuate enormously, and rates vary significantly over periods of a decade or two, the return on stocks over very long periods has been quite stable (Siegel 1999).[10] Despite that long-run stability, great uncertainty surrounds both a projection for any particular period and the relevance of returns in any short period of time for projecting returns over the long run.

The equity premium is the difference between the rate of return on stocks and on an alternative asset—Treasury bonds, for the purpose of this article. There are two concepts of equity premiums. One is the *realized* equity premium, which is measured by the actual rates of return. The other is the *required* equity premium, which equals the premium that investors expect to get in exchange for holding available quantities of assets. The two concepts are closely related but different—significantly different in some circumstances.

The realized equity premium for stocks relative to bonds has been 3.5 percent for the two centuries of available data, but it has increased over time (Table 2).[11, 12] That increase has resulted from a significant decline in bond returns over the past

Table 1.
Compound annual real returns, by type of investment, 1802-1998 (in percent)

Period	Stocks	Bonds	Bills	Gold	Inflation
1802-1998	7.0	3.5	2.9	-0.1	1.3
1802-1870	7.0	4.8	5.1	0.2	0.1
1871-1925	6.6	3.7	3.2	-0.8	0.6
1926-1998	7.4	2.2	0.7	0.2	3.1
1946-1998	7.8	1.3	0.6	-0.7	4.2

Source: Siegel (1999).

Table 2.
Equity premiums: Differences in annual rates of return between stocks and fixed-income assets, 1802-1998

	Equity premium (percent)	
Period	With bonds	With bills
1802-1998	3.5	5.1
1802-1870	2.2	1.9
1871-1925	2.9	3.4
1926-1998	5.2	6.7
1946-1998	6.5	7.2

Source: Siegel (1999).

200 years. The decline is not surprising considering investors' changing perceptions of default risk as the United States went from being a less-developed country (and one with a major civil war) to its current economic and political position, where default risk is seen to be virtually zero.[13]

These historical trends can provide a starting point for thinking about what assumptions to use for the future. Given the relative stability of stock returns over time, one might initially choose a 7.0 percent assumption for the return on stocks—the average over the entire 200-year period. In contrast, since bond returns have tended to decline over time, the 200-year number does not seem to be an equally good basis for selecting a long-term bond yield. Instead, one might choose an assumption that approximates the experience of the past 75 years—2.2 percent, which suggests an equity premium of around 5.0 percent. However, other evidence, discussed below, argues for a somewhat lower value.[14]

Why Future Returns May Differ from Past Returns

Equilibrium and Long-Run Projected Rates of Return

The historical data provide one way to think about rates of return. However, thinking about how the future may be different from the past requires an underlying theory about how those returns are determined. This section lists some of the actions by investors, firms, and government that combine to determine equilibrium; it can be skipped without loss of continuity.

In asset markets, the demand by individual and institutional investors reflects a choice among purchasing stocks, purchasing Treasury bonds, and making other investments.[15] On the supply side, corporations determine the supplies of stocks and corporate bonds through decisions on dividends, new issues, share repurchases, and borrowing. Firms also choose investment levels. The supplies of Treasury bills and bonds depend on the government's budget and debt management policies as well as monetary policy. Whatever the supplies of stocks and bonds, their prices will be determined so that the available amounts are purchased and held by investors in the aggregate.

The story becomes more complicated, however, when one recognizes that investors base decisions about portfolios on their projections of both future prices of assets and future dividends.[16] In addition, market participants need to pay transactions costs to invest in assets, including administrative charges, brokerage commissions, and the bid-ask spread. The risk premium relevant for investors' decisions should be calculated net of transactions costs. Thus, the greater cost of investing in equities than in Treasuries must be factored into any discussion of the equity premium.[17] Differences in tax treatments of different types of income are also relevant (Gordon 1985; Kaplow 1994).

In addition to determining the supplies of corporate stocks and bonds, corporations also choose a debt/equity mix that affects the risk characteristics of both bonds and stocks. Financing a given level of investment more by debt and less by equity leaves a larger interest cost to be paid from the income of corporations before determining dividends. That makes both the debt and the equity more risky. Thus, changes in the debt/equity mix (possibly in response to prevailing stock market prices) should affect risk and, therefore, the equilibrium equity premium.[18]

Since individuals and institutions are generally risk averse when investing, greater expected variation in possible future yields tends to make an asset less valuable. Thus, a sensible expectation about long-run equilibrium is that the expected yield on equities will exceed that on Treasury bonds. The question at hand is how much more stocks should be expected to yield.[19] That is, assuming that volatility in the future will be roughly similar to volatility in the past, how much more of a return from stocks would investors need to expect in order to be willing to hold the available supply of stocks. Unless one thought that stock market volatility would collapse, it seems plausible that the premium should be significant. For example, equilibrium with a premium of 70 basis points (as suggested by Baker 1999a) seems improbable, especially since transactions costs are higher for stock than for bond investments. In considering this issue, one needs to recognize that a greater willingness to bear the risk associated with stocks is likely to be accompanied by greater volatility of stock prices if bond rates are unchanged. That is, fluctuations in expected growth in corporate profits will have bigger impacts on expected discounted returns (which approximate prices) when the equity premium, and so the discount rate, is lower.[20]

Although stocks should earn a significant premium, economists do not have a fully satisfactory explanation of why stocks have yielded so much more than bonds historically, a fact that has been called the equity-premium puzzle (Mehra and Prescott 1985; Cochrane 1997). Ongoing research is trying to develop more satisfactory explanations, but the theory still has inadequacies.[21] Nevertheless, to explain why the future may be different from the past, one needs to rely on some theoretical explanation of the past in order to have a basis for projecting a different future.

Commentators have put forth three reasons as to why future returns may be different from those in the historical record. First, past and future long-run trends in the capital market may imply a decline in the equity premium. Second, the current valuation of stocks, which is historically high relative to various benchmarks, may signal a lower future rate of return on equities. Third, the projection of slower economic growth may suggest a lower long-run marginal product of capital, which is the source of returns to financial assets. The first two issues are discussed in the context of financial markets; the third, in the context of physical assets. One should distinguish between arguments that suggest a lower equity premium and those that suggest lower returns to financial assets generally.

Equity Premium and Developments in the Capital Market

The capital market has experienced two related trends—the decrease in the cost of acquiring a diversified portfolio of stocks and the spread of stock ownership more widely in the economy. The relevant equity premium for investors is the equity premium net of the costs of investing. Thus, if the cost of investing in some asset decreases, that asset should have a higher price and a lower expected return gross of investment costs. The availability of mutual funds and the decrease in the cost of purchasing them should lower the equity premium in the future relative to long-term historical values. Arguments have also been raised about investors' time horizons and their understanding of financial markets, but the implications of those arguments are less clear.

Mutual Funds. In the absence of mutual funds, small investors would need to make many small purchases in different companies in order to acquire a widely diversified portfolio. Mutual funds provide an opportunity to acquire a diversified portfolio at a lower cost by taking advantage of the economies of scale in investing. At the same time, these funds add another layer of intermediation, with its costs, including the costs of marketing the funds.

Nevertheless, as the large growth of mutual funds indicates, many investors find them a valuable way to invest. That suggests that the equity premium should be lower in the future than in the past, since greater diversification means less risk for investors. However, the significance of the growth of mutual funds depends on the importance in total equity demand of "small" investors who purchase them, since this argument is much less important for large investors, particularly large institutional investors. According to recent data, mutual funds own less than 20 percent of U.S. equity outstanding (Investment Company Institute 1999).

A second development is that the average cost of investing in mutual funds has decreased. Rea and Reid (1998) report a drop of 76 basis points (from 225 to 149) in the average annual charge of equity mutual funds from 1980 to 1997. They attribute the bulk of the decline to a decrease in the importance of front-loaded funds (funds that charge an initial fee when making a deposit in addition to annual charges). The development and growth of index funds should also reduce costs, since index funds charge investors considerably less on average than do managed funds while doing roughly as well in gross rates of return. In a separate analysis, Rea and Reid (1999) also report a decline of 38 basis points (from 154 to 116) in the cost of bond mutual funds over the same period, a smaller drop than with equity mutual funds. Thus, since the cost of stock funds has fallen more than the cost of bond funds, it is plausible to expect a decrease in the equity premium relative to historical values. The importance of that decline is limited, however, by the fact that the largest cost savings do not apply to large institutional investors, who have always faced considerably lower charges.

A period with a declining required equity premium is likely to have a temporary increase in the realized equity premium. Assuming no anticipation of an ongoing trend, the divergence occurs because a greater willingness to hold stocks, relative to bonds, tends to increase the price of stocks. Such a price rise may yield a realized return that is higher than the required return.[22] The high realized equity premium since World War II may be partially caused by a decline in the required equity premium over that period. During such a transition period, therefore, it would be a mistake to extrapolate what may be a temporarily high realized return.

Spread of Stock Ownership. Another trend that would tend to decrease the equity premium is the rising fraction of the American public investing in stocks either directly or indirectly through mutual funds and retirement accounts (such as 401(k) plans). Developments in tax law, pension provision, and the capital markets have expanded the base of the population who are sharing in the risks associated with the return to corporate stock. The share of households investing in stocks in any form increased from 32 percent in 1989 to 41 percent in 1995 (Kennickell, Starr-McCluer, and Sundén 1997). Numerous studies have concluded that widening the pool of investors sharing in stock market risk should lower the equilibrium risk premium (Mankiw and Zeldes 1991; Brav and Geczy 1996; Vissing-Jorgensen 1997; Diamond and Geanakoplos 1999; Heaton and Lucas 2000). The importance of that trend must be weighted by the low size of investment by such new investors.[23]

Investors' Time Horizons. A further issue relevant to the future of the equity premium is whether the time horizons of investors, on average, have changed or will change.[24] Although the question of how time horizons should affect demands for assets raises subtle theoretical issues (Samuelson 1989), longer horizons and sufficient risk aversion should lead to greater willingness to hold stocks given the tendency for stock prices to revert toward their long-term trend (Campbell and Viceira 1999).[25]

The evidence on trends in investors' time horizons is mixed. For example, the growth of explicit individual retirement savings vehicles, such as individual retirement accounts (IRAs) and 401(k)s, suggests that the average time horizons of individual investors may have lengthened. However, some of that growth is at the expense of defined benefit plans, which may have longer horizons. Another factor that might suggest a longer investment horizon is the increase in equities held by institutional investors, particularly through defined benefit pension plans. However, the relevant time horizon for such holdings may not be the open-ended life of the plan but rather the horizon of the plans' asset managers, who may have career concerns that shorten the relevant horizon.

Other developments may tend to lower the average horizon. Although the retirement savings of baby boomers may currently add to the horizon, their aging and the aging of the population generally will tend to shorten horizons. Finally, individual stock ownership has become less concentrated (Poterba and Samwick 1995), which suggests a shorter time horizon because less wealthy investors might be less concerned about passing assets on to younger generations. Overall, without detailed calculations that would go beyond the scope of this article, it is not clear how changing time horizons should affect projections.

Investors' Understanding. Another factor that may affect the equity premium is investors' understanding of the properties of stock and bond investments. The demand for stocks might be affected by the popular presentation of material, such as Siegel (1998), explaining to the general public the difference between short- and long-run risks. In particular, Siegel highlights the risks, in real terms, of holding nominal bonds. While the creation of inflation-indexed Treasury bonds might affect behavior, the lack of wide interest in those bonds (in both the United States and the United Kingdom) and the failure to fully adjust future amounts for inflation generally (Shafir, Diamond, and Tversky 1997) suggest that nominal bonds will continue to be a major part of portfolios. Perceptions that those bonds are riskier than previously believed would then tend to decrease the required equity premium.

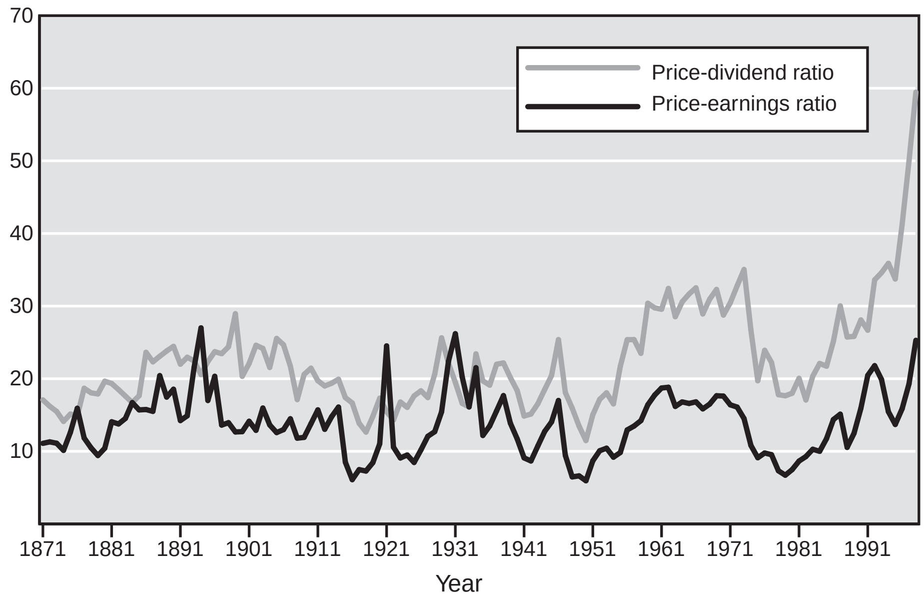
Popular perceptions may, however, be excessively influenced by recent events—both the high returns on equity and the low rates of inflation. Some evidence suggests that a segment of the public generally expects recent rates of increase in the prices of assets to continue, even when those rates seem highly implausible for a longer term (Case and Shiller 1988). The possibility of such extrapolative expectations is also connected with the historical link between stock prices and inflation. Historically, real stock prices have been adversely affected by inflation in the short run. Thus, the decline in inflation expectations over the past two decades would be associated with a rise in real stock prices if the historical pattern held. If investors and analysts fail to consider such a connection, they might expect robust growth in stock prices to continue without recognizing that further declines in inflation are unlikely. Sharpe (1999) reports evidence that stock analysts' forecasts of real growth in corporate earnings include extrapolations that may be implausibly high. If so, expectations of continuing rapid growth in stock prices suggest that the required equity premium may not have declined.

On balance, the continued growth and development of mutual funds and the broader participation in the stock market should contribute to a drop in future equity premiums relative to the historical premium, but the drop is limited.[26] Other factors, such as investors' time horizons and understanding, have less clear-cut implications for the equity premium.

Equity Premium and Current Market Values

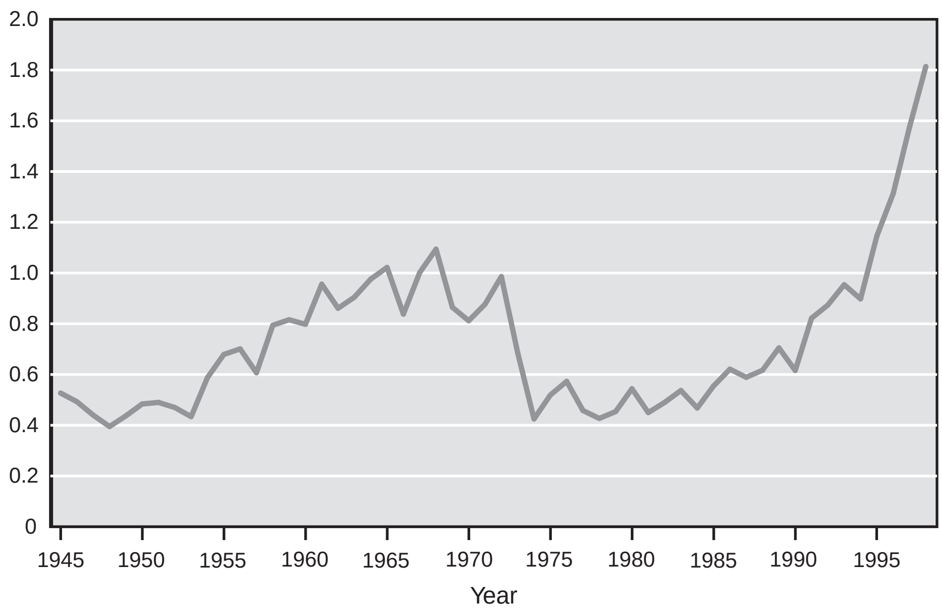
At present, stock prices are very high relative to a number of different indicators, such as earnings, dividends, book values, and gross domestic product (GDP) (Charts 1 and 2). Some

Chart 1.
Price-dividend ratio and price-earnings ratio, 1871-1998



Source: Robert Shiller, Yale University. Available at www.econ.yale.edu/~shiller/data/chapt26.html.
Note: These ratios are based on Standard and Poor's Composite Stock Price Index.

Chart 2.
Ratio of market value of stocks to gross domestic product,1945-1998



Source: Bureau of Economic Analysis data from the national income and product accounts and federal flow of funds.

critics, such as Baker (1998), argue that this high market value, combined with projected slow economic growth, is not consistent with a 7.0 percent return. Possible implications of the high prices have also been the subject of considerable discussion in the finance community (see, for example, Campbell and Shiller 1998; Cochrane 1997; Philips 1999; and Siegel 1999).

The inconsistency of current share prices and 7.0 percent real returns, given OCACT's assumptions for GDP growth, can be illustrated in two ways. The first way is to project the ratio of the stock market's value to GDP, starting with today's values and given assumptions about the future. The second way is to ask what must be true if today's values represent a steady state in the ratio of stock values to GDP.

The first calculation requires assumptions for stock returns, adjusted dividends (dividends plus net share repurchases),[27] and GDP growth. For stock returns, the 7.0 percent assumption is used. For GDP growth rates, OCACT's projections are used. For adjusted dividends, one approach is to assume that the ratio of the aggregate adjusted dividend to GDP would remain the same as the current level. However, as discussed in the accompanying box, the current ratio seems too low to use for projection purposes. Even adopting a higher, more plausible level of adjusted dividends, such as 2.5 percent or 3.0 percent, leads to an implausible rise in the ratio of stock value to GDP—in this case, a more than 20-fold increase over the next 75 years. The calculation derives each year's capital gains by subtracting projected adjusted dividends from the total cash flow to shareholders needed to return 7.0 percent on that year's share values. (See Appendix A for an alternative method of calculating this ratio using a continuous-time differential equation.)

A second way to consider the link between stock market value, stock returns, and GDP is to look at a steady-state relationship. The Gordon formula says that stock returns equal the ratio of adjusted dividends to prices (or the adjusted dividend yield) plus the growth rate of stock prices.[28] In a steady state,

Projecting Future Adjusted Dividends

This article uses the concept of adjusted dividends to estimate the dividend yield. The adjustment begins by adding the value of net share repurchases to actual dividends, since that also represents a cash flow to stockholders in aggregate. A further adjustment is then made to reflect the extent to which the current situation might not be typical of the relationship between dividends and gross domestic product (GDP) in the future. Three pieces of evidence suggest that the current ratio of dividends to GDP is abnormally low and therefore not appropriate to use for projection purposes.

First, dividends are currently very low relative to corporate earnings—roughly 40 percent of earnings compared with a historical average of 60 percent. Because dividends tend to be much more stable over time than earnings, the dividend-earnings ratio declines in a period of high growth of corporate earnings. If future earnings grow at the same rate as GDP, dividends will probably grow faster than GDP to move toward the historical ratio.[1] On the other hand, earnings, which are high relative to GDP, might grow more slowly than GDP. But then, corporate earnings, which have a sizable international component, might grow faster than GDP.

Second, corporations are repurchasing their outstanding shares at a high rate. Liang and Sharpe (1999) report on share repurchases by the 144 largest (nonbank) firms in the Standard and Poor's 500. From 1994 to 1998, approximately 2 percent of share value was repurchased, although Liang and Sharpe anticipate a lower value in the future. At the same time, those firms were issuing shares because employees were exercising stock options at prices below the share values, thus offsetting much of the increase in the number of shares outstanding. Such transfers of net wealth to employees presumably reflect past services. In addition, initial public offerings (IPOs) represent a negative cash flow from stockholders as a whole. Not only the amount paid for stocks but also the value of the shares held by insiders represents a dilution relative to a base for long-run returns on all stocks. As a result, some value needs to be added to the current dividend ratio to adjust for net share repurchases, but the exact amount is unclear. However, in part, the high rate of share repurchase may be just another reflection of the low level of dividends, making it inappropriate to both project much higher dividends in the near term and assume that all of the higher share repurchases will continue. Exactly how to project current numbers into the next decade is not clear.

Finally, projected slow GDP growth, which will plausibly lower investment levels, could be a reason for lower retained earnings in the future. A stable level of earnings relative to GDP and lower retained earnings would increase the ratio of adjusted dividends to GDP.[2]

In summary, the evidence suggests using an "adjusted" dividend yield that is larger than the current level. Therefore, the illustrative calculations in this article use adjusted dividend yields of 2.0 percent, 2.5 percent, 3.0 percent, and 3.5 percent. (The current level of dividends without adjustment for share repurchases is between 1.0 percent and 2.0 percent.)

---

[1] For example, Baker and Weisbrot (1999) appear to make no adjustment for share repurchases or for current dividends being low. However, they use a dividend payout of 2.0 percent, while Dudley and others (1999) report a current dividend yield on the Wilshire 5000 of 1.3 percent.

[2] Firms might change their overall financing package by changing the fraction of net earnings they retain. The implications of such a change would depend on why they were making it. A long-run decrease in retained earnings might merely be increases in dividends and borrowing, with investment held constant. That case, to a first approximation, is another application of the Modigliani-Miller theorem, and the total stock value would be expected to fall by the decrease in retained earnings. Alternatively, a change in retained earnings might signal a change in investment. Again, there is ambiguity. Firms might be retaining a smaller fraction of earnings because investment opportunities were less attractive or because investment had become more productive. These issues tie together two parts of the analysis in this article. If slower growth is associated with lower investment that leaves the return on capital relatively unchanged, then what financial behavior of corporations is required for consistency? Baker (1999b) makes such a calculation; it is not examined here.

the growth rate of prices can be assumed to equal that of GDP. Assuming an adjusted dividend yield of roughly 2.5 percent to 3.0 percent and projected GDP growth of 1.5 percent, the Gordon equation implies a stock return of roughly 4.0 percent to 4.5 percent, not 7.0 percent. Those lower values would imply an equity premium of 1.0 percent to 1.5 percent, given OCACT's assumption of a 3.0 percent yield on Treasury bonds. Making the equation work with a 7.0 percent stock return, assuming no change in projected GDP growth, would require an adjusted dividend yield of roughly 5.5 percent—about double today's level.[29]

For such a large jump in the dividend yield to occur, one of two things would have to happen—adjusted dividends would have to grow much more rapidly than the economy, or stock prices would have to grow much less rapidly than the economy (or even decline). But a consistent projection would take a very large jump in adjusted dividends, assuming that stock prices grew along with GDP starting at today's value. Estimates of recent values of the adjusted dividend yield range from 2.10 percent to 2.55 percent (Dudley and others 1999; Wadhwani 1998).[30]

Even with reasons for additional growth in the dividend yield, which are discussed in the box on projecting future dividends, an implausible growth of adjusted dividends is needed if the short- and long-term returns on stocks are to be 7.0 percent. Moreover, historically, very low values of the dividend yield and earnings-price ratio have been followed primarily by adjustments in stock prices, not in dividends and earnings (Campbell and Shiller 1998).

If the ratio of aggregate adjusted dividends to GDP is unlikely to change substantially, there are three ways out of the internal inconsistency between the market's current value and OCACT's assumptions for economic growth and stock returns. One can:

- Assume higher GDP growth, which would decrease the implausibility of the calculations described above for either the ratio of market value to GDP or the steady state under the Gordon equation. (The possibility of more rapid GDP growth is not explored further in this article.[31])
- Adopt a long-run stock return that is considerably less than 7.0 percent.
- Lower the rate of return during an intermediate period so that a 7.0 percent return could be applied to a lower market value base thereafter.

A combination of the latter two alternatives is also possible.

In considering the prospect of a near-term market decline, the Gordon equation can be used to compute the magnitude of the drop required over, for example, the next 10 years in order for stock returns to average 7.0 percent over the remaining 65 years of OCACT's projection period (see Appendix B). A long-run return of 7.0 percent would require a drop in real prices of between 21 percent and 55 percent, depending on the assumed value of adjusted dividends (Table 3).[32] That calculation is relatively sensitive to the assumed rate of return—for example, with a long-run return of 6.5 percent, the required drop in the market falls to a range of 13 percent to 51 percent.[33]

The two different ways of restoring consistency—a lower stock return in all years or a near-term decline followed by a return to the historical yield—have different implications for Social Security finances. To illustrate the difference, consider the contrast between a scenario with a steady yield of 4.25 percent derived by using current values for the Gordon equation as described above (the steady-state scenario) and a scenario in which stock prices drop by half immediately and the yield on stocks is 7.0 percent thereafter (the market-correction scenario).[34] First, dollars newly invested in the future (that is, after any drop in share prices) earn only 4.25 percent per year under the steady-state scenario, compared with 7.0 percent per year under the market-correction scenario. Second, even for dollars currently in the market, the long-run yield differs under the two scenarios when the returns on stocks are being reinvested. Under the steady-state scenario, the yield on dollars currently in the market is 4.25 percent per year over any projected time period; under the market-correction scenario, the annual rate of return depends on the time horizon used for the calculation.[35] After one year, the latter scenario has a rate of return of –46 percent. By the end of 10 years, the annual rate of return with the latter scenario is –0.2 percent; by the end of 35 years, 4.9 percent; and by the end of 75 years, 6.0 percent. Proposals for Social Security generally envision a gradual buildup of stock investments, which suggests that those investments would fare better under the market-correction scenario. The importance of the difference between scenarios depends also on the choice of additional changes to Social Security, which affect how long the money can stay invested until it is needed to pay benefits.

Given the different impacts of these scenarios, which one is more likely to occur? The key issue is whether the current stock

Table 3.
Required percentage decline in real stock prices over the next 10 years to justify a return of 7.0, 6.5, and 6.0 percent thereafter

Adjusted dividend yield	Percentage decline to justify a long-run return of— 7.0	6.5	6.0
2.0	55	51	45
2.5	44	38	31
3.0	33	26	18
3.5	21	13	4

Source: Author's calculations.

Note: Derived from the Gordon formula. Dividends are assumed to grow in line with gross domestic product (GDP), which the Office of the Chief Actuary (OCACT) assumes is 2.0 percent over the next 10 years. For long-run GDP growth, OCACT assumes 1.5 percent.

market is overvalued in the sense that rates of return are likely to be lower in the intermediate term than in the long run. Economists have divergent views on this issue.

One possible conclusion is that current stock prices signal a significant drop in the long-run required equity premium. For example, Glassman and Hassett (1999) have argued that the equity premium will be dramatically lower in the future than it has been in the past, so that the current market is not overvalued in the sense of signaling lower returns in the near term than in the long run.[36] Indeed, they even raise the possibility that the market is "undervalued" in the sense that the rate of return in the intermediate period will be higher than in the long run, reflecting a possible continuing decline in the required equity premium. If their view is right, then a 7.0 percent long-run return, together with a 4.0 percent equity premium, would be too high.

Others argue that the current stock market values include a significant price component that will disappear at some point, although no one can predict when or whether it will happen abruptly or slowly. Indeed, Campbell and Shiller (1998) and Cochrane (1997) have shown that when stock prices (normalized by earnings, dividends, or book values) have been far above historical ratios, the rate of return over the following decade has tended to be low, and the low return is associated primarily with the price of stocks, not the growth of dividends or earnings.[37] Thus, to project a steady rate of return in the future, one needs to argue that this historical pattern will not repeat itself. The values in Table 3 are in the range suggested by the historical relationship between future stock prices and current price-earnings and price-dividend ratios (see, for example, Campbell and Shiller 1998).

Therefore, either the stock market is overvalued and requires a correction to justify a 7.0 percent return thereafter, or it is correctly valued and the long-run return is substantially lower than 7.0 percent. (Some combination of the two is also possible.) Under either scenario, stock returns would be lower than 7.0 percent for at least a portion of the next 75 years. Some evidence suggests, however, that investors have not adequately considered that possibility.[38] The former view is more convincing, since accepting the "correctly valued" hypothesis implies an implausibly small long-run equity premium. Moreover, when stock values (compared with earnings or dividends) have been far above historical ratios, returns over the following decade have tended to be low. Since this discussion has no direct bearing on bond returns, assuming a lower return for stocks over the near or long term also means assuming a lower equity premium.

In short, given current stock values, a constant 7.0 percent return is not consistent with OCACT's projected GDP growth.[39] However, OCACT could assume lower returns for a decade, followed by a return equal to or about 7.0 percent.[40] In that case, OCACT could treat equity returns as it does Treasury rates, using different projection methods for the first 10 years and for the following 65. This conclusion is not meant to suggest that anyone is capable of predicting the timing of annual stock returns, but rather that this is an approach to financially consistent assumptions. Alternatively, OCACT could adopt a lower rate of return for the entire 75-year period.

Marginal Product of Capital and Slow Growth

In its long-term projections, OCACT assumes a slower rate of economic growth than the U.S. economy has experienced over an extended period. That projection reflects both the slowdown in labor force growth expected over the next few decades and the slowdown in productivity growth since 1973.[41] Some critics have suggested that slower growth implies lower projected rates of return on both stocks and bonds, since the returns to financial assets must reflect the returns on capital investment over the long run. That issue can be addressed by considering either the return to stocks directly, as discussed above, or the marginal product of capital in the context of a model of economic growth.[42]

For the long run, the returns to financial assets must reflect the returns on the physical assets that support the financial assets. Thus, the question is whether projecting slower economic growth is a reason to expect a lower marginal product of capital. As noted above, this argument speaks to rates of return generally, not necessarily to the equity premium.

The standard (Solow) model of economic growth implies that slower long-run economic growth with a constant savings rate will yield a lower marginal product of capital, and the relationship may be roughly point-for-point (see Appendix C). However, the evidence suggests that savings rates are not unaffected by growth rates. Indeed, growth may be more important for savings rates than savings are for growth rates. Bosworth and Burtless (1998) have observed that savings rates and long-term rates of income growth have a persistent positive association, both across countries and over time. That observation suggests that if future economic growth is slower than in the past, savings will also be lower. In the Solow model, low savings raise the marginal product of capital, with each percentage-point decrease in the savings rate increasing the marginal product by roughly one-half of a percentage point in the long run. Since growth has fluctuated in the past, the stability in real rates of return to stocks, as shown in Table 1, suggests an offsetting savings effect, preserving the stability in the rate of return.[43]

Focusing directly on demographic structure and the rate of return rather than on labor force growth and savings rates, Poterba (1998) does not find a robust relationship between demographic structure and asset returns. He does recognize the limited power of statistical tests based on the few "effective degrees of freedom" in the historical record. Poterba suggests that the connection between demography and returns is not simple and direct, although such a connection has been raised as a possible reason for high current stock values, as baby boomers save for retirement, and for projecting low future stock values, as they finance retirement consumption. Goyal (1999) estimates equity premium regressions and finds that changes in population age structure add significant explanatory power. Nevertheless, using a vector autoregression approach, his analysis predicts no significant increase in *average* outflows over the next 52 years. That occurs despite the retirement of baby boomers. Thus, both papers reach the same conclusion—that demography is not likely to effect large changes in the long-run rate of return.

Another factor to consider in assessing the connection between growth and rates of return is the increasing openness of the world economy. Currently, U.S. corporations earn income from production and trade abroad, and individual investors, while primarily investing at home, also invest abroad. It is not clear that putting the growth issue in a global context makes much difference. On the one hand, since other advanced economies are also aging, increased economic connections with other advanced countries do not alter the basic analysis. On the other hand, although investment in the less-developed countries may preserve higher rates, it is not clear either how much investment opportunities will increase or how to adjust for political risk. Increasing openness further weakens the argument for a significant drop in the marginal product of capital, but the opportunities abroad may or may not be realized as a better rate of return.

On balance, slower projected growth may reduce the return on capital, but the effect is probably considerably less than one-for-one. Moreover, this argument relates to the overall return to capital in an economy, not just stock returns. Any impact would therefore tend to affect returns on both stocks and bonds similarly, with no directly implied change in the equity premium.[44]

Other Issues

This paper has considered the gross rate of return to equities and the equity premium generally. Two additional issues arise in considering the prospect of equity investment for Social Security: how gross returns depend on investment strategy and how they differ from net returns; and the degree of risk associated with adding stock investments to a current all-bond portfolio.

Gross and Net Returns

A gross rate of return differs from a net return because it includes transactions costs such as brokerage charges, bid-ask spreads, and fees for asset management.[45]

If the Social Security trust fund invests directly in equities, the investment is likely to be in an index fund representing almost all of the equities outstanding in the United States. Thus, the analysis above holds for that type of investment. Although some critics have expressed concern that political influence might cause deviations from a broad-based indexing strategy, the evidence suggests that such considerations would have little impact on the expected rate of return (Munnell and Sundén 1999).

If the investment in stocks is made through individual accounts, then individuals may be given some choice either about the makeup of stock investment or about varying the mix of stocks and bonds over time. In order to consider the rate of return on stocks held in such individual accounts, one must consider the kind of portfolio choices individuals might make, both in the composition of the stock portfolio and in the timing of purchases and sales. Given the opportunity, many individuals would engage in numerous transactions, both among stocks and between stocks and other assets (attempts to time the market).

The evidence suggests that such transactions reduce gross returns relative to risks, even before factoring in transactions costs (Odean 1998). Therefore, both the presence of individual accounts with choice and the details of their regulation are likely to affect gross returns. On average, individual accounts with choice are likely to have lower gross returns from stocks than would direct trust fund investment.

Similarly, the cost of administration as a percentage of managed assets varies depending on whether there are individual accounts and how they are organized and regulated (National Academy of Social Insurance 1998; Diamond 2000). Estimates of that cost vary from 0.5 basis points for direct trust fund investment to 100 to150 basis points for individually organized individual accounts, with government-organized individual accounts somewhere in between.

Investment Risk of Stocks

The Office of the Chief Actuary's projections are projections of plausible long-run scenarios (ignoring fluctuations). As such, they are useful for identifying a sizable probability of future financial needs for Social Security. However, they do not address different probabilities for the trust fund's financial condition under different policies.[46] Nor are they sufficient for normative evaluation of policies that have different distributional or risk characteristics.

Although investment in stocks entails riskiness in the rate of return, investment in Treasury bonds also entails risk. Therefore, a comparison of those risks should consider the distribution of outcomes—concern about risk should not be separated from the compensation for bearing risk. That is, one needs to consider the probabilities of both doing better and doing worse as a result of holding some stocks. Merely observing that stocks are risky is an inadequate basis for policy evaluations. Indeed, studies of the historical pattern of returns show that portfolio risk decreases when some stocks are added to a portfolio consisting only of nominal bonds (Siegel 1998). Furthermore, many risks affect the financial future of Social Security, and investing a small portion of the trust fund in stocks is a small risk for the system as a whole relative to economic and demographic risks (Thompson 1998).

As long as the differences in risk and expected return are being determined in a market and reflect the risk aversion of market participants, the suitability of the trust fund's portfolio can be considered in terms of whether Social Security has more or less risk aversion than current investors. Of course, the "risk aversion" of Social Security is a derived concept, based on the risks to be borne by future beneficiaries and taxpayers, who will incur some risk whatever portfolio Social Security holds. Thus, the question is whether the balance of risks and returns looks better with one portfolio than with another. The answer is somewhat complex, since it depends on how policy changes in taxes and benefits respond to economic and demographic outcomes. Nevertheless, since individuals are normally advised to hold at least some stocks in their own portfolios, it seems appropriate for Social Security to also hold some stocks when investing on their behalf, at least in the long run, regardless of the rates of return used for projection purposes (Diamond and Geanakoplos 1999).[47]

Conclusion

Of the three main bases for criticizing OCACT's assumptions, by far the most important one is the argument that a constant 7.0 percent stock return is not consistent with the value of today's stock market and projected slow economic growth. The other two arguments—pertaining to developments in financial markets and the marginal product of capital—have merit, but neither suggests a dramatic change in the equity premium.

Given the high value of today's stock market and an expectation of slower economic growth in the future, OCACT could adjust its stock return projections in one of two ways. It could assume a decline in the stock market sometime over the next decade, followed by a 7.0 percent return for the remainder of the projection period. That approach would treat equity returns like Treasury rates, using different short- and long-run projection methods for the first 10 years and the following 65 years. Alternatively, OCACT could adopt a lower rate of return for the entire 75-year period. That approach may be more acceptable politically, but it obscures the expected pattern of returns and may produce misleading assessments of alternative financing proposals, since the appropriate uniform rate to use for projection purposes depends on the investment policy being evaluated.

Notes

Peter Diamond is Institute Professor at the Massachusetts Institute of Technology, where he has taught since 1966. He is a member of the Board of the National Academy of Social Insurance, where he has been President, Chair of the Board, and Chair of the Panel on Privatization of Social Security. He has written on public finance, macroeconomics, and the economics of uncertainty.

Acknowledgments: The author is grateful to John Campbell, Alicia Munnell, and Jim Poterba for extended discussions and to Andy Abel, Dean Baker, Olivier Blanchard, John Cochrane, Andy Eschtruth, Steve Goss, Joyce Manchester, Peter Orszag, Bernie Saffran, Jeremy Siegel, Tim Smeeding, Peter Temin, and Joe White for helpful comments. The views and remaining errors are those of the author.

[1] This 7.0 percent real rate of return is gross of administrative charges.

[2] To generate short-run returns on stocks, the Social Security Administration's Office of the Chief Actuary (OCACT) multiplied the ratio of one plus the ultimate yield on stocks to one plus the ultimate yield on bonds by the annual bond assumptions in the short run.

[3] An exception was the use of 6.75 percent for the President's proposal evaluated in a memorandum on January 26, 1999.

[4] This report is formally called the *1999 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.*

[5] For OCACT's short-run bond projections, see Table II.D.1 in the 1999 Social Security Trustees Report.

[6] This article was written in the summer of 1999 and uses numbers appropriate at the time. The 2000 Trustees Report uses the same assumptions of 6.3 percent for the nominal interest rate and 3.3 percent for the annual percentage change in the consumer price index. The real wage is assumed to grow at 1.0 percent, as opposed to 0.9 percent in the 1999 report.

[7] See, for example, Baker (1999a) and Baker and Weisbrot (1999). This article only considers return assumptions given economic growth assumptions and does not consider growth assumptions.

[8] This article does not analyze the policy issues related to stock market investment either by the trust fund or through individual accounts. Such an analysis needs to recognize that higher expected returns in the U.S. capital market come with higher risk. For the issues relevant for such a policy analysis, see National Academy of Social Insurance (1998).

[9] Ideally, one would want the yield on the special Treasury bonds held by Social Security. However, this article simply refers to published long-run bond rates.

[10] Because annual rates of return on stocks fluctuate so much, a wide band of uncertainty surrounds the best statistical estimate of the average rate of return. For example, Cochrane (1997) notes that over the 50 years from 1947 to 1996, the excess return of stocks over Treasury bills was 8 percent, but, assuming that annual returns are statistically independent, the standard statistical confidence interval extends from 3 percent to 13 percent. Using a data set covering a longer period lowers the size of the confidence interval, provided one is willing to assume that the stochastic process describing rates of return is stable for the longer period. This article is not concerned with that uncertainty, only with the appropriate rate of return to use for a central (or intermediate) projection. For policy purposes, one must also look at stochastic projections (see, for example, Copeland, VanDerhei, and Salisbury 1999; and Lee and Tuljapurkar 1998). Despite the value of stochastic projections, OCACT's central projection plays an important role in thinking about policy and in the political process. Nevertheless, when making a long-run projection, one must realize that great uncertainty surrounds any single projection and the relevance of returns in any short period of time.

[11] Table 2 also shows the equity premiums relative to Treasury bills. Those numbers are included only because they arise in other discussions; they are not referred to in this article.

[12] For determining the equity premium shown in Table 2, the rate of return is calculated assuming that a dollar is invested at the start of a period and the returns are reinvested until the end of the period. In contrast to that geometric average, an arithmetic average is the average of the annual rates of return for each of the years in a period. The arithmetic average is larger than the geometric average. Assume, for example, that a dollar doubles in value in year 1 and then halves in value from year 1 to year 2. The geometric average over the 2-year period is zero; the arithmetic average of +100 percent and –50 percent annual rates of return is +25 percent. For projection purposes, one looks for an estimated rate of return that is suitable for investment over a long period. Presumably the best approach would be to take the arithmetic average of the rates of return that were each the geometric average for different historical periods of the same length as

the average investment period within the projection period. That calculation would be close to the geometric average, since the variation in 35- or 40-year geometric rates of return, which is the source of the difference between arithmetic and geometric averages, would not be so large.

[13] In considering recent data, some adjustment should be made for bond rates being artificially low in the 1940s as a consequence of war and postwar policies.

[14] Also relevant is the fact that the real rate on 30-year Treasury bonds is currently above 3.0 percent.

[15] Finance theory relates the willingness to hold alternative assets to the expected risks and returns (in real terms) of the different assets, recognizing that expectations about risk and return are likely to vary with the time horizon of the investor. Indeed, time horizon is an oversimplification, since people are also uncertain about when they will want to have access to the proceeds of those investments. Thus, finance theory is primarily about the difference in returns to different assets (the equity premium) and needs to be supplemented by other analyses to consider the expected return to stocks.

[16] With Treasury bonds, investors can easily project future nominal returns (since default risk is taken to be virtually zero), although expected real returns depend on projected inflation outcomes given nominal yields. With inflation-protected Treasury bonds, investors can purchase bonds with a known real interest rate. Since those bonds were introduced only recently, they do not play a role in interpreting the historical record for projection purposes. Moreover, their importance in future portfolio choices is unclear.

[17] In theory, for determining asset prices at which markets clear, one wants to consider marginal investments. Those investments are made up of a mix of marginal portfolio allocations by all investors and by marginal investors who become participants (or nonparticipants) in the stock and/or bond markets.

[18] This conclusion does not contradict the Modigliani-Miller theorem. Different firms with the same total return distributions but different amounts of debt outstanding will have the same total value (stock plus bond) and so the same total expected return. A firm with more debt outstanding will have a higher expected return on its stock in order to preserve the total expected return.

[19] Consideration of equilibrium suggests an alternative approach to analyzing the historical record. Rather than looking at realized rates of return, one could construct estimates of expected rates of return and see how they have varied in the past. That approach has been taken by Blanchard (1993). He concluded that the equity premium (measured by expectations) was unusually high in the late 1930s and 1940s and, since the 1950s, has experienced a long decline from that unusually high level. The high realized rates of return over this period are, in part, a consequence of a decline in the equity premium needed for people to be willing to hold stocks. In addition, the real expected returns on bonds have risen since the 1950s, which should have moderated the impact of a declining equity premium on expected stock returns. Blanchard examines the importance of inflation expectations and attributes some of the recent trend to a decline in expected inflation. He concluded that the premium in 1993 appeared to be around 2 percent to 3 percent and would probably not move much if inflation expectations remain low. He also concluded that decreases in the equity premium were likely to involve both increases in expected bond rates and decreases in expected rates of return on stocks.

[20] If current cash returns to stockholders are expected to grow at rate g, with projected returns discounted at rate r, this fundamental value is the current return divided by $(r - g)$. If r is smaller, fluctuations in long-run projections of g result in larger fluctuations in the fundamental value.

[21] Several explanations have been put forth, including: (1) the United States has been lucky, compared with stock investment in other countries, and realized returns include a premium for the possibility that the U.S. experience might have been different; (2) returns to actual investors are considerably less than the returns on indexes that have been used in analyses; and (3) individual preferences are different from the simple models that have been used in examining the puzzle.

[22] The timing of realized returns that are higher than required returns is somewhat more complicated, since recognizing and projecting such a trend will tend to boost the price of equities when the trend is recognized, not when it is realized.

[23] Nonprofit institutions, such as universities, and defined benefit plans for public employees now hold more stock than in the past. Attributing the risk associated with that portfolio to the beneficiaries of those institutions would further expand the pool sharing in the risk.

[24] More generally, the equity premium depends on the investment strategies being followed by investors.

[25] This tendency, known as mean reversion, implies that a short period of above-average stock returns is likely to be followed by a period of below-average returns.

[26] To quantify the importance of these developments, one would want to model corporate behavior as well as investor behavior. A decline in the equity premium reflects a drop to corporations in the "cost of risk" in the process of acquiring funds for risky investment. If the "price per unit of risk" goes down, corporations might respond by selecting riskier investments (those with a higher expected return), thereby somewhat restoring the equity premium associated with investing in corporations.

[27] In considering the return to an individual from investing in stocks, the return is made up of dividends and a (possible) capital gain from a rise in the value of the shares purchased. When considering the return to all investment in stocks, one needs to consider the entire cash flow to stockholders, including dividends and net share repurchases by the firms. That suggests two methods of examining the consistency of any assumed rate of return on stocks. One is to consider the value of all stocks outstanding. If one assumes that the value of all stocks outstanding grows at the same rate as the economy (in the long run), then the return to all stocks outstanding is that rate of growth plus the sum of dividends and net share repurchases, relative to total share value. Alternatively, one can consider ownership of a single share. The assumed rate of return minus the rate of dividend payment then implies a rate of capital gain on the single share. However, the relationship between the growth of value of a single share and the growth of the economy depends on the rate of share repurchase. As shares are being repurchased, remaining shares should grow in value relative to the growth of the economy. Either approach can be calculated in a consistent manner. What must be avoided is an inconsistent mix, considering only dividends and also assuming that the value of a single share grows at the same rate as the economy.

[28] Gordon (1962). For an exposition, see Campbell, Lo, and MacKinlay (1997).

[29] The implausibility refers to total stock values, not the value of single shares—thus, the relevance of net share repurchases. For example, Dudley and others (1999) view a steady equity premium in the range of 1.0 percent to 3.0 percent as consistent with current stock prices and their projections. They assume 3.0 percent GDP growth and a 3.5 percent real bond return, both higher than the assumptions used by OCACT. Wadhwani (1998) finds that if the S&P 500 is correctly valued, he has to assume a negative risk premium. He considers various adjustments that lead to a higher premium, with his "best guess" estimate being 1.6 percent. That still seems too low.

[30] Dudley and others (1999) report a current dividend yield on the Wilshire 5000 of 1.3 percent. They then make an adjustment that is equivalent to adding 80 basis points to that rate for share repurchases, for which they cite Campbell and Shiller (1998). Wadhwani (1998) finds a current expected dividend yield of 1.65 percent for the S&P 500, which he adjusts to 2.55 percent to account for share repurchases. For a discussion of share repurchases, see Cole, Helwege, and Laster (1996).

[31] Stock prices reflect investors' assumptions about economic growth. If their assumptions differ from those used by OCACT, then it becomes difficult to have a consistent projection that does not assume that investors will be surprised.

[32] In considering these values, note the observation that a fall of 20 percent to 30 percent in advance of recessions is typical for the U.S. stock market (Wadhwani 1998). With OCACT assuming a 27 percent rise in the price level over the next decade, a 21 percent decline in real stock prices would yield the same nominal prices as at present.

[33] The importance of the assumed growth rate of GDP can be seen by redoing the calculations in Table 3 for a growth rate that is one-half of a percent larger in both the short and long runs. Compared with the original calculations, such a change would increase the ratios by 16 percent.

[34] Both scenarios are consistent with the Gordon formula, assuming a 2.75 percent adjusted dividend yield (without a drop in share prices) and a growth of dividends of 1.5 percent per year.

[35] With the steady-state scenario, a dollar in the market at the start of the steady state is worth 1.0425^t dollars t years later, if the returns are continuously reinvested. In contrast, under the market-correction scenario, a dollar in the market at the time of the drop in prices is worth $(1/2)(1.07^t)$ dollars t years later.

[36] The authors appear to assume that the Treasury rate will not change significantly, so that changes in the equity premium and in the return to stocks are similar.

[37] One could use equations estimated on historical prices to check the plausibility of intermediate-run stock values with the intermediate-run values needed for plausibility for the long-run assumptions. Such a calculation is not considered in this article. Another approach is to consider the value of stocks relative to the replacement cost of the capital that corporations hold, referred to as Tobin's q. That ratio has fluctuated considerably and is currently unusually high. Robertson and Wright (1998) have analyzed the ratio and concluded that a cumulative real decline in the stock market over the first decades of the 21st century has a high probability.

[38] As Wadhwani (1998, p. 36) notes, "Surveys of individual investors in the United States regularly suggest that they expect returns above 20 percent, which is obviously unsustainable. For example, in a survey conducted by Montgomery Asset Management in 1997, the typical mutual fund investor expected annual returns from the stock market of 34 percent over the next 10 years! Most U.S. pension funds operate under actuarial assumptions of equity returns in the 8-10 percent area, which, with a dividend yield under 2 percent and nominal GNP growth unlikely to exceed 5 percent, is again, unsustainably high."

[39] There is no necessary connection between the rate of return on stocks and the rate of growth of the economy. There is a connection among the rate of return on stocks, the current stock prices, dividends relative to GDP, and the rate of growth of the economy.

[40] The impact of such a change in assumptions on actuarial balance depends on the amount that is invested in stocks in the short term relative to the amount invested in the long term. The levels of holdings at different times depend on both the speed of initial investment and whether stock holdings are sold before very long (as would happen with no other policy changes) or whether, instead, additional policies are adopted that result in a longer holding period, possibly including a sustained sizable portfolio of stocks. Such an outcome would follow if Social Security switched to a sustained level of funding in excess of the historical long-run target of just a contingency reserve equal to a single year's expenditures.

[41] "The annual rate of growth in total labor force decreased from an average of about 2.0 percent per year during the 1970s and 1980s to about 1.1 percent from 1990 to 1998. After 1998 the labor force is projected to increase about 0.9 percent per year, on average, through 2008, and to increase much more slowly after that, ultimately reaching 0.1 percent toward the end of the 75-year projection period" (Social Security Trustees Report, p. 55). "The Trustees assume an intermediate trend growth rate of labor productivity of 1.3 percent per year, roughly in line with the average rate of growth of productivity over the last 30 years" (Social Security Trustees Report, p. 55).

[42] Two approaches are available to answer this question. Since the Gordon formula, given above, shows that the return to stocks equals the adjusted dividend yield plus the growth of stock prices, one needs to consider how the dividend yield is affected by slower growth. In turn, that relationship will depend on investment levels relative to corporate earnings. Baker (1999b) makes such a calculation, which is not examined here. Another approach is to consider the return on physical capital directly, which is the one examined in this article.

[43] Using the Granger test of causation (Granger 1969), Carroll and Weil (1994) find that growth causes saving but saving does not cause growth. That is, changes in growth rates tend to precede changes in savings rates but not vice versa. For a recent discussion of savings and growth, see Carroll, Overland, and Weil (2000).

[44] One can also ask how a change in policy designed to build and maintain a larger trust fund in a way that significantly increases national saving might affect future returns. Such a change would plausibly tend to lower rates of return. The size of that effect depends on the size of investment increases relative to available investment opportunities, both in the United States and worldwide. Moreover, it depends on the response of private saving to the policy, including the effect that would come through any change in the rate of return. There is plausibly an effect here, although this article does not explore it. Again, the argument speaks to the level of rates of return generally, not to the equity premium.

[45] One can also ask how changed policies might affect future returns. A change in portfolio policy that included stocks (whether in the trust fund or in individual accounts) would plausibly lower the equity premium somewhat. That effect could come about through a combination of a rise in the Treasury rate (thereby requiring a change

in tax and/or expenditure policy) and a fall in expected returns on stocks. The latter depends on both the underlying technology of available returns to real investments and the effect of portfolio policy on national saving. At this time, research on this issue has been limited, although it is plausible that the effect is not large (Bohn 1998; Abel 1999; Diamond and Geanakoplos 1999).

[46] For stochastic projections, see Copeland, VanDerhei, and Salisbury (1999); and Lee and Tuljapurkar (1998). OCACT generally provides sensitivity analysis by doing projections with several different rates of return on stocks.

[47] Cochrane (1997, p. 32) reaches a similar conclusion relative to individual investment: "We could interpret the recent run-up in the market as the result of people finally figuring out how good an investment stocks have been for the last century, and building institutions that allow wise participation in the stock market. If so, future returns are likely to be much lower, but there is not much one can do about it but sigh and join the parade."

References

Abel, Andrew B. 1999. *The Social Security Trust Fund, The Riskless Interest Rate, and Capital Accumulation.* NBER Working Paper 6991. Cambridge, Mass.: National Bureau of Economic Research.

Advisory Council on Social Security. 1997. *Report of the 1994-1996 Advisory Council on Social Security*. Vol. 1, *Findings and Recommendations*. Washington, D.C.

Baker, Dean. 1998. "Saving Social Security in Three Steps." Briefing Paper, Economic Policy Institute, Washington, D.C.

_______. 1999a. "Two Cheers for Clinton's Social Security Plan." *The American Prospect,* Vol. 10, No. 44, pp. 82-85.

_______.1999b. Letter to Martin Feldstein, May 15. Available at www.preamble.org.

Baker, Dean, and Mark Weisbrot. 1999. *Social Security: The Phony Crisis*. Chicago, Ill.: University of Chicago Press.

Blanchard, Olivier J. 1993. "Movements in the Equity Premium." *Brookings Papers on Economic Activity*, No. 2, pp. 75-138.

Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds. 1999. *1999 Annual Report.* Washington, D.C.: U.S. Government Printing Office.

Bohn, Henning. 1998. "Social Security Reform and Financial Markets." In Federal Reserve Bank of Boston, *Social Security Reform: Links to Saving, Investment, and Growth*. Boston, Mass.: Federal Reserve Bank of Boston.

Bosworth, Barry, and Gary Burtless. 1998. "Social Security Reform in a Global Context." In Federal Reserve Bank of Boston, *Social Security Reform: Links to Saving, Investment, and Growth*. Boston, Mass.: Federal Reserve Bank of Boston.

Brav, Alon, and Christopher C. Geczy. 1996. "An Empirical Resurrection of the Simple Consumption CAPM with Power Utility." Working Paper, University of Chicago.

Campbell, John Y.; Andrew W. Lo; and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, N.J.: Princeton University Press.

Campbell, John Y., and Robert J. Shiller. 1998. "Valuation Ratios and the Long-Run Stock Market Outlook." *Journal of Portfolio Management*, Vol. 24, No. 2, pp. 11-26.

Campbell, John Y., and Luis M. Viceira. 1999. "Consumption and Portfolio Decisions When Expected Returns Are Time Varying." *Quarterly Journal of Economics* (May), pp. 433-495.

Carroll, Christopher D.; Jody Overland; and David N. Weil. 2000. "Saving and Growth with Habit Formation." *American Economic Review*, Vol. 90, No. 3, pp. 341-355.

Carroll, Christopher D., and David N. Weil. 1994. "Saving and Growth: A Reinterpretation." *Carnegie-Rochester Conference Series on Public Policy,* No. 40, pp. 133-192.

Case, Karl E., and Robert J. Shiller. 1988. "The Behavior of Home Buyers in Boom and Post-Boom Markets." *New England Economic Review* (November/December), pp. 29-46.

Cochrane, John H. 1997. "Where is the Market Going? Uncertain Facts and Novel Theories." *Economic Perspectives,* Vol. 21, Federal Reserve Bank of Chicago. (Also published as NBER Working Paper 6207, Cambridge, Mass.: National Bureau of Economic Research.)

Cole, Kevin; Jean Helwege; and David Laster. 1996. "Stock Market Valuation Indicators: Is This Time Different?" *Financial Analysts Journal,* Vol. 52, No. 3, pp. 56-64.

Copeland, Craig; Jack VanDerhei; and Dallas L. Salisbury. 1999. *Social Security Reform: Evaluating Current Proposals: Latest Results of the EBRI SSASIM2 Policy Simulation Model*. EBRI Issue Brief 210. Washington, D.C.: Employee Benefit Research Institute.

Diamond, Peter. 1999. *Issues in Privatizing Social Security*. Cambridge, Mass.: MIT Press.

_______. 2000. "Administrative Costs and Equilibrium Charges with Individual Accounts." In *Administrative Aspects of Investment-Based Social Security Reform*, edited by John B. Shoven. Chicago and London: University of Chicago Press, pp. 137-162.

Diamond, Peter, and John Geanakoplos. 1999. *Social Security Investment in Equities I: Linear Case*. Working Paper 2. Center for Retirement Research at Boston College, Chestnut Hill, Mass.

Dudley, William; Jan Hatzius; Joanne Hill; Rajib Pal; and Maria Tsu. 1999. "An Analysis of Social Security Trust Fund Investment into Equities." Goldman Sachs Global Economics Paper 18, New York.

Glassman, James K., and Kevin A. Hassett. 1999. *Dow 36,000*. New York: Times Business.

Gordon, Myron. 1962. *The Investment, Financing, and Valuation of the Corporation*. Homewood, Ill.: Irwin.

Gordon, Roger. 1985. "Taxation of Corporate Capital Income: Tax Revenues versus Tax Distributions." *Quarterly Journal of Economics,* Vol. 100, No. 1, pp. 1-27.

Goyal, Amit. 1999. "Demographics, Stock Market Flows, and Stock Returns." Anderson Graduate School of Management, University of California at Los Angeles. Unpublished.

Granger, C.W.J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, Vol. 37, No. 3, pp. 424-438.

Heaton, John, and Deborah Lucas. 2000. "Stock Prices and Fundamentals." In *NBER Macroeconomics Annual,* edited by Ben S. Bernanke and Julio Rotemberg. Cambridge, Mass.: MIT Press.

Investment Company Institute. 1999. *Mutual Fund Fact Book* (May), p. 22. Available at www.ici.org.

Kaplow, Louis. 1994. "Taxation and Risk-taking: A General Equilibrium Perspective." *National Tax Journal,* Vol. 47, pp. 789-798.

Kennickell, Arthur B.; Martha Starr-McCluer; and Annika Sundén. 1997. "Family Finances in the U.S.: Recent Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin,* Vol. 83 (June), pp. 1-24.

Lee, Ronald, and Shripad Tuljapurkar. 1998. "Stochastic Forecasts for Social Security." In *Frontiers in the Economics of Aging*, edited by David Wise. Chicago, Ill.: University of Chicago Press, pp. 393-420.

Liang, J. Nellie, and Steven A. Sharpe. 1999. "Share Repurchases and Employee Stock Options and Their Implications for S&P 500 Share Retirements and Expected Returns." Division of Research and Statistics, Federal Reserve Board, Washington, D.C. Unpublished.

Mehra, Rajnish, and Edward Prescott. 1985. "The Equity Premium Puzzle." *Journal of Monetary Economics*, Vol. 15, pp. 145-161.

Mankiw, N.G., and S. Zeldes. 1991. "The Consumption of Stockholders and Non-Stockholders." *Journal of Financial Economics*, Vol.17, pp. 211-219.

Munnell, Alicia, and Annika Sundén (with the assistance of Cynthia Perry and Ryan Kling). 1999. *Investment Practices of State and Local Pension Funds: Implications for Social Security*. Working Paper 1. Center for Retirement Research at Boston College, Chestnut Hill, Mass.

National Academy of Social Insurance. 1998. *Evaluating Issues in Privatizing Social Security.* Report of the Panel on Privatization of Social Security. Available at www.nasi.org and as Diamond 1999.

Odean, Terrance. 1998. "Do Investors Trade Too Much?" Graduate School of Management,University of California, Davis. Unpublished.

Philips, Thomas K. 1999. "Why Do Valuation Ratios Forecast Long-Run Equity Returns?" *Journal of Portfolio Management*, Vol. 25, No. 3, pp. 39-44.

Poterba, James M. 1998. "Population Age Structure and Asset Returns: An Empirical Investigation." Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass. Unpublished.

Poterba, James M., and Andrew A. Samwick. 1995. "Stock Ownership Patterns, Stock Market Fluctuations, and Consumption." *Brookings Papers on Economic Activity*, No. 2, pp. 295-372.

Rea, John D., and Brian K. Reid. 1998. *Trends in the Ownership Cost of Equity Mutual Funds*. Perspective, Vol. 4, No. 3. Washington, D.C.: Investment Company Institute.

_______.1999. *Total Shareholder Cost of Bond and Money Market Mutual Funds*. Perspective, Vol. 5, No. 3. Washington, D.C.: Investment Company Institute.

Robertson, Donald, and Stephen Wright. 1998. "The Good News and the Bad News about Long-Run Stock Market Returns." Faculty of Economics and Politics, University of Cambridge. Unpublished.

Samuelson, Paul A. 1989. "A Case at Last for Age-Phased Reduction in Equity." *Proceedings of the National Academy of Sciences*, Vol. 86, pp. 9048-9051.

Shafir, Eldar; Peter Diamond; and Amos Tversky. 1997. "Money Illusion." *Quarterly Journal of Economics*, Vol. 112, No. 2, pp. 341-374.

Sharpe, Steven. 1999. "Stock Prices, Expected Returns, and Inflation." Federal Reserve Board, Washington, D.C. Unpublished.

Siegel, Jeremy J. 1998. *Stocks for the Long Run*. New York: McGraw Hill, p. 318.

_______. 1999. "The Shrinking Equity Premium: Historical Facts and Future Forecasts." Wharton School, University of Pennsylvania, Philadelphia, Pa. Unpublished.

Thompson, Lawrence. 1998. *Older and Wiser: The Economics of Public Pensions*. Washington, D.C.: Urban Institute Press.

Vissing-Jorgensen, Annette. 1997. "Limited Stock Market Participation." Department of Economics, Massachusetts Institute of Technology, Cambridge, Mass. Unpublished.

Wadhwani, Sushil B. 1998. "The U.S. Stock Market and the Global Economic Crisis." Tudor Investment Corporation, New York. Unpublished.

Appendix A: Alternative Method for Determining the Ratio of Stock Value to GDP

Variables

r …… rate of return on stocks

g …… rate of growth of both GDP and dividends

a …… adjusted dividend yield at time 0

$P(t)$ … aggregate stock value at time t

$Y(t)$ … GDP at time t

$D(t)$ … dividends at time t

Equations

$$Y(t) = Y(0)e^{gt}$$

$$D(t) = D(0)e^{gt} = aP(0)e^{gt}$$

$$dP(t)/dt = rP - D(t) = rP - aP(0)e^{gt}$$

Solving the differential equation, we have:

$$\begin{aligned} P(t) &= P(0)\{(r-g-a)e^{rt} + ae^{gt}\}/(r-g) \\ &= P(0)\{e^{rt} - (a/(r-g))(e^{rt} - e^{gt})\} \end{aligned}$$

Taking the ratio of prices to GDP, we have:

$$\begin{aligned} P(t)/Y(t) &= \{P(0)/Y(0)\}\{(r-g-a)e^{(r-g)t} + a\}/(r-g) \\ &= \{P(0)/Y(0)\}\{(e^{(r-g)t} - (a/(r-g))(e^{(r-g)t} - 1)\} \end{aligned}$$

Consistent with the Gordon formula, a constant ratio of P/Y (that is, a steady state) follows from $r = g + a$. As a non-steady-state example—with values of .07 for r, .015 for g, and .03 for a—$P(75)/Y(75) = 28.7P(0)/Y(0)$.

Appendix B: Calculation Using the Gordon Equation

In discrete time, once we are in a steady state, the Gordon growth model relates a stock price P at time t to the expected dividend D in the following period, the rate of growth of dividends G, and the rate of return on the stock R. Therefore, we have:

$$P_t = D_{t+1}/(R-G) = (1+G)D_t/(R-G)$$

We denote values after a decade (when we are assumed to be in a steady state) by P' and D' and use an "adjusted" initial dividend that starts at a ratio X times current stock prices. Thus, we assume that dividends grow at the rate G from the "adjusted" current value for 10 years, where G coincides with GDP growth over the decade. We assume that dividends grow at G' thereafter, which coincides with long-run GDP growth. Thus, we have:

$$\begin{aligned} P'/P &= (1+G')D'/((R-G')P) \\ &= (1+G')D(1+G)^{10}/((R-G')P) \\ &= X(1+G')(1+G)^{10}/(R-G') \end{aligned}$$

For the basic calculation, we assume that R is .07, G is .02, G' is .015. In this case, we have:

$$P'/P = 22.5X$$

Thus, for initial ratios of adjusted dividends to stock prices of .02, .025, .03, and .035, P'/P equals .45, .56, .67 and .79, respectively. Subtracting those numbers from 1 yields the required decline in the real value of stock prices as shown in the first column of Table 3. Converting them into nominal values by multiplying by 1.27, we have values of .57, .71, and .86. If the long-run stock return is assumed to be 6.5 percent instead of 7.0 percent, the ratio P'/P is higher and the required decline is smaller. Increasing GDP growth also reduces the required decline. Note that the required declines in stock values in Table 3 is the decline in real values; the decline in nominal terms would be less.

Appendix C: A Cobb-Douglas Solow Growth Model in Steady State

Variables

Y ……… output
K ……… capital
L ………. labor
a ………. growth rate of Solow residual
g ………. growth rate of both K and Y
n ………. growth rate of labor
b ………. share of labor
s ………. savings rate
c ………. depreciation rate
$MP(K)$ … marginal product of capital

Equations

$\log[Y] = at + b\log[L] + (1-b)\log[K]$
$(dL/dt)/L = n$
$(dY/dt)/Y = (dK/dt)/K = g$
$dK/dt = sY - cK$
$(dK/dt)/K = sY/K - c$
$Y/K = (g + c)/s$
$MP(K) = (1-b)Y/K = (1-b)(g + c)/s$
$g = a + bn + (1-b)g$
$g = (a + bn)/b$
$MP(K) = (1-b)\{(a + bn)/(bs) + c/s\}$
$dMP(K)/da = (1-b)/(bs)$
$dg/da = 1/b$

Assume that the share of labor is .75 and the gross savings rate is .2. Then the change in the marginal product of capital from a change in the growth rate is:

$$dMP(K)/dg = (dMP(K)/da)/(dg/da) = (1-b)/s == .25/.2$$

(Note that these are gross savings, not net savings. But the corporate income tax reduces the return to savers relative to the return to corporate capital, so the derivative should be multiplied by roughly 2/3.)

Similarly, we can consider the effect of a slowdown in labor force growth on the marginal product of capital:

$$dMP(K)/dn = (1-b)/s$$

$$dg/dn = 1$$

$$dMP(K)/dg = (dMP(K)/dn)/(dg/dn) = (1-b)/s == .25/.2$$

(This is the same expression as when the slowdown in economic growth comes from a drop in technical progress.)

Turning to the effects of changes in the savings rate, we have:

$$dMP(K)/ds = -MP(K)/s == .5$$

Thus, the savings rate has a large impact on the marginal product of capital as well.

Both of these effects are attenuated to the extent that the economy is open and rates of return in the United States change less because some of the effect occurs abroad.

Estimating the Ex Ante Equity Premium

R. Glen Donaldson
Sauder School of Business
University of British Columbia
glen.donaldson AT sauder.ubc.ca

Mark J. Kamstra
Schulich School of Business
York University
mkamstra AT schulich.yorku.ca

Lisa A. Kramer
Rotman School of Management
University of Toronto
Lkramer AT rotman.utoronto.ca

October 2007

Keywords: equity risk premium; simulated method of moments; SMM

JEL classifications: G12, C13, C15, C22

We have benefited from the suggestions of Wayne Ferson, Ian Garrett, Mark Fisher, Joel Hasbrouck, Raymond Kan, Patrick Kelly, Alan Kraus, Tom McCurdy, Federico Nardari, Cesare Robotti, Jacob Sagi, Tan Wang, participants of the Western Finance Association Meetings, the Northern Finance Association Meetings, the Canadian Econometrics Study Group, the European Econometric Society Meetings, the investment conference of the University of Colorado at Boulder's Burridge Center for Securities Analysis and Valuation, and seminar participants at the Board of Governors of the Federal Reserve System, Emory University, the Federal Reserve Bank of Atlanta, Queen's University, the US Securities & Exchange Commission, and the University of British Columbia. We thank the Social Sciences and Humanities Research Council of Canada for financial support. Any remaining errors are our own. A previous version of this paper was titled "Stare Down the Barrel and Center the Crosshairs: Targeting the Ex Ante Equity Premium." The most recent version of this paper is available from: http://ssrn.com/abstract=945192. Copyright ©2007 by R. Glen Donaldson, Mark Kamstra, and Lisa Kramer. All rights reserved.

Estimating the Ex Ante Equity Premium

Abstract

We find that the true ex ante equity premium very likely lies within 50 basis points of 3.5%. This estimate is similar to values obtained in some recent studies but is considerably more precise. In addition to narrowing the range of plausible ex ante equity premia, we also find that equity premium models that allow for time-variation, breaks, and/or trends are the models that best match the experience of US markets and are the only models not rejected by our specification tests. This suggests that time-variation, breaks, and/or trends are critical features of the equity premium process. Our approach involves simulating the distribution from which interest rates, dividend growth rates, and equity premia are drawn and determining the prices and returns consistent with these distributions. We achieve the narrower range of ex ante equity premium values and the narrower set of plausible models by comparing statistics that arise from our simulations with key financial characteristics of the US economy, including the mean dividend yield, return volatility, and mean return. Our findings are achieved in part with the imposition of more structure than is typically exploited in the literature. In order to mitigate the potential for misspecification with this additional structure, we consider a broad collection of models that variously do or do not incorporate features such as an adjustment in dividend growth rates to account for recently increased share repurchase activity, sampling uncertainty in generating model parameters, and cross-correlation between interest rates, dividend growth rates, and equity premia.

Estimating the Ex Ante Equity Premium

Financial economic theory is often concerned with the premium that investors demand ex ante, when they first decide whether to purchase risky stocks instead of risk-free debt. In contrast, empirical tests of the equity premium often focus on the return investors received ex post.[1] It is well known that estimates of the ex ante equity premium based on ex post data can be very imprecise; such estimates have very wide margins of error, as wide as 1000 basis points in typical studies and 320 basis points in some recent studies. This fact makes it challenging to employ the equity premium estimates for common practical purposes, including evaluating the equity premium puzzle, performing valuation, and conducting capital budgeting. The imprecision of traditional equity premium estimates also makes it difficult to determine if the equity premium has changed over time. Our goals, therefore, are to develop a more precise estimate of the ex ante equity premium and to determine what kind of equity premium model can be supported by the experience of US markets. We accomplish these goals by employing simulation techniques that identify a range of models of the equity premium and the values of the ex ante equity premium that are consistent with values of several key financial statistics that are observed in US market data, including dividend growth rates, interest rates, Sharpe ratios, price-dividend ratios, volatilities, and of course the ex post equity premium.

Our results suggest that the mean ex ante equity premium lies within 50 basis points of 3.5%. These results stand even when we allow for investors' uncertainty about the true state of the world. The tightened bounds are achieved in part with the imposition of more structure than has been commonly employed in the equity premium literature. In order to mitigate the potential for misspecification with this additional structure, we consider a broad collection of models that variously do or do not incorporate features such as a conditionally time-varying equity premium, a downward trend in the equity premium, a structural break in the equity premium, an adjustment in dividend growth rates to account for increased share repurchase activity in the last 25 years, sampling uncertainty in generating model parameters, a range of time series models, and cross-correlation between interest rates, dividend growth rates, and equity premia. We also find that

[1]The equity premium literature is large, continuously growing, and much too vast to fully cite here. For recent work, see Bansal and Yaron (2004), Graham and Harvey (2005), and Jain (2005). For excellent surveys see Kocherlakota (1996), Siegel and Thaler (1997), Mehra and Prescott (2003), and Mehra (2003).

equity premium models that allow for time-variation, breaks, and/or trends in the equity premium process are the models that best match the experience of US markets and are the only models not rejected by our specification tests. This suggests that time-variation, breaks, and/or trends are critical features of the equity premium process, itself an important finding.

We draw on two relatively new techniques in order to provide a more precise estimate of the equity premium than is currently available. The first technique builds on the fundamental valuation dividend discounting method of Donaldson and Kamstra (1996). This technique permits the simulation of fundamental prices, returns, and return volatility for a given ex ante equity premium. Donaldson and Kamstra find that if we allow dividend growth rates and discount rates to be time-varying and dependent, as well as cross-correlated, the fundamental prices and returns that come out of dividend discounting match observed prices and returns, even during extreme events like stock market crashes. The second technique is simulated method of moments (SMM).[2] An attractive feature of SMM is that the estimation of parameters requires only that the model, with a given set of parameters, can generate data. SMM forms estimates of model parameters by using a given model with a given set of parameter values to simulate moments of the data (for instance means or volatilities), measuring the distance between the simulated moments and the actual data moments, and repeating with new parameter values until the parameter values that minimize the (weighted) distance are found.[3] The parameter estimates that minimize this distance are consistent for the true values, are asymptotically normally distributed, and display the attractive feature of permitting tests that can reject misspecified models. The SMM technique has been described as "estimating on one group of moments, testing on another." See Cochrane (2001, Section 11.6). We use SMM rather than GMM because, as we show below, the economic model we use is nonlinear in the parameters and cannot be solved without the use of SMM.

We exploit the dividend discounting method of Donaldson and Kamstra to generate simulated fundamental prices, dividends, returns, and derivative moments such as the mean ex post equity

[2]Simulated method of moments was developed by McFadden (1989) and Pakes and Pollard (1989), and a helpful introduction to the technique is provided in Carrasco and Florens (2002). Examples of papers that employ SMM in an asset pricing context are Duffie and Singleton (1993) and Corradi and Swanson (2005).

[3]The typical implementation of SMM is to weight the moments inversely to their estimated precision; that is minimize the product of the moments weighted by the inverse of the covariance matrix of the moments. This is the approach we adopt.

premium, mean dividend yield, and return volatility for a given ex ante equity premium. We minimize (by choice of the ex ante equity premium) the distance between the simulated moments that the model produces and the moments observed in US stock markets over the past half century. That is, given various characteristics of the US economic experience (such as low interest rates and a high ex post equity premium, high Sharpe ratios and low dividend yields, *etc.*), we determine the range of values of the ex ante equity premium and the set of equity premium models that are most likely to have generated the observed collection of sample moments.

To undertake our study, we consider a broad collection of models, including models with and without conditional time-variation in the equity premium process, with and without trends in the equity premium, with and without breaks in the equity premium, with and without breaks in the dividend growth rate, as well as various autoregressive specifications for dividend growth rates, interest rates, and the equity premium. Virtually every model we consider achieves a minimum distance between the simulated moments and the actual data moments by setting the ex ante equity premium between 3% and 4%, typically very close to 3.5%. That is, the equity premium estimate is very close to 3.5% across our models. Further, the range of ex ante equity premium values that can be supported by the US data for a given model is typically within plus or minus 50 basis points of 3.5%. Our models of fundamentals, which capture the dynamics of actual US dividend and interest rate data, imply that the true ex ante equity premium is 3.5% plus or minus 50 basis points. Simpler models of fundamental valuation, such as the Gordon (1962) constant dividend growth model, are overwhelmingly rejected by the data. Models of the equity premium which do not allow time-variation, trends, or breaks are also rejected by the SMM model specification tests. While we restrict our attention to a stock market index in this study, the technique we employ is more broadly applicable to estimating the equity premium of an individual firm.

In the literature to date, empirical work investigating the equity premium has largely consisted of a series of innovations around a common theme: producing a better estimate of the mean ex ante equity premium. Recent work in the area has included insights such as exploiting dividend yields or earnings yields to provide new, more precise estimates of the return to holding stocks (see Fama and French, 2002, and Jagannathan, McGrattan, and Scherbina, 2000), looking across many countries to account for survivorship issues (see Jorion and Goetzmann, 1999), looking across many countries to decompose the equity premium into dividend growth, price-dividend ratio, dividend yield, and real exchange rate components (see Dimson, Marsh, and Staunton, 2007), modeling equity premium structural breaks in a Bayesian econometric framework (see Pástor and Stambaugh, 2001), or computing out-of-sample forecasts of the distribution of excess returns, allowing for structural breaks which are identified in real time (see Maheu and McCurdy, 2007). Most of this work estimates the ex ante equity premium by considering one moment of the data at a time, typically the mean difference between an estimate of the return to holding equity and a risk free rate, though Maheu and McCurdy (2007) consider higher-order moments of the excess return distribution and Pástor and Stambaugh (2001) incorporate return volatility and direction of price movements through their use of priors.

Unfortunately, the equity premium is still estimated without much precision. Pástor and Stambaugh (2001), exploiting extra information from return volatility and prices, narrow a two standard deviation confidence interval around the value of the ex ante equity premium to plus or minus roughly 280 basis points around a mean premium estimate of roughly 4.8% (a range that spans 2% to 7.6%) and determine that the data strongly support at least one break in the equity premium in the last half century. Fama and French (2002), based on data from 1951 to 2000, provide point estimates of the ex post equity premium of 4.32% (based on earnings growth rate fundamentals) plus or minus roughly 400 basis points (again, two standard deviations) and of 2.55% (based on dividend growth rate fundamentals) plus or minus roughly 160 basis points: a range of approximately 0.95% to 4.15%. That is, the plausible range of equity premia that emerge from Fama and French's study occupy a confidence bound with a width of anywhere from 320 to 800 basis points. Claus and Thomas (2001), like Fama and French (2002), make use of fundamental information to form lower estimates of the ex post equity premium, but their study covers a shorter time period relative to the Fama and French study – 14 years versus 50 years – yielding point estimates that are subject to at least as much variability as the Fama and French estimates.

Not only are the point estimates from the existing literature imprecisely estimated in terms of their standard error, there is also less of an emerging consensus than one would hope. Fama and French (2002) produce point estimates of 2.55% (using dividend yields) and 4.78% (using earnings yields), Pástor and Stambaugh (2001) estimate the equity premium at the end of the 1990s to be 4.8%, and Claus and Thomas (2001) estimate the equity premium to be *no more* than 3%. Welch (2000), surveying academic financial economists, estimates the consensus equity premium to be between 6% and 7% (depending on the horizon). Based on a survey of US CFOs, Graham and Harvey (2005) estimate the ten-year equity premium to be 3.66%. We believe that the lack of consensus across the literature is intimately tied to the imprecision of techniques typically used to estimate the equity premium, such as the simple average excess return. That is, the various estimates cited above all fall within two standard errors of the sample mean estimate of the equity premium, based on US data. Further, the studies that provide these estimates do not explicitly consider which models of the equity premium process can be rejected by actual data, though Pástor and Stambaugh's analysis strongly supports a model that incorporates breaks in the equity premium process.

The remainder of our paper proceeds as follows. The basic methodology of our simulation approach to estimating equity premia is presented in Section 1, along with important details on estimating the equity premium. (Appendices to the paper provide detailed explanations of the technical aspects of our simulations, including calibration of key model parameters.) In Section 2 we compare univariate financial statistics that arise in our simulations with US market data, including dividend yields, Sharpe ratios, and conditional moments including ARCH coefficients. Our results confirm that the simulations generate data broadly consistent with the US market data and, taken one-at-a-time, these financial statistics imply that the ex ante equity premium lies in a range much narrower than between 2% and 8%. We determine how much narrower in Section 3 by exploiting the full power of the simulation methodology. We compare joint multivariate distributions of our simulated data with observed US data, yielding a very precise estimate of the ex ante equity premium and providing strong rejections of models of the equity premium process that fail to incorporate time variation, breaks, and/or trends. We find the range of ex ante equity premium values is very narrow: 3.5% plus or minus 50 basis points. Our consideration of a broad collection of possible data generating processes and models lends confidence to the findings. Section 4 concludes.

I Methodology

Consider a stock for which the price P_t is set at the beginning of each period t and which pays a dividend D_{t+1} at the end of period t. The return to holding this stock (denoted R_t) is defined as

$$R_t = \frac{D_{t+1} + P_{t+1} - P_t}{P_t}.$$

The risk-free rate, set at the beginning of each period, is denoted $r_{t,f}$. The ex ante equity premium, π, is defined as the difference between the expected return on risky assets, $E\{R_t\}$, and the expected risk-free rate, $E\{r_{t,f}\}$:[4]

$$\pi \equiv E\{R_t\} - E\{r_{t,f}\}. \tag{1}$$

We do not observe this ex ante equity premium. Empirically, we only observe the returns that investors actually receive ex post, after they have purchased the stock and held it over some period of time during which random economic shocks impact prices. Hence, the ex post equity premium is typically estimated using historical equity returns and risk-free rates. Define $\overline{R}$ as the average historical annual return on the S&P 500 and $\overline{r}_f$ as the average historical return on US T-bills. Then we can calculate the estimated ex post equity premium, $\hat{\pi}$, as follows:

$$\hat{\pi} \equiv \overline{R} - \overline{r}_f. \tag{2}$$

Given that the world almost never unfolds exactly as one expects, there is no reason to believe that the stock return we estimate ex post is exactly the same as the return investors anticipated ex ante. It is therefore difficult to argue that just because we observe a 6% ex post equity premium in the US data, the premium that investors demand ex ante is also 6% and thus a puzzling challenge to economic theory. So we ask the following question: If investors' true ex ante premium is π, what is the probability that the US economy could randomly produce an ex post premium of at least 6%? The answer to this question has implications for whether or not the 6% ex post premium

[4]See, for instance, Mehra and Prescott (1985), Equation (14). We will consider time-varying equity premium models below.

observed in the US data is consistent with various ex ante premium values, π, with which standard economic theory may be more compatible. We also ask a deeper question: If investors' true ex ante premium is π, what is the probability that we would observe the various *combinations* of key financial statistics and yields that have been realized in the US, such as high Sharpe ratios and low dividend yields, high return volatility and a high ex post equity premium, and so on? The analysis of multivariate distributions of these statistics allows us to narrow substantially the range of equity premia consistent with the US market data, especially relative to previous studies that have considered univariate distributions.

Because the *empirical* joint distribution of the financial statistics we wish to consider is difficult or impossible to estimate accurately, in particular the joint distribution *conditional* on various ex ante equity premium values, we use simulation techniques to estimate this distribution. The simulated joint distribution allows us to conduct formal statistical tests that a given ex ante equity premium could have produced the US experience. Most of our models employ a time-varying ex ante equity premium, so that a simulation described as having an ex ante equity premium of 2.75% actually has a mean ex ante equity premium of 2.75%, while period-by-period the ex ante equity premium can vary somewhat from this mean value. In what follows we refer to the ex ante equity premium and the mean ex ante equity premium interchangeably.

A Matching Moments

Consider the valuation of a stock. Define $1+r_t$ as the gross rate investors use to discount payments received during period t. The price of the stock is then given by Equation (3),

$$P_t = E_t\left\{\frac{D_{t+1}+P_{t+1}}{1+r_t}\right\}, \tag{3}$$

where E_t is the conditional expectations operator incorporating information available to the market when P_t is formed, up to but not including the beginning of period t (*i.e.*, information from the end of period $t-1$ and earlier).

Assuming the usual transversality conditions, we can derive Equation (4) by recursively substituting out for future prices in Equation (3):

$$P_t = E_t \left\{ \sum_{j=0}^{\infty} \left(\Pi_{i=0}^{j} \frac{1}{1 + r_{t+i}} \right) D_{t+j+1} \right\}. \tag{4}$$

Defining the growth rate of dividends over the period t as $g_t \equiv (D_{t+1} - D_t)/D_t$, we can re-write Equation (4) as

$$P_t = D_t E_t \left\{ \sum_{j=0}^{\infty} \left(\Pi_{i=0}^{j} \left[\frac{1 + g_{t+i}}{1 + r_{t+i}} \right] \right) \right\}. \tag{5}$$

Hence we can re-write Equation (1) as

$$\pi \equiv E \left\{ \frac{D_{t+1} + D_{t+1} E_{t+1} \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+1+i}}{1+r_{t+1+i}} \right\} - D_t E_t \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+i}}{1+r_{t+i}} \right\}}{D_t E_t \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+i}}{1+r_{t+i}} \right\}} - r_{t,f} \right\} \tag{6}$$

or

$$\pi \equiv E \left\{ \frac{(1 + g_t) \left(1 + E_{t+1} \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+1+i}}{1+r_{t+1+i}} \right\} \right) - E_t \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+i}}{1+r_{t+i}} \right\}}{E_t \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+i}}{1+r_{t+i}} \right\}} - r_{t,f} \right\}. \tag{7}$$

In the case of a constant equity premium π and a possibly time-varying risk-free interest rate we can re-write Equation (7) as

$$\pi \equiv E \left\{ \frac{(1 + g_t) \left(1 + E_{t+1} \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+1+i}}{1+\pi+r_{t+1+i,f}} \right\} \right) - E_t \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+i}}{1+\pi+r_{t+i,f}} \right\}}{E_t \left\{ \sum_{j=0}^{\infty} \Pi_{i=0}^{j} \frac{1+g_{t+i}}{1+\pi+r_{t+i,f}} \right\}} - r_{t,f} \right\}. \tag{8}$$

Under interesting conditions, such as risk-free rates and dividend growth rates that conditionally time-vary and covary (we consider, for instance, ARMA models and correlated errors for dividend growth rates and interest rates), the individual conditional expectations in Equation (8) are analytically intractable. The difference between the sample mean return and the sample mean risk-free interest rate provides a consistent estimate of π, as shown by Mehra and Prescott (1985), but unfortunately the sample mean difference is very imprecisely estimated, even based on more than 100 years of data.

We note that another consistent estimator of π is one that directly exploits the method of Donaldson and Kamstra (1996), hereafter referred to as the DK method. The DK method uses (ARMA) models for dividend growth rates and interest rates to simulate the conditional expectations $E_t\left\{\sum_{j=0}^{\infty}\Pi_{i=0}^{j}\frac{1+g_{t+i}}{1+\pi+r_{t+i,f}}\right\}$ and $E_{t+1}\left\{\sum_{j=0}^{\infty}\Pi_{i=0}^{j}\frac{1+g_{t+1+i}}{1+\pi+r_{t+1+i,f}}\right\}$. The DK method allows us, for a given *ex ante* equity premium (or time-varying equity premium process), to simulate the conditional expectations in Equation (8) as well as related (unconditional) moments, including the expected dividend yield, return volatility, ex post equity premium, and Sharpe ratio. Our estimate of π is produced by finding the value of π that minimizes the distance between the collection of simulated moments (produced by the DK procedure) and the analogous sample moments (from the US experience over the last half century). The estimation of these expectations relies on the exact form of the conditional models for dividend growth rates and interest rates, that is, the parameters that characterize these models. A joint estimation of these models' parameters and π (*i.e.* minimizing the distance between simulated and sample moments by varying all the model's parameters and π at once) would be computationally very difficult. We utilize a two-step procedure in which first, for a given ex ante equity premium, we jointly estimate the parameters that characterize the evolution of dividend growth rates and interest rates. We use these models to simulate data to compare with realized S&P 500 data. Second, we do a grid search over values of the ex ante equity premium to find our SMM estimate of π.

It is helpful to consider some examples of estimators based on our simulation technique. The simplest estimator would have us considering only the ex ante equity premium moment, $\pi = E\left[R_t\right]-E\left[r_{f,t}\right]$, ignoring other potentially informative moments of the data, such as the dividend yield and return volatility. Exploiting the DK procedure, we would find that the π in Equation (8) which matches the ex post equity premium (the sample moment analogue of Equation (8)) is the sample estimate of the ex post equity premium, roughly 6%. That is, in this simplest case, when we minimize the distance between the sample moment and the simulated moment and find that the estimate of the ex ante equity premium is the ex post equity premium, we do so by construction. If the DK method is internally consistent, and if we are fitting only the ex post equity premium sample moment, then the difference must be zero at the value of π equal to the ex post equity premium. This DK estimator of π, considering only one moment of the data, would offer no advantage over the ex post equity premium, which is the traditional estimate of the ex ante equity premium. Adding a second moment to our estimation procedure, say the dividend yield, and minimizing the distance between the simulated and sample moments for the ex post equity premium and the dividend yield *jointly*, would likely lead to a somewhat different ex ante equity premium estimate. Furthermore, the estimate would be more precisely estimated (*i.e.,* with a smaller standard error) since two moments are exploited to estimate the ex ante equity premium, not just one moment, at least if the extra moment of the data provided some unique information about the value of the parameter π.

The DK method provides simulated dividend yields, ex post equity premia, and any other statistic that is derivative to returns and prices, such as return volatility, resulting in a broad collection of simulated moments with which to compare moments of the actual US data in order to derive an estimator. The large collection of available moments makes it likely that our analysis can provide a tighter bound on the value of the ex ante equity premium than has been achieved previously.

B The Simulation

To estimate the joint distribution of the financial quantities of interest, we consider models calibrated to the US economy. (We calibrate to US data over 1952 through 2004, with the starting year of 1952 motivated by the US Federal Reserve Board's adoption of a modern monetary policy regime in 1951.) We provide specific details on the nature of the models we consider and how we conduct our simulations in Appendices 1 and 2. Our entire procedure can be generally summarized in the following five steps:

Step 1: Specify assumptions about the ex ante equity premium demanded by investors. Is the premium constant or time-varying? If constant, what value does it take? If time-varying, how does the value change over time? Are there any structural breaks in the equity premium process over time? Pástor and Stambaugh (2001), among others, provide evidence that the equity premium has been trending downward over the sample period we study, finding a modest downward trend of roughly 0.80% in total since the early 1950s. Pástor and Stambaugh (2001) also find fairly strong support for there having been a structural break over the 1990s which led to a 0.5% drop in the equity premium.[5]

Once the process driving the ex ante equity premium is defined, we can specify the discount rate (which equals the risk-free rate plus the equity premium) that an investor would rationally apply to a forecasted dividend stream in order to calculate the present value of a dividend-paying stock. Note that if the equity premium varies over time, then the models generated in the next step are calibrated to mimic the degree of covariation between interest rates, dividend growth rates, and equity premia observed in the US data.

Step 2: Estimate econometric models for the time-series processes driving actual dividends and interest rates in the US economy, allowing for autocorrelation and covariation as observed in the US data. These models will later be used to Monte-Carlo simulate a variety of potential paths for US dividends and interest rates. The simulated dividend and interest rate paths are of course different in each of these simulated economies because different sequences of random innovations are applied to the common stochastic processes in each case. However, the key drivers of the simulated economies themselves are all still identical to those of the US economy since all economies share common stochastic processes fitted to US data.

Some of the models we consider assume that all cashflows received by investors come in the form of dividends (the standard assumption). Another set of models we consider embed higher cashflows and cashflow growth rates than observed in the US S&P 500 dividend data, to account for the observation of Bagwell and Shoven (1989), Fama and French (2002), and others, that dividends under-report total cashflows to shareholders. As reported by these authors, firms have been increasingly distributing cash to shareholders via share repurchases instead of via dividends, a phenomenon commonly known as disappearing dividends, a practice adopted widely beginning in the late 1970s. Fama and French find evidence that the disappearance of dividends is in part due to an increase in the inflow of new listing to US stock exchanges, representing mostly young companies

[5] A falling equity premium is thought to come from several sources, including the declining cost of diversifying through mutual funds over the last half century, the infeasibility before the advent of mutual funds to hold fully diversified portfolios (hence higher returns required by investors to hold relatively undiversified positions), and the broader pool of investors now participating in equity ownership, sharing in the market risk and presumably lowering the required rate of return to risky assets. See Siegel (1999) and Diamond (2000).

with the characteristics of firms that would not be expected to pay dividends, and in part due to a decline in the propensity of firms to pay dividends.

Thus, for some models in our simulations, we adopt higher cashflows than would be indicated by considering US dividend data alone. On a broad set of data, Grullon and Michaely (2002) find that total payouts to shareholders have remained fairly flat, not growing over the period we consider. To the extent that this is true of the S&P 500 data, the models we consider with upward-trending dividend growth are overly aggressive, but as we show below, the higher dividend growth rate only widens the range of plausible ex ante equity premia, meaning our estimate of the precision of our approach is conservative.

Step 3: Allow for the possibility of estimation error in the parameter values for the dividend growth rate, interest rate, and equity premium time-series models. That is, incorporate into the simulations uncertainty about the true parameter values. This allows for some models with more autocorrelation in the dividend growth, interest rate, and equity premium series, some with less, some with more correlation between the processes, some with less, some with a higher variance or mean of dividend growth and interest rates, some with less, and so on. This uncertainty is measured using the estimated covariance of the parameter estimates from our models generated in Steps 1 and 2, and the procedure to randomly select parameters from the estimated joint distribution of the parameters is detailed in Appendix 1. We also account for investor uncertainty about the true fundamental processes underlying prices and returns by performing tests insensitive to this uncertainty and its impact on prices and returns, as we describe below.

Further details about Steps 1 through 3 are contained in Appendix 1. Before continuing with summarizing Steps 4 and 5 of our methodology, it is worth identifying some models that emerge from various combinations of the assumptions embedded in Steps 1 through 3. The key models we consider in this paper are shown in Table I. The first column of Table I indicates numbering that we assign to the models. The second column specifies the time-series process used to generate the interest rate and dividend growth rate series, corresponding to Step 2. The next three columns relate to Step 1 above, indicating whether or not the ex ante equity premium process incorporates a downward trend over time (and if so, how much the mean ex ante equity premium in 1952 differs from the value in 2004), whether or not there is a structural break (consisting of a 50 basis point drop) in the equity premium consistent with the findings of Pástor and Stambaugh (2001), and whether or not there is a break in the dividend growth rate process, consistent with the Bagwell and Shoven (1989) and Fama and French (2002) finding of an increase in share repurchases from the late 1970s onward.[6] The last column corresponds to Step 3, showing which models incorporate uncertainty in generating parameters. We consider a selection of 12 representative models, ranging from a simple model with no breaks or trends in the equity premium process (Model 1) to very complex models.[7] Each model is fully explored in the sections that follow. We now continue describing the two final steps of our basic methodology.

Table I goes about here.

Step 4: Calculate the fundamental stock returns (and hence ex post equity premia) that arise in each simulated economy, using a discounted-dividend-growth-rate model and based on assumptions about the ex ante equity premium from Step 1, the dividend growth rate and interest rate processes specified in Step 2, and the possible parameter uncertainty specified in Step 3. The model is rolled out to produce 53 annual observations of returns, prices, dividends, interest rates, and so on, mimicking the 53 years of annual US data available to us for comparison. Keep in mind the fact that the assumptions made in Steps 1 through 3 are the same for all simulated economies in a given experiment. That is, all economies in a given experiment have the same ex ante equity premium model (for instance a constant ex ante equity premium, or perhaps an ex ante equity premium that time-varies between a starting and ending value) and yet all economies in the set of simulations have different ex post equity premia. Given the returns and ex post equity premia for each economy, as well as the means of the interest rates and dividend growth rates produced for each economy, we are able to calculate various other important characteristics, including return volatility,

[6]In each case where we consider model specifications intended to capture real-world features like breaks and trends in rates and premia, we adopt parameterizations that bias our results to be more conservative (*i.e.* to produce a wider confidence interval for the ex ante equity premium). This allows us to avoid over-stating the gains in precision possible with our technique. For example, while Pástor and Stambaugh (2001) find evidence that there was a break in the equity premium process across several years in the 1990s, we concentrate the entire break into one year (1990). Allowing the break to be spread across several years would lead to a narrower bound on the ex ante equity premium than we find. See Appendix 1 for more details.

[7]For the sake of brevity, the Gordon (1962) constant dividend growth model is excluded from the set of models we explore in this paper. We did analyze the Gordon model and found it to perform very poorly. The model itself is rejected at every value of the ex ante equity premium, even more strongly than any other simple model considered in this paper is rejected.

dividend yields, and Sharpe ratios. There is nothing in our experimental design to exclude (rational) market crashes and dramatic price reversals. Indeed our simulations do produce such movements on occasion. The details of Step 4 are provided in Appendix 2.

Step 5: Examine the distributions of variables of interest, including ex post equity premia, Sharpe ratios, dividend yields, and regression coefficients (from estimating AR(1) and ARCH models for returns) that arise conditional on various mean values and various time-series characteristics of the ex ante equity premia. Comparing the performance of the US economy with various univariate and multivariate distributions of these quantities and conducting joint hypothesis tests allows us to determine a narrow range of equity premia consistent with the US market data. That is, only a small range of mean ex ante equity premia and time-varying equity premium models could have yielded the outcome of the past half century of high mean return and return standard deviation, low dividend yield, high ex post equity premium, *etc.*

A large literature makes use of similar techniques in many asset pricing applications, directly or indirectly simulating stock prices and dividends under various assumptions to investigate price and dividend behavior.[8] However, these studies typically employ restrictions on the dividend and discount rate processes in order to obtain prices from some variant of the Gordon (1962) model and/or some log-linear approximating framework. For instance, the present value (price, defined as P_0) of an infinite stream of expected discounted future dividends can be simplified under the Gordon model as

$$P_0 = D_1/(r - g), \tag{9}$$

where D_1 is the coming dividend, r is the constant discount rate, and g is the constant dividend growth rate. That is, by assuming constant r and g, one can analytically solve for the price. If, however, discount rates or dividend growth rates are in fact conditionally time-varying, then the infinite stream of expected discounted future dividends in Equation (5) cannot be simplified into Equation (9), and it is difficult or impossible to solve prices analytically without imposing other simplifying assumptions.

[8] See, for example, Scott (1985), Kleidon (1986), West (1988a,b), Campbell (1991), Gregory and Smith (1991), Mankiw, Romer, and Shapiro (1991), Hodrick (1992), Timmermann (1993, 1995), and Campbell and Shiller (1998).

Rather than employ approximations to solve our price calculations analytically, we instead simulate the dividend growth and discount rate processes directly, and evaluate the expectation through Monte Carlo integration techniques, adopting the DK method.[9] In the setting of time-varying dividend growth rates and interest rates which conditionally covary, this technique allows us to evaluate prices, returns, and other financial quantities without approximation error.[10] We also take extra care to calibrate our models to the time-series properties of actual market data. For example, annual dividend growth is strongly autocorrelated in the S&P 500 stock market data, counter to the assumption of a logarithmic random walk for dividends sometimes employed for tractability in other applications. Furthermore, interest rates are autocorrelated and cross-correlated with dividend growth rates. Thus we incorporate these properties in our 12 models (shown in Table I), which we use to produce our simulated dividend growth rates, interest rates, and, ultimately, our estimate of the ex ante equity premium.

We estimated each of the 12 models over a grid of discrete values of the ex ante equity premium, with the grid as fine as an eighth of a percent in the vicinity of a 3.5% equity premium, and no coarser than 100 basis points for equity premium values exceeding 5%. The entire exercise was conducted using distributed computing across a grid of 30 high-end, modern-generation computers over the course of a month. On a modern stand-alone computer, estimation of a single model for a single assumed value of the ex ante equity premium would take roughly one week to estimate (and, as stated above, we consider many values of the ex ante equity premium for each of our models).

II Univariate Conditional Distributions For Model 1

All of the results in this section of the paper are based on Model 1, as defined in Table I. Model 1 incorporates interest rates that follow an AR(1) process and dividend growth rates that follow a MA(1) process. The ex ante equity premium in Model 1 follows an AR(1) process (that emerges from Merton's (1980) conditional CAPM, as detailed in Appendix 1), with no trends or breaks in either the equity premium process or dividend growth rate process. We start with this "plain

[9]The Dondaldson and Kamstra (1996) method nests other fundamental dividend-discounting valuation methods as special cases. For instance, in a Gordon (1962) world of constant dividend growth rates and interest rates, the DK method produces the Gordon model price, albeit through numerical integration rather than analytically.

[10]There is still Monte Carlo simulation error, but that is random, unlike most types of approximation error, and it can also be measured explicitly and controlled to be very small, which we do, as explained in Appendix 2.

vanilla" model because it provides a good illustration of how well dividend-discounting models that incorporate time-varying autocorrelated dividend growth and discount rate processes can produce prices and returns that fit the experience of the last half century in the US. This model also provides a good starting point to contrast with models employing breaks and trends in equity premium and dividend growth processes. We consider more complex and arguably more realistic models incorporating trends and breaks later in the paper.

It is well known that the ex ante equity premium is estimated with error. See, for instance, Merton (1980), Gregory and Smith (1991), and Fama and French (1997). Any particular realization of the equity premium is drawn from a distribution, implying that given key information about the distribution (such as its mean and standard deviation), one can construct a confidence interval of statistically similar values and determine whether a particular estimate is outside the confidence interval. As mentioned above, an implication of this estimation error is that most studies have produced imprecise estimates of the mean equity premium. For instance, a typical study might yield an 800 basis point 95% confidence interval around the ex ante equity premium.[11] Studies including Fama and French (2002) have introduced innovations that make it possible to narrow the range. One of our goals is to further sharpen the estimate of the mean ex ante equity premium.

We first consider what we can learn by looking at the univariate statistics that emerge from our simulations. We can use the univariate distributions to place loose bounds on plausible values of the mean ex ante equity premium. While the analysis in this section based on univariate empirical distributions is somewhat casual, in Section III we conduct formal analysis based on χ^2 statistics and the joint distributions of the data, yielding very tight bounds on plausible values of the mean ex ante equity premium and identifying plausible models of the equity premium process, representing our main contributions.

Consider the following: conditional on a particular value of the ex ante equity premium, how unusual is an observed realization of the ex post equity premium? How unusual is an observed realization of the mean dividend yield? Each simulated economy produces a set of financial statistics based on the simulated *annual* time-series observations, and these financial statistics can be

[11]This particular range is based on the simple difference between mean realized equity returns and the average riskfree rate based on the last 130 years of data, as summarized in Table I of Fama and French (2002).

compared and contrasted with the US experience of the last half century. By considering not only the mean of a financial statistic across simulated economies, such as the mean ex post equity premium, but also conditional moments and higher moments including the standard deviation of excess returns produced in our simulations, we can determine with high refinement the ability of our simulated data to match characteristics of the US economy. For instance, market returns, to be discussed below, are volatile. Thus it is interesting to examine the degree to which our simulations are able to produce volatile returns and to look at the distribution of return variance as we vary the mean ex ante equity premium in our simulated economies.

We can compare any financial statistic from the last half century to our simulated economies provided the statistic is based on returns or dividends or prices, as these are data that the simulation produces. We could also consider moments based on interest rates or dividend growth rates, but since we calibrate our models to interest rates and dividend growth rates, all our simulations should (and do) fit these moments well by construction. We choose moments based on two considerations. First, the moments should be familiar and the significance of the moments to economic theory should be obvious. Second, the moments should be precisely estimated; if the moments are too "noisy," they will not help us narrow the range of ex ante equity premia. For instance, return skew and kurtosis are very imprecisely estimated with even 50 years of data, so that these moments are largely uninformative. The moments must also be well-defined; moments must be finite, for instance. The expected value of the price of equity is undefined, but we can use prices in concert with a cointegrated variable like lagged price (to form returns) or dividends (to form dividend yields).

Rather than presenting copious volumes of tabled results, we summarize the simulation results with concise plots of probability distributions of the simulated data for various interesting financial statistics. This permits us to determine if a particular ex ante equity premium produces financial statistics similar to what has been seen over the last half century in the US.

Figure 1 contains four panels, and in each panel we present the probability distribution function for one of various financial statistics (ex post equity premia, dividend yield, Sharpe ratio, and return volatility) based on each of four different ex ante equity premium settings. We also indicate the realized value for the actual US data. Comparison of the simulated distribution with realized values in these plots permits a very quick, if casual, first assessment of how well the realized US data agree with the simulated data, and which assumed values of the ex ante equity premium appear inconsistent with the experience of the last half century of US data.

Panels A through D of Figure 1 contain probability distribution functions (PDFs) corresponding to the mean ex post equity premium, the mean dividend yield, the Sharpe ratio, and return volatility respectively, based on assumed mean ex ante equity premia of 2.75%, 3.75%, 5%, and 8%. For the sake of clarity, the dotted lines depicting the PDFs in Figure 1 are thinnest for the 2.75% case and become progressively thicker for the 3.75%, 5%, and 8% cases. The actual US realized data is denoted in each panel with a solid vertical line.

The actual US mean equity premium, displayed in Panel A, is furthest in the right tail of the distribution corresponding to a 2.75% ex ante equity premium, and furthest in the left tail for the ex ante premium of 8%. The wide range of the distribution of the mean ex post equity premia for each assumed value of the ex ante equity premium is consistent with the experience of the last half century in the US, in which the mean ex post equity premium has a 95% confidence interval spanning plus or minus roughly 4% or 5%. The actual dividend yield of 3.4%, displayed in Panel B, is unusually low for the 5% and 8% ex ante equity premium cases, but it is near the center of the distribution for the ex ante premium values of 2.75% and 3.75%. In Panel C, only the Sharpe ratios generated with an ex ante equity premium of 8% appear inconsistent with the US experience of the last half century. The return volatility, displayed in Panel D, clearly indicates that the experience of the US over the last half century is somewhat unusual for all ex ante equity premia considered, though least unusual for the lowest ex ante equity premium. Casual observation, based on only the evidence in these univariate plots, implies that the ex ante equity premium which could have generated the actual high ex post equity premium and low dividend yield of the last half century of the US experience likely lies above 2.75% and below 5%.

Figure 1 goes about here.

We constructed similar plots for the mean return and for conditional moments, including the return first order autocorrelation coefficient estimate (the OLS parameter estimate from regressing returns on lagged returns and a constant, *i.e.,* the AR(1) coefficient), the return first order autoregressive conditional heteroskedasticity coefficient estimate (the OLS parameter estimate from regressing squared residuals on lagged squared residuals and a constant, *i.e.,* the ARCH(1) coefficient), and the price-dividend ratio's first order autocorrelation coefficient estimate (the OLS parameter estimate from regressing the price-dividend ratio on the lagged price-dividend ratio and a constant). The mean return distributions are similar to the ex post equity premium distributions shown in Figure 1, and all choices of the ex ante equity premium produce returns and price-dividend ratios that have conditional time-series properties matching the US data, so these results are not presented here.

Figure 1 has two central implications of interest to us. First, the financial variable statistics produced in our simulations are broadly consistent with what has been observed in the US economy over the past five decades. Most simulated statistics match the magnitudes of financial quantities from the actual US data, even though we do not calibrate to prices or returns.[12] Second, the results suggest that the 2.75% through 8% interval we present here likely contains the ex ante equity premium consistent with the US economy. Univariate results for Models 2 through 10 are qualitatively very similar to those presented for Model 1. Univariate results for Models 11 and 12, in contrast, are grossly rejected by the experience of the US economy. Detailed univariate results for Models 2 through 12 are omitted for the sake of brevity, but the poor performance of Models 11 and 12 will be evident in multivariate results reported below.

To narrow further the range of plausible ex ante equity premium values, we need to exploit the full power of our simulation procedure by considering the *joint distributions* of statistics that arise in our simulations and comparing them to empirical moments of the observed data. We consider the multivariate distributions of several moments of the data, including ex post equity premia, dividend yields, and return volatility. This exercise allows for inference that is not feasible with the univariate analysis conducted above, and it leads to a very precise estimate of the ex ante equity premium. We turn to this task in the next section, where we also broaden the class of models we consider.

[12]This in itself is noteworthy, as analytically tractable models, such as the Gordon (1962) growth model, typically imply constant or near-constant dividend yields and very little return volatility. In contrast, dividend yields observed in practice vary considerably over time and are strongly autocorrelated, and returns exhibit considerable volatility.

III Model Extensions, Multivariate Analysis, and Tests

The central focus in this section is on *joint* distributions of the financial statistics that emerge from our simulations: combinations of the returns, ex post equity premia, Sharpe ratios, dividend yields, *etc.*, and tests on the value of the ex ante equity premium using these joint distributions. We focus primarily on three moments of the data: the mean ex post equity premium, the excess return volatility, and the mean dividend yield. These three moments have the advantage of being the most precisely estimated and hence most informative for the value of the ex ante equity premium. Other moments that we could have considered are either largely redundant (such as the Sharpe Ratio which is a direct function of excess returns and the excess return standard deviation), or are so imprecisely estimated (for example, the ARCH(1) or AR(1) coefficients) that they would not help sharpen our estimates of the ex ante equity premium. Of course, we also do not consider the distributions of financial variables to which we calibrate our simulations (interest rates and dividend growth rates), as the simulated mean, variance, and covariance of these variables are, by construction, identical to the corresponding moments of the actual data to which we calibrate.

Our purpose in considering joint distributions is two-fold. First, multivariate tests are used to form a tight confidence bound on the true value of the ex ante equity premium. These tests strongly reject our models if the ex ante equity premium is outside of a narrow range around 3.5%. This range is not sensitive to even fairly substantial changes in the model specification, which suggests that the 3.5% finding is robust. Second, this analysis leads us to reject model specifications that fail to incorporate certain features, such as trends and breaks in the equity premium. Interestingly, even when a model specification is rejected, we find the most plausible ex ante equity premium still lies in the same range as the rest of our models, very near 3.5%.

Up to this point we have considered detailed results for Model 1 exclusively. The Model 1 simulation incorporates some appealing basic features, such as parameter uncertainty and calibrated time-series models for equity premia, interest rates, and dividend growth rates. It does not, however, incorporate some features of the equity premium process that have been indicated by other researchers. One omitted feature is a gradual downward trend in the equity premium, as documented in many studies, including Jagannathan, McGrattan, and Scherbina (2000), Pástor and

Stambaugh (2001), Bansal and Lundblad (2002), and Fama and French (2002). Another is a structural break in the equity premium process over the early 1990s, as shown by Pástor and Stambaugh (2001). An increase in the growth rate of cashflows (but not dividends) to investors starting in the late 1970s, as documented by Bagwell and Shoven (1989), Fama and French (2001) and others, is also a feature that Model 1 fails to incorporate. Therefore, in this section we consider models which incorporate one, two, or all three of these features, as well as different time-series models for interest rates and equity premia. We also consider stripped-down models to assess the marginal contribution of model features such as parameter uncertainty and the specification of the time-series process used to model dividend growth rates and interest rates.

In Figures 2 through 8 (to be fully discussed below), we present χ^2 test statistics for the null hypothesis that the US experience during 1952 through 2004 could have been a random draw from the simulated distribution of the mean ex post equity premium, the excess return volatility, and the mean dividend yield.[13]

A significant test statistic, in this context, suggests that the combination of financial statistics observed for the US economy is significantly unusual compared to the collection of simulated data, leading us to reject the null hypothesis that the given model and assumed ex ante equity premium value could have generated the US data of the last half century. It is possible to reject every ex ante equity premium value if we use models of the equity premium that are misspecified (the rejection of the null hypothesis can be interpreted as a rejection of the model). It is also possible that a very wide range of ex ante equity premium values are not rejected for a collection of models, thwarting our efforts to provide a precise estimate of the ex ante equity premium or a small range of allowable equity premium models.

As it happens, models that ignore breaks and trends in the equity premium are rejected for

[13]The χ^2 tests are based on joint normality of sample estimates of moments of the simulated data, which follow an asymptotic normal distribution based on a law of large numbers (see White, 1984, for details). In the case of the excess return volatility, we consider the *cube root* of the return variance, which is approximately normally distributed (see page 399 of Kendall and Stuart, 1977, for further details). We also estimate the probability of rejection using bootstrapped p-values, to guard against deviations from normality. These bootstrapped values are qualitatively identical to the asymptotic distribution p-values. Finally, when performing tests that include the dividend yield moment, if the simulation includes a break in dividends corresponding to an increase in cash payouts starting in 1978 in the US data (again, see Fama and French, 2001), we also adjust the US data to reflect the increase in mean payout levels. This makes for a small difference in the mean US payout ratio and no qualitative change to our results if ignored.

virtually every value of the ex ante equity premium we consider. But for a group of sophisticated models that incorporate trends and breaks in the equity premium, we cannot reject a narrow range of ex ante equity premia, roughly between 3% and 4%. We also find that models tend to be rejected if the impact on cashflows to shareholders from share repurchases are ignored. We begin with some simple models, then consider models that are arguably more realistic as they incorporate equity premium and cashflow trends and breaks, and finish by considering a host of related issues, including the impact of parameter estimation error and, separately, *investor* uncertainty about the fundamental value of equities.

A Simple (One-at-a-Time) Model Extensions

We now consider extensions to Model 1, each extension adding a single feature to the base model. Recall that the features of each model are summarized in Table I. For Model 2, an 80 basis point downward trend is incorporated in the equity premium process. For Model 3, a 50 basis point drop in year 39 of the simulation (corresponding to 1990 for the S&P 500 data) is incorporated in the equity premium process. For Model 4, the dividend growth rate process is shifted gradually upward a total of 100 basis points, starting in year 27 of the simulation (corresponding to 1978 for the S&P 500 data) and continuing for 20 years at a rate of 5 basis points per year. These one-at-a-time feature additions help us evaluate if one or another feature documented in the literature can markedly improve model performance over the simple base model.

Panel A of Figure 2 and Panel A of Figure 3 display plots of the value of joint χ^2 tests on three moments of the data, the mean ex post equity premium, the excess return volatility, and the mean dividend yield, for Models 1 though 4, and shows how the test statistic varies as the ex ante equity premium varies from 2.25% to 8% in increments as small as an eighth of a percent toward the lower end of that range. Panels B through D of Figures 2 and 3 display the univariate Student t-test statistics for each of these three moments of the data, again showing how the test statistic varies with the assumed value of the ex ante equity premium. The values of the ex ante equity premia indicated on the horizontal axis represent the *ending* values of the ex ante equity premium in each set of simulations. For models which incorporate a downward trend or a structural break in the equity premium, the ending value of the ex ante equity premium differs from the starting value.

So, for instance, Model 2 has a starting ex ante equity premium that is 80 basis points higher than that displayed in Figure 2, as Model 2 has an 80 basis point trend downward in the ex ante equity premium. For Model 1 the value of the ex ante equity premium is the same at the *end* of the 53-year simulation period as it is at the *start* of the 53-year period, as Model 1 does not incorporate a downward trend or structural break in the equity premium process. Critical values of the test statistics corresponding to statistical significance at the 10%, 5%, and 1% levels are indicated by thin dotted horizontal lines in each panel, with the lowest line indicating significance at the 10% level and the highest line the 1% significance level.

Figures 2 and 3 go about here.

Consider now specifically Panel A of Figures 2 and 3. (Note that we use a log scale for the vertical axis of the plots in Panel A of Figures 2 through 8 for clarity of presentation. Note as well that we postpone further discussion of Panels B through D until after we have introduced results for all the models, 1 through 12.) On the basis of Panel A of Figures 2 and 3, we see that only in the case of Model 4 do we observe χ^2 test statistics lower than the cutoff value implied by a 10% significance level (again, indicated by the lowest horizontal dotted line in the plot). The test statistics dip (barely) below the 10% cutoff line only for values of the ex ante equity premium within about 25 basis points of 4%. Models 1-3, in contrast, are rejected at the 10% level for *every* ex ante equity premium value. If we allow fairly substantial departures of the S&P 500 data from the expected distribution, say test statistics that are unusual at the 1% level of significance (the upper horizontal dotted line in the plot), then all the models indicate ranges of equity premia that are not rejected, in each case centered roughly between 3.5% and 4%. Recall that the equity premium plotted is the *ending* value, so if the model has a downward trend or decline because of a break in the equity premium, its ending value is below its average ex ante equity premium.

One conclusion to draw from the relative performance of these four competing models is that each additional feature over the base model, the dividend growth acceleration in the late 1970s and the trends and breaks in the equity premium, lead to better performance relative to the base model, but each in isolation is still inadequate. The model most easily rejected is clearly that which does not account for trends and breaks in the equity premium and cashflow processes.

B Further Model Extensions (Two or More at a Time)

We turn now to joint tests based on Models 5 though 10. These models incorporate the basic features of Model 1, including time-varying and dependent dividend growth and interest rates, parameter uncertainty, and, with the exception of Model 10, an equity premium process derived from the Merton (1980) conditional CAPM (detailed in Appendix 1). These models also permit trends and/or breaks in the equity premium and dividend growth rate processes two or more at-a-time and incorporate alternative time-series models for the interest rate and the equity premium processes. Models 1 through 4 demonstrate that it is not sufficient to model the equity premium as an autoregressive time-varying process, and that one-at-a-time augmentation with trends or breaks in the equity premium process is also not sufficient, though the augmentations do lead to improvements over the base model in our ability to match sample moments from the US experience of the last half century. Models 5 through 10 allow us to explore questions like: do we need a conditionally time-varying equity premium model built on the Merton conditional CAPM model, or is it sufficient to have an equity premium that simply trends downward with a break? If we have a break, a trend, and time-variation in the equity premium process, is it still essential to account for the disappearing dividends of the last 25 years? Are our results sensitive to the time-series model specifications we employ in our base model?

Model 5 is the base model, Model 1, augmented to include an 80 basis point gradual downward trend in the equity premium and a 100 basis point gradual upward trend in the dividend growth rate. Model 6 is the base model adjusted to incorporate a 30 basis point gradual downward trend in the equity premium, a 50 basis point abrupt decline in the equity premium, and a 100 basis point gradual upward trend in the dividend growth rate. Model 7 is the best model as indicated by the Bayesian Information Criterion (BIC),[14] augmenting the equity premium process with a 30 basis point gradual downward trend and a 50 basis point abrupt decline and adding a 100 basis point gradual upward trend in the dividend growth rate. Model 8 takes the second-best BIC model

[14]For Models 7 and 8 we employ the BIC to select the order of the ARMA model driving each of the interest rate, equity premium, and dividend growth rate processes. The order of each AR process and each MA process for each series is chosen over a (0, 1, 2) grid. The BIC has been shown by Hannan (1980) to provide consistent estimation of the order of linear ARMA models. We employ the BIC instead of alternative criteria because it delivers relatively parsimonious specifications and because it is widely used in the literature (*e.g.*, Nelson, 1991, uses the BIC to select EGARCH models).

and incorporates a 30 basis point gradual downward trend in the equity premium, a 50 basis point abrupt decline in the equity premium, and a 100 basis point gradual upward trend in the dividend growth rate. Model 9 is the base model adjusted to incorporate a 30 basis point gradual downward trend in the equity premium and a 50 basis point abrupt decline in the equity premium. Model 10 has the equity premium model following a deterministic downward trend with a 50 basis point structural break, interest rates following an AR(1), and dividend growth rates following an MA(1).

Given the existing evidence in support of a gradual downward trend in the equity premium, a structural break in the equity premium process over the early 1990s, and an increase in the growth rate of non-dividend cashflows to investors (such as share repurchases) starting in the late 1970s, we believe Models 6, 7, and 8 to be the best calibrated and therefore perhaps the most plausible among all the models we consider, and Model 5 to be a close alternative.

In Panel A of Figures 4, 5, and 6 we present plots of the χ^2 test statistics on three moments of the data, the mean ex post equity premium, the excess return volatility, and the mean dividend yield. Again, we consider Panels B through D later. We see in Panel A of Figures 4 and 5 that for Models 5 through 8 we cannot reject a range of ex ante equity premium values at the 5% level. These models produce test statistics that drop well below even the 10% critical value (recall that Panel A's scale is logarithmic, and thus compressed). These models all embed the increased cashflow feature and either an eighty basis point downward trend in the equity premium, or both a break and a trend in the equity premium, adding to an eighty basis point decline over the last half century. The range of ex ante equity premia supported (not rejected) is narrowest for Model 7 (the best model indicated by BIC) and Model 8 (the second best model indicated by BIC) with a range less than 75 basis points at the 10% level. The range is slightly wider for Models 5 and 6, roughly 75 to 100 basis points. In each case, the ex ante equity premium that yields the minimum joint test statistic, corresponding to our estimate of π, is centered between 3.25% and 3.75%.

For the models which exclude the cashflow increase, Models 9 and 10, displayed in Figure 6, we see that we can reject at the 10% level all ex ante equity premium values. Model 9 is best compared to Model 6, as it is equivalent to Model 6 with the sole difference of excluding the cashflow increase. We see from Panel A of Figures 4 and 6 that excluding the cashflow increase flattens the trough of the plot of χ^2 statistics, and approximately doubles the test statistic value, from a little over 3 for

Model 6 in Figure 4 to a little over 6 for Model 9 in Figure 6 (recall that the scale is compressed in Panel A as we use a log scale). Model 10 is identical to Model 9 apart from the sole difference that Model 10 excludes the Merton CAPM conditionally-varying equity premium process. Exclusion of this conditional time variation (modeled as a first order autoregressive process) worsens the ability of the model to match moments to the US experience at every value of the ex ante equity premium. The difference in performance leads us to reject a model excluding a conditionally-varying equity premium.

Figures 4, 5, and 6 go about here.

On the basis of our most plausible models, Models 6, 7, and 8, we can conservatively conclude that the ex ante equity premium is within 50 basis points of 3.5%. We can also conclude that models that allow for breaks and/or trends in the equity premium process are the only models that are not rejected by the data. Simple equity premium processes, those that rule out any one of a downward break and/or trend or a Merton (1980) CAPM conditionally-varying equity premium process, cannot easily account for the observed low dividend yields, high returns, and high return volatility. Ignoring the impact of share repurchases on cashflows to investors over the last 25 years also compromises our ability to match the experience of US prices and returns of the last half century.

C Is Sampling Variability (Uncertainty) in Generating Parameters Important?

All of the models we have considered so far, Models 1-10, incorporate parameter value uncertainty. This uncertainty is measured using the estimated covariance of the parameter estimates from our models. We generate model parameters by randomly drawing values from the joint distribution of the parameters, exploiting the asymptotic result that our full information maximum likelihood procedure produces parameter estimates that are jointly normally distributed, with an easily computed variance-covariance structure.

Now we consider two models that have no parameter sampling variability built into them, Models 11 and 12. In these models the point estimates from our ARMA estimation on the S&P 500 data are used for each and every simulation. Ignoring uncertainty about the true values for the parameters of the ARMA processes for interest rates, dividend growth rates, and the equity premium should dampen the variability of the generated financial statistics from these simulations, and potentially understate the range of ex ante equity premia supported by the last half century of US data. Model 11 is the base model augmented to incorporate a 30 basis point gradual downward trend in the equity premium, a 50 basis point abrupt decline in the equity premium, and a 100 basis point gradual upward trend in the dividend growth rate, with no parameter uncertainty. (Model 11 is identical to Model 6 apart from ignoring parameter uncertainty.) Model 12 is the base model, Model 1, with no parameter uncertainty.

Figure 7 goes about here.

In Panel A of Figure 7 we present plots of the χ^2 test statistics on three moments of the data, the mean ex post equity premium, the excess return volatility, and the mean dividend yield. Again, we consider Panels B through D later. We see in Panel A that both Models 11 and 12 are rejected for all values of the ex ante equity premium, though Model 11, which allows for trends and breaks, performs better than Model 12. The log scale for the vertical axis compresses the values, but the minimum χ^2 statistic for Model 12 is close to 30, indicating very strong rejection of the model, while the minimum χ^2 statistic for Model 11 is roughly 10. In each case, the ex ante equity premium that yields the minimum joint test statistic, corresponding to our estimate of π, is centered around 3%. It is apparent that parameter uncertainty is an important model feature. Ignoring parameter uncertainty leads to model rejection, even at the ex ante equity premium setting that corresponds to the minimum test statistic.

D The Moments That Matter

An interesting question that arises with regard to the joint tests is, where does the test power come from? That is, which variables give us the power to reject certain ranges of the ex ante equity premium in our joint χ^2 tests? An examination of the ranges of the ex ante equity premium consistent with the *individual* moments can shed some light on the source of the power of the joint tests. Panels B, C, and D of Figures 2 through 7 display plots of the univariate t-test statistics based on each of the variables we consider in the joint tests plotted in Panel A of these figures. Panel B of each figure plots t-test statistics on the ex post equity premium, Panel C of each figure plots t-test statistics on the excess return volatility, and Panel D of each figure plots t-test statistics on the price-dividend ratio.

Consider first Panel B of Figures 2 through 7. Virtually all of the models have a minimum t-test statistic at a point that is associated with an ex ante equity premium close to 6%.[15] Because our method involves minimizing the distance between the ex post equity premium based on the actual S&P 500 value (which is a little over 6%) and the ex post equity premium estimate based on the simulated data, it is not surprising that the minimum distance is achieved for models when they are set to have an ex ante equity premium close to 6%. The t-test on the mean ex post equity premium rises linearly as the ex ante equity premium setting departs from 6% for each model, but does not typically reject ex ante equity premium values at the 10% level until they deviate quite far from the ex ante value at which the minimum t-test is observed. For example, in Panel B of Figure 4 the ending ex ante equity premium must be as low as 2.25% or as high as 7% before we see a rejection at the 10% level. This wide range reflects the imprecision of the estimate of the ex post equity premium which is also evident in the actual S&P 500 data.

The t-tests on the excess return volatility, presented in Panel C of Figures 2 through 7, indicate that lower ex ante equity premium values lead to models that are better able to match the S&P 500 experience of volatile returns.[16] Note that as the ex ante equity premium decreases, the volatility of returns *increases*, so high ex ante equity premia lead to simulated return volatilities that are much lower than the actual S&P 500 return volatility we have witnessed over the last half century. The test statistic, however, rises slowly as the ex ante equity premium grows larger, in contrast to the joint test statistics plotted in Panel A of Figures 2 through 7, in which the χ^2 test statistic

[15] Recall that the ex ante equity premium values shown on the horizontal axes are *ending* values, so if the model has a downward trend or break in the equity premium process, its ending value is below the mean equity premium. For instance, Model 11 has a data generating process that incorporates trends and breaks that lead to an ending equity premium lower than the starting value. Accordingly, for this model we observe (in Panel B of Figure 7) a minimum t-test at an *ending* value of the ex ante equity premium which is below the 6% *average* equity premium. The coarseness of the grid of ex ante equity premium values around 6% prevents this feature from being more obvious for some of the other models.

[16] The intuition behind this result is easiest to see by making reference to the Gordon (1962) constant dividend growth model, shown above in Equation 9. As the discount rate, r, declines in magnitude, the Gordon price increases. The variable r equals the risk-free rate plus the equity premium in our simulations, so low values of the equity premium lead to values of the discount rate that are closer to the dividend growth rate, resulting in higher prices. When the value of the equity premium is low, small increases in the dividend growth rate or small decreases in the risk-free rate lead to large changes in the Gordon price. In our simulations (where the conditional mean dividend growth rate and conditional mean risk-free rate change over time), when the value of the equity premium is low, small changes in the conditional means of dividend growth rates or risk-free rates also lead to large prices changes, *i.e.* volatility.

rises *sharply* as the ex ante equity premium grows larger (recall that the Panel A vertical axis has a compressed log scale in Figures 2 through 7). Given these contrasting patterns, the return volatility moment is unlikely, *by itself,* to be causing the sharply rising joint test statistic.

Consider now the t-test statistics on the price-dividend ratio, plotted in Panel D of Figures 2 through 7. Notice that in all cases the t-test on the price-dividend ratio jumps up sharply as the ex ante equity premium rises above 3%. Thus the sharply increasing χ^2 statistics we saw in Panel A of the three figures are likely due in large part to information contained in the price-dividend ratio. However, return volatility reinforces and amplifies the sharp rejection of premia above 4% that the dividend yield also leads us to. In terms of the three moments we have considered in the joint χ^2 and univariate t-test statistics, it is evident that the upper range of ex ante equity premia consistent with the experience of the last half century in the US is limited by the high average S&P 500 price-dividend ratio (or equivalently, the low average S&P 500 dividend yield) together with the high volatility of returns. This result is invariant to the way we model dividend growth, interest rates, or the equity premium process. Even an ex ante equity premium of 5% produces economies with price-dividend ratios and return volatilities so low that they are greatly at odds with the high return volatility and high average price-dividend ratio observed over the past half century in the US.

D.1 Sensitivity to Declining Dividends Through Use of the Price-Dividend Ratio

To ensure that our results are not driven by a single moment of the data, in particular a moment of the data possibly impacted by declining dividend payments in the US, we perform two checks. First, in Models 4 through 8 we incorporate higher dividends and dividend growth rates than observed in US corporate dividends. This is to adjust for the practice, adopted widely beginning in the late 1970s, of US firms delivering cashflows to investors in ways (such as share repurchases) which are not recorded as corporate dividends. As we previously reported, Models 4 through 8 (the models that incorporate higher cashflows to investors than recorded by S&P 500 dividend payments, *i.e.*, the models that use cashflows *including* share repurchases) are best able to account for the observed US data. Reassuringly, the estimate of the equity premium emerging from Models 4 through 8 is virtually identical to that produced by the models that exclude share repurchases.

Our second check is to perform joint tests excluding the price-dividend ratio. Any sensitivity to mismeasurement of the price-dividend ratio should be mitigated if we consider joint test statistics that are based only the ex post equity premium and return volatility, excluding the price-dividend ratio. These (unreported) joint tests confirm two facts. First, when the joint tests exclude the price-dividend ratios, the value of the χ^2 statistic rises less sharply for values of the ex ante equity premium above 4%. Essentially, this indicates that using two moments of the data (excluding the price-dividend ratio) rather than all three makes it more difficult to identify the minimum test statistic value and thus more difficult to identify our estimate of the ex ante equity premium. This confirms our earlier intuition that the price-dividend ratio is instrumental in determining the steep rise of the joint test statistic in Panel A of Figures 2 through 7. Second, and most importantly, the minimum test statistic is still typically achieved for models with an ex ante equity premium value between 3% and 4%. For some of the models, the minimum test statistic is 25 or 50 basis points lower than that found when basing joint tests on the full set of three moments. For a few models, the minimum test statistic is 25 or 50 basis points higher. Again Models 1 through 3 are rejected for every value of the ex ante equity premium, and again for Models 4 through 8 the range of ex ante equity premia that are not rejected is narrow.

E Investors' Model Uncertainty

We have been careful to explore the impact of estimation uncertainty by simulating from the sampling distribution of our model parameters, and to explore the impact of model specification choice (and implicitly model *mis*specification) by looking at a variety of models for interest rates, dividend growth rates, and equity premium, ranging from constant rate models to various ARMA specifications, with and without trends and breaks in the equity premium and dividend growth rates. Comparing distributions of financial statistics emerging from this range of models to the outcome observed in the US over the last half century leads us to the conclusion that the range of true ex ante equity premia that could have generated the US experience is fairly narrow, under 100 basis points, centered roughly on 3.5%. We have not yet addressed, however, the impact of *investor* uncertainty regarding the true fundamental value of the assets being priced. Up to this point, all simulated prices and returns have been generated with knowledge of the (fundamental) processes generating interest rates and dividends.

It is impossible to be definitive in resolving the impact of investor uncertainty on prices and returns. To do so we would have to know what (incorrect) model of fundamental valuation investors are actually using. We can nonetheless focus our attention on procedures likely to be less affected by investor uncertainty than others. Up to this point, the joint tests we have used to identify the plausible range of ex ante equity premia have employed the observed return volatility over the last half century in the US and the volatility of returns produced in our simulated economies. However, investor uncertainty could cause market prices to over- and under-shoot fundamental prices, impacting return volatility, perhaps significantly. A joint test statistic based on only the mean equity premium and the mean price-dividend ratio, however, should be relatively immune to the impact of investor uncertainty. (In the absence of extended price bubbles, *mean* yields should not be impacted greatly by temporary pricing errors.) Thus we now consider the joint χ^2 test statistic based on only the mean return and the mean price-dividend ratio. Figure 8, Panel A plots the test statistics for Models 1, 2, and 3, Panel B plots the test statistics for Models 4, 5, and 6, Panel C plots the test statistics for Models 7, 8, and 9, and Panel D plots the test statistics for Models 10, 11, and 12, with a log scale for the vertical axis in all cases.

Figure 8 goes about here.

First consider results for Models 1 through 4, shown in Panels A and B of Figure 8. These are the base model with no trends or breaks, and models which incorporate only one feature (trend or break in the equity premium or dividend growth rate) at a time. We see again that Model 1 is rejected outright for every value of the ex ante equity premium, at the 10% level of significance, and we see again that adding trends or breaks, even one-at-a-time, improves performance. Now Model 2 (incorporating an 80 basis point downward trend in the equity premium) and Model 4 (incorporating the increased cashflow growth rate) are not rejected over narrow ranges at the 10% significance level. We find that Models 5, 6, 7, and 8, all incorporating trends and breaks in the equity premium and dividend growth rate processes and shown in Panels B and C of Figure 8, deliver a wide range of ex ante equity premia which cannot be rejected at any conventional level of statistical significance. We also see that Model 9 in Panel C, incorporating a trend (of 30 basis points) and a break (of 50 basis points) in the equity premium, performs similarly to Model 2, which has only a trend of 80 basis points (neither model incorporates a cashflow change). In Panel D we see Model 10 which has a deterministic equity premium with trends and breaks. This model's performance is also similar to Model 2, but slightly worse, rejected at the 10% level at every ex ante equity premium. Also in Panel D we see that Models 11 and 12, which do not incorporate parameter estimation uncertainty, are almost everywhere rejected. (In contrast to the joint test shown in Panel A of Figure 7, based on all three moments, we find that Model 11 is not rejected only for the 3% value of the ex ante equity premium.)

Overall, the value of the ex ante equity premium at which the joint test statistic is minimized (*i.e.*, our estimate of the ex ante equity premium) is not particularly affected by our having based the joint tests on two moments of the data rather than the original three, nor is our selection of plausible models for the equity premium process. Across the models, the highest estimate of the ex ante equity premium is roughly 4% (for Model 4) and the lowest is 3% (for Models 11 and 12). With the joint tests based on two moments, all models support (*i.e.*, do not reject) broader ranges of the ex ante equity premium, with the range widest for Models 4 through 8 (now spanning roughly 200 basis points for any given model, from ex ante equity premium values as low as 2.25% for Model 7 to values as high as 4.5% for Model 4). This widening of the range of plausible ex ante equity premia is consistent with a decline in the power of our joint test, presumably from omitting an important moment of the data, the return volatility. The widening of the range of plausible ex ante equity premia is also consistent with investors being uncertain about the true fundamental value of the assets being priced. The last half century of data from the US will be less informative as investor uncertainty about the processes governing fundamentals exaggerates the volatility of returns and hence reduces the precision of estimates of the ex ante equity premium.

To the extent that market prices are set in an efficient market dominated by participants with models of dividend growth rates and interest rates that reflect reality, these ranges of plausible ex ante equity premia based on only the two-moment joint test are overly wide. Still these ranges are useful for putting a loose bound on the likely range of the ex ante equity premium.

F Bootstrapped Test Statistics

Up to this point, all of our test statistics have relied on asymptotic distribution theory for critical values. The asymptotic distributions should be reliable both because we are looking at averages over independent events (our simulations are by construction independent) and because we have many simulations over which to average (2,000). Nonetheless, it is straightforward to use our simulated test statistics to bootstrap the distribution of the test statistics, thus we do so. While use of the bootstrap produces small quantitative changes to our results, our main findings remain unchanged. The best estimate of the mean ex ante equity premium and the range of plausible ex ante equity premia and equity premium models do not budge.

IV Conclusions

The equity premium of interest in theoretical models is the extra return investors anticipate when purchasing risky stock instead of risk-free debt. Unfortunately, we do not observe this ex ante equity premium in the data. We only observe the returns that investors actually receive ex post, after they purchase the stock and hold it over some period of time during which random economic shocks impact prices. US stocks have historically returned roughly 6% more than risk-free debt. Ex post estimates provided by recent papers suggest the US equity premium may be falling in recent years. However, all of these estimates are imprecise, and there is little consensus emerging about the true value of the ex ante equity premium. The imprecision and lack of consensus both hamper efforts to use equity premium estimates in practice, for instance to conduct valuation or to perform capital budgeting. The imprecision of equity premium estimates also complicates resolution of the equity premium puzzle and makes it difficult to determine if the equity premium changes over time.

In order to determine the most plausible value of the ex ante equity premium and the most plausible restrictions on how the equity premium evolves over time, we have exploited information not just on the ex post equity premium and the precision of this estimate, but also on related financial statistics that define the era in which this ex post equity premium was estimated. The idea of looking at related fundamental information in order to improve the estimate of the mean ex ante equity premium follows recent work on the equity premium which has also sought improvements through the use fundamental information like the dividend and earnings yields (Fama and French, 2002, and Jagannathan, McGrattan, and Scherbina, 2000), higher-order moments of the excess return distribution (Maheu and McCurdy, 2007) and return volatility and price movement directions (Pástor and Stambaugh, 2001).

Our central insight is that the knowledge that a low dividend yield, high ex post equity premium, high return volatility, and high Sharpe ratio all occurred together over the last five decades tells us something about the mean ex ante equity premium and the likelihood that the equity premium is time-varying with trends and breaks. Certainly, if sets of these financial statistics are considered together, we should be able to estimate the equity premium more accurately than if we were to look only at the ex post equity premium. This insight relies on the imposition of some structure from economic models, but our result is quite robust to a wide range of model structures, lending confidence to our conclusions.

We employ the simulated method of moments technique and build on the dividend discounting method of fundamental valuation of Donaldson and Kamstra (1996) to estimate the ex ante equity premium. We reject as inconsistent with the US experience all but a narrow range of values of the mean ex ante equity premium and all but a small number equity premium time-series models. We do so while incorporating model estimation uncertainty and allowing for investor uncertainty about the true state of the world. The range of ex ante equity premia that is most plausible is centered very close to 3.5% for virtually every model we consider. The models of the equity premium not rejected by our model specification tests – that is, consistent with the experience of the US over the last half century – incorporate substantial autocorrelation, a structural break, and/or a gradual downward trend in the equity premium process. For these models, the range of ex ante equity premia supported by our tests is very narrow, plus or minus 50 basis points around 3.5%. All together, our tests strongly support the notion that the equity premium process over the last half century in the US was very unlikely to have been constant, was likely to have demonstrated at least one sharp downward break, and was likely to have demonstrated a gradual downward trend.

References

Bagwell, L., Shoven, J., 1989, Cash distributions to shareholders, *Journal of Economic Perspectives* 3, 129-149.

Bansal, Ravi and Amir Yaron, 2004, Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance* 59(4), 1481-1509.

Bansal, Ravi and Christian Lundblad, 2002, Market efficiency, asset returns, and the size of the risk premium in global equity markets, *Journal of Econometrics* 109, 195-237.

Bekaert, Geert and Campbell R. Harvey, 1995, Time-varying world market integration, *Journal of Finance* 50, 403-444.

Campbell, John Y., 1991, A Variance Decomposition for Stock Returns, *The Economic Journal* 101, 157-179.

Campbell, John Y. and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195-228.

Campbell, John. Y., and Robert. J. Shiller, 1998, Valuation Ratios and the Long-Run Stock Market Outlook, *Journal of Portfolio Management* 24(6), 11-26.

Carrasco, Marine and Jean-Pierre Florens, 2002, Simulation-based method of moments and efficiency, *Journal of Business and Economic Statistics* 20(4), 482-492.

Claus, James and Jacob Thomas, 2001, Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets, *Journal of Finance* 56(5), 1629-1666.

Cochrane, John H., 2001 *Asset Pricing,* Princeton NJ, Princeton University Press.

Corradi, Valentina and Norman R. Swanson, 2005, Bootstrap specification tests for diffusion processes, *Journal of Econometrics* 124(1), 117-148.

Diamond, Peter A., 2000, What stock market returns to expect for the future? *Social Security Bulletin* 63(2), 38-52.

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2007, The Worldwide Equity Premium: A Smaller Puzzle, in Rajnish Mehra, ed.: *Handbook of Investments: Equity Risk Premium*, Amsterdam: Elsevier.

Donaldson, R. Glen and Mark J. Kamstra, 1996, A New Dividend Forecasting Procedure That Rejects Bubbles in Asset Prices, *Review of Financial Studies* 9, 333-383.

Duffie, Darrell and Kenneth J. Singleton, 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica* 61(4), 929-952.

Fama, Eugene F. and Kenneth R. French, 1997, Industry Costs of Equity, *Journal of Financial Economics* 43, 153-193.

Fama, Eugene F. and Kenneth R. French, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay, *Journal of Financial Economics* 60, 3-43.

Fama, Eugene F. and Kenneth R. French, 2002, The Equity Premium, *Journal of Finance* 57 (2), 637-659.

Graham, John, and Campbell Harvey, 2005, The Long-Run Equity Risk Premium, *Finance Research Letters* 2, 185-194.

Gregory, Allan W. and Gregor W. Smith, 1991, Calibration as Testing: Inference in Simulated Macroeconomic Models. *Journal of Business and Economic Statistics* 9(3), 297-303.

Gordon, Myron, 1962, *The Investment, Financing and Valuation of the Corporation*, Homewood, IL: Irwin.

Grullon, Gustavo and Roni Michaely, 2002, Dividends, Share Repurchases, and the Substitution Hypothesis, *Journal of Finance* 57(4), 1649-1684.

Hannan, E.J., 1980, The Estimation of the Order of an ARMA Process, *The Annals of Statistics* 8, 1071-1081.

Hodrick, Robert J., 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for inference and Measurement, *Review of Financial Studies* 5(3), 357-386.

Hull, John, 1993, *Options, Futures and Other Derivative Securities,* Englewood Cliffs, NJ: Prentice-Hall.

Jagannathan, Ravi, Ellen R. McGrattan, and Anna Scherbina, 2000, The Declining US Equity Premium, NBER Working Paper W8172, *Federal Reserve Bank of Minneapolis Quarterly Review* 24, 3-19.

Jain, Pankaj K. 2005, Financial Market Design and the Equity Premium: Electronic versus Floor Trading, *Journal of Finance* 60(6), p2955-2985

Jorion, Philippe and William N. Goetzmann, 1999, Global Stock Markets in the Twentieth Century, *Journal of Finance* 54(3), 953-980.

Kendall, Maurice and Alan Stuart, 1977, *The Advanced Theory of Statistics,* Volume 1, 4th edition, New York: MacMillan Publishing.

Kleidon, Allan W., 1986, Variance Bound Tests and Stock Price Valuation Models, *Journal of Political Economy* 94, 953-1001.

Kocherlakota, Narayana R., 1996, The Equity Premium: It's Still a Puzzle, *Journal of Economic Literature* 34(1), 42-71.

Maheu, John M. and Thomas H. McCurdy, 2007, How useful are historical data for forecasting the long-run equity return distribution? forthcoming, *Journal of Business and Economic Statistics*.

Mankiw, N. Gregory, David H. Romer, and Matthew D. Shapiro, 1991, Stock Market Forecastability and Volatility: A Statistical Appraisal, *Review of Economic Studies* 58(3), 455-477.

McFadden, Daniel, 1989, A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, *Econometrica* 47, 995-1026.

Mehra, Rajnish, 2003, The Equity Premium: Why Is It A Puzzle? *Financial Analysts Journal* 59(1) 54-69.

Mehra, Rajnish and Edward C. Prescott, 1985, The Equity Premium: A Puzzle, *Journal of Monetary Economics* 15, 145-161.

Mehra, Rajnish and Edward C. Prescott, 2003, The Equity Premium in Retrospect, in Constantinides, G.M., M. Harris, and R. Stulz, eds.: *Handbook of the Economics of Finance* (Elsevier).

Merton, Robert C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics* 8, 323-361.

Nelson, Daniel B., 1991, Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

Newey, Whitney K. and Kenneth D. West, 1987, A Simple, Positive, Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.

Pakes, Ariel and David Pollard, 1989, Simulation and the Asymptotics of Optimization Estimators, *Econometrica* 57, 1027-1057.

Pástor, Ľuboš and Robert Stambaugh, 2001, The Equity Premium and Structural Breaks, *Journal of Finance* 56(1), 1207-1239.

Rozeff, Michael S., 1984, Dividend yields are equity risk premiums, *Journal of Portfolio Management* 11(1), 68-75.

Scott, Louis O., 1985, The Present Value Model of Stock Prices: Regression Tests and Monte Carlo Results, *The Review of Economics and Statistics* 67(4), 599-605.

Siegel, Jeremy J., 1999, The shrinking equity premium, *Journal of Portfolio Management*, 26(1), 10-17.

Shiller, Robert J., 1984, Stock Prices and Social Dynamics, *Brookings Papers on Economic Activity*, 457-498.

Siegel, Jeremy J. and Richard H. Thaler, 1997, Anomalies: The Equity Premium Puzzle, *Journal of Economic Perspectives* 11(1), 191-200.

Timmermann, Allen G., 1993, How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices, *The Quarterly Journal of Economics* 108(4), 1135-1145.

Timmermann, Allen G., 1995, Tests of Present Value Models with a Time-Varying Discount Factor, *Journal of Applied Econometrics* 10(1), 17-31.

Welch, I., 2000, Views of Financial Economists on the Equity Premium and on Professional Controversies, *Journal of Business* 73(4), 501-537.

West, Kenneth D., 1988a, Bubbles, Fads and Stock Price Volatility Tests: A Partial Evaluation, *Journal of Finance* 43, 639-656.

West, Kenneth D., 1988b, Dividend Innovations and Stock Price Volatility, *Econometrica* 56, 37-61.

White, Halbert, 1984, *Asymptotic Theory for Econometricians,* Orlando: Academic Press.

Appendices

Appendix 1: Models for Generating Data

In creating distributions of financial variables modeled on the US economy, we must generate the fundamental factors that drive asset prices: dividends and discount rates (where the discount rate is defined as the risk-free rate plus a possibly time-varying equity premium). Thus we must specify time-series models for dividend growth, interest rates, and ex ante equity premia so that our Monte Carlo simulations will generate dividends and discount rates that share key features with observed S&P 500 dividends and US discount rates. We consider a range of models to generate data in our simulations, as outlined in Table I. Each model incorporates specific characteristics that define the way we generate interest rates and dividend growth rates, and each model makes specific assumptions about the way the ex ante equity premium evolves over time, if indeed it does evolve over time. In providing further information about these defining aspects of our models, we consider each model feature from Table I in turn, starting with the time-series processes for interest rates, dividend growth rates, and the ex ante equity premium.

A1.1 Processes for the Interest Rate, Dividend Growth Rate and the Ex Ante Equity Premium

The interest rate and dividend growth rate series we generate are calibrated to the time-series properties of data observed in the US over the period 1952 to 2004. We considered the ability of various time-series models to eliminate residual autocorrelation and ARCH (evaluated with LM tests for residual autocorrelation and for ARCH, both using 5 lags), and we evaluated the log likelihood function and Bayesian Information Criterion (BIC) across models. Although we will describe the process of model selection one variable at-a-time, our final models were chosen using a Full Information Maximum Likelihood (FIML) systems equation estimation and a joint-system BIC optimization.

Economic theory admits a wide range of possible processes for the risk-free interest rate, from constant to autoregressive and highly non-linear heteroskedastic forms. We find that in practice, both AR(1) and ARMA(1,1) models of the logarithm of interest rates, based on the model of Hull (1993, page 408), perform well in capturing the time-series properties of observed interest rates. We also find the AR(1) and ARMA(1,1) specifications perform comparably to one another, markedly dominating the performance of other specifications including higher order models like ARMA(2,2). An attractive feature of modeling the log of interest rates is that doing so restricts nominal interest rates to be positive. Finally, we find standard tests for normality of the error term (and hence conditional log-normality of interest rates) do not reject the null of normality.

Since dividend growth rates have a minimum value of -100% and no theoretical maximum, a natural choice for their distribution is the log-normal. Thus we model the log of 1 plus the dividend growth rate, and we find that both a MA(1) and an AR(1) specification fit the data well, removing evidence of residual autocorrelation and ARCH at five lags. These specifications are preferred on the basis of the same criteria used to choose the specification for modeling interest rates. As with the interest rate data, we find standard tests for normality of the error term (and hence conditional log-normality of dividend growth rates) do not reject the null of normality.

Most of our models incorporate an ex ante equity premium that follows an ARMA process emerging from Merton's (1980) conditional CAPM. Merton's conditional CAPM is expressed in terms of returns in excess of the risk-free rate, or, in other words, the period-by-period equity premium. For the i^{th} asset,

$$E_t(r_{i,t}) = \lambda \; cov_{t-1}(r_{i,t} r_{m,t}), \tag{10}$$

where $r_{i,t}$ are excess returns on the asset, $r_{m,t}$ are excess returns on the market portfolio, cov_{t-1} is the time-varying conditional covariance between excess returns on the asset and on the market portfolio, and E_t is the conditional-expectations operator incorporating information available to the market up to but not including the beginning of period t. λ is a parameter of the model, described below.

For the expected excess market return, (10) becomes

$$E_t(r_{m,t}) = \lambda \; var_{t-1}(r_{m,t}) \tag{11}$$

where var_{t-1} is the market time-varying conditional variance. Merton (1980) argues that λ in (11) is the weighted sum of the reciprocal of each investor's coefficient of relative risk aversion, with the weight being related to the distribution of wealth among individuals.

Equation (11) defines a time-varying equity premium but has the equity premium varying only as a function of time-varying conditional variance. Following Bekaert and Harvey (1995), it is possible to allow λ in Equation (11) to vary over time by making it a parametric function of conditioning variables (indicated below as Z_{t-1}). The functional form Bekaert and Harvey employ (in Equation (12) of their paper) is exponential, restricting the price of risk to be positive:

$$\lambda_{t-1} = exp\left(\delta' Z_{t-1}\right). \tag{12}$$

Shiller (1984), Rozeff (1984), Campbell and Shiller (1988), Hodrick (1992), and Bekaert and Harvey (1995) all document the usefulness of dividend yields to predict returns, so we use lagged dividend yields as our conditioning variable. We make use of a simple ARCH specification to model $var_{t-1}(r_{m,t})$. Once again we calibrate to the S&P 500 over 1952 to 2004, estimating the following model:

$$r_{m,t} = \lambda_{t-1}\ var_{t-1}(r_{m,t}) + e_{m,t} \tag{13}$$

$$var_{t-1}(r_{m,t}) = \omega + \alpha e^2_{m,t-1} \tag{14}$$

$$\lambda_{t-1} = exp\left(\delta_0 + \delta_1 \frac{D_{t-1}}{P_{t-1}}\right). \tag{15}$$

The values of estimated parameters are $\delta_0 = -3.93$, $\delta_1 = 0.277$, $\omega = 0.0194$, and $\alpha = 0.542$. The R^2 of this model is 2.8%.

For our simulations, we model the time-series process of the ex ante time-varying equity premium (denoted π_t) by using the excess return as a proxy for the equity premium:

$$\hat{\pi}_t = \hat{\lambda}_{t-1} v\hat{a}r_{t-1}(r_{m,t}), \tag{16}$$

where $\hat{\lambda}_{t-1} = exp\left(-3.93 + 0.277\frac{D_{t-1}}{P_{t-1}}\right)$, $v\hat{a}r_{t-1}(r_{m,t}) = 0.0194 + 0.542\hat{e}^2_{m,t-1}$, and $\hat{e}_{m,t-1} = r_{m,t-1} - \hat{\pi}_{t-1}$. The time-varying equity premium we estimate here, $\hat{\pi}_t$, follows a strong AR(1) time-series process, similar to that of the risk-free interest rate,[17] so that when the equity premium is perturbated it reverts to its mean slowly. This permits slightly more volatile returns in our simulations than would otherwise be the case. The best way to see the impact of this slow mean reversion of the equity premium on our simulations is to compare Models 9 and 10. Model 9 has a conditionally time-varying equity premium (together with a trend and break in the premium) while Model 10 is identical except the equity premium does not conditionally vary. We find standard tests for normality of the error term (and hence conditional log-normality of the equity premium) show some evidence of non-normality when estimated as a single equation, but less or no evidence if estimated in a system of equations with the interest rate and dividend growth rate equations.

Hence we generate the ex ante equity premia, interest rate, and dividend growth rate series as autocorrelated series with jointly normal error terms, calibrated to the degree of autocorrelation observed in the US data. The processes we simulate also mimic the covariance structure between the residuals from the time-series models of equity premia, interest rates, and dividend growth rates as estimated using US data. We adjust the mean and the standard deviation of these log-normal processes to generate the desired level and variability for each when they are transformed back into levels. The coefficients and error covariance structure are estimated with FIML (very similar results are obtained using iterative GMM and Newey and West, 1987, heteroskedasticity and autocorrelation consistent covariance estimation).

To give a sense for what our estimated models for interest rates, dividend growth rates, and the equity premium look like, we present in Table A.I the estimated parameters of Model 1, which incorporates an AR(1) model for interest rates (r), a MA(1) model for dividend growth rates (g), and an AR(1) model for the ex ante equity premium (π).

[17] The mean of the estimated equity premium from this model is 5.8% and its standard deviation is 2.2%. An AR(1) model of the natural logarithm of the equity premium has a coefficient of 0.79 on the lagged equity premium, with a standard error of 0.050 and an R^2 of 0.83.

Table A.I
Estimated Parameters of Model 1

$log(r_t)$	$=$	-0.214	$+0.929\ log(r_{t-1})$	$+\epsilon_{r,t}$
		(0.262)	(0.086)	
$log(1+g_t)$	$=$	0.0516	$+0.454\ \epsilon_{g,t-1}$	$+\epsilon_{g,t}$
		(0.0063)	(0.084)	
$log(\hat{\pi}_t)$	$=$	-0.562	$+0.851\ log(\hat{\pi}_{t-1})$	$+\epsilon_{\pi,t}$
		(0.230)	(0.070)	

In Table A.I, standard errors of the estimated coefficients are shown in parentheses. The covariance of $\epsilon_{r,t}$ and $\epsilon_{g,t}$ equals 0.00240, the covariance of $\epsilon_{r,t}$ and $\epsilon_{\pi,t}$ equals -0.0117, and the covariance of $\epsilon_{g,t}$ and $\epsilon_{\pi,t}$ equals 0.0018. The variance of $\epsilon_{r,t}$ equals 0.0890, the variance of $\epsilon_{g,t}$ equals 0.000986, and the variance of $\epsilon_{\pi,t}$ equals 0.0648. The adjusted R^2 for the interest rate equation is 72.9%, the adjusted R^2 for the dividend growth rate equation is 30.0%, and the adjusted R^2 for the equity premium equation is 79.5%.

A1.2 Allowing a Downward Trend in the Ex Ante Equity Premium Process

Pástor and Stambaugh (2001), among others, provide evidence that the equity premium has been trending downward over the sample period we study, finding a modest downward trend of roughly 0.80% in total since the early 1950s, with much of the difference coming from a steep decline in the 1990s. Their study of the equity premium has the premium fluctuating between about 4% and 6% since 1834. Given this evidence and the fact that we calibrate to data starting in the 1950s, we investigate a 0.80% trend in the equity premium, and when modeling a trend with a break we limit ourselves to a 0.30% trend with an additional 50 basis point break, as discussed below. This is accomplished in conjunction with setting the ex ante equity premium to follow an AR(1) process.

A1.3 Allowing a Structural Break in the Equity Premium Process

Pástor and Stambaugh (2001) estimate the probability of a structural break in the equity premium over the last two centuries. They find fairly strong support for there having been a structural break over the 1990s which led to a 0.5% drop in the equity premium. An aggressive interpretation of their results would have the majority of the drop in the equity premium over the 1990s occurring at once. We decide to adopt a one-time-drop specification because doing so makes our results more conservative (*i.e.* produces a wider confidence interval for the ex ante equity premium). Spreading the drop in the premium across several years serves only to narrow the range of ex ante equity premium consistent with the US returns data over the last 50 years, which would only bolster our claims to provide a much tighter confidence interval about the estimate of the ex ante equity premium. Thus we incorporate an abrupt 50 basis point drop in the equity premium in some of the models we consider. We time the drop to coincide with 1990, 39 years into our simulation period. This feature of the equity premium process can be accomplished with or without incorporating other features discussed above.

A1.4 Allowing for Sampling Variability in Generating Parameters

Our experiments are motivated by the large sampling variability of the ex post equity premium, but when we produce our simulations we have to first estimate the parameter values for the time-series models of dividend growth rates, interest rates, and ex ante equity premia. These estimates themselves incorporate sampling variability. Fortunately, estimates of the sampling variability are available to us through the covariance matrix of our parameters, so we can incorporate uncertainty about the true values of these parameters into our simulations. We estimate our system of equations (the dividend growth rate, interest rate, and the ex ante equity premium equation) jointly with FIML, and generate for *each* simulation an independent set of parameters drawn randomly from the joint limiting normal distribution of these parameter estimates (including the variance and covariance of the equation residuals) subject to some technical considerations[18] and data consistency checks.[19] This process accounts for possible variability in the true state of the world that generates dividends, interest rates, and ex ante equity premia.

To illustrate, for Model 1 reported in Table A.I,

[18]The time-series models must exhibit stationarity, the growth rate of dividends must be strictly less than the discount rate, and the residual variances must be greater than zero.

[19]The parameters must generate mean interest rates, dividend growth rates, and ex post equity premia that lie within three standard deviations of the US data sample mean. Also, the limiting price-dividend ratio must be within 50 standard deviations of the mean US price-dividend ratio. This last consistency check rules out some extreme simulations generated when the random draw of parameters leads to near unit root behavior. The vast majority of simulations do not exhibit price-dividend ratios that are more than a few standard deviations from the mean of the US data.

$$
\begin{array}{lclll}
log(r_t) & = & \alpha_r & +\rho_r \; log(r_{t-1}) & +\epsilon_{r,t} \\
log(1+g_t) & = & \alpha_g & +\theta_g \; \epsilon_{g,t-1} & +\epsilon_{g,t} \\
log(\hat{\pi}_t) & = & \alpha_\pi & +\rho_\pi \; log(\hat{\pi}_{t-1}) & +\epsilon_{\pi,t},
\end{array}
$$

the estimated covariance matrix of the parameter estimates is shown in Table A.II.

Table A.II
Estimated Covariance Matrix for Model 1 Parameters

	α_r	ρ_r	α_g	θ_g	α_π	ρ_π
α_r	0.068705	0.022307	-.000051933	.000226443	-0.012165	-0.003511
ρ_r	0.022307	0.007436	-.000040346	.000114831	-0.004730	-0.001401
α_g	-0.000052	-0.000040	0.000039674	.000025651	0.000153	0.000031
θ_g	0.000226	0.000115	0.000025651	.007086714	0.001699	0.000454
α_π	-0.012165	-0.004730	0.000153376	.001699151	0.052664	0.015791
ρ_π	-0.003511	-0.001401	0.000031495	.000453874	0.015791	0.004844

The top-left element of Table A.II, equal to 0.068705, is the variance of the parameter estimate of α_r. The entry below the top-left element, equal to 0.022307, is the covariance between the estimate of α_r and ρ_r, and so on. The *estimated covariance matrix* of the equation residual variances is shown in Table A.III. (The variances themselves are reported in Section A1.1, as are the parameter estimates of the mean.)

Table A.III
Estimated Covariance Matrix of Model 1 Residual Variances

	ϵ_r^2	$\epsilon_r\epsilon_g$	$\epsilon_r\epsilon_\pi$	ϵ_g^2	$\epsilon_g\epsilon_\pi$	ϵ_π^2
ϵ_r^2	0.0000944	$1.9729{\cdot}10^{-6}$	$-8.351{\cdot}10^{-7}$	$-1.902{\cdot}10^{-7}$	$-1.564{\cdot}10^{-6}$	$-1.69{\cdot}10^{6}$
$\epsilon_r\epsilon_g$	$1.9729{\cdot}10^{-6}$	$8.5163{\cdot}10^{-7}$	$1.0437{\cdot}10^{-6}$	$4.3066{\cdot}10^{-8}$	$-1.602{\cdot}10^{-7}$	$9.1448{\cdot}10^{-7}$
$\epsilon_r\epsilon_\pi$	$-8.351{\cdot}10^{-7}$	$1.0437{\cdot}10^{-6}$	0.0000797	$1.8827{\cdot}10^{-7}$	$5.001{\cdot}10^{-6}$	-0.000044
ϵ_g^2	$-1.902{\cdot}10^{-7}$	$4.3066{\cdot}10^{-8}$	$1.8827{\cdot}10^{-7}$	$4.8337{\cdot}10^{-8}$	$9.6885{\cdot}10^{-8}$	$1.3458{\cdot}10^{-6}$
$\epsilon_g\epsilon_\pi$	$-1.564{\cdot}10^{-6}$	$-1.602{\cdot}10^{-7}$	$5.001{\cdot}10^{-6}$	$9.6885{\cdot}10^{-8}$	$3.5567{\cdot}10^{-6}$	0.0000203
ϵ_π^2	$-1.69{\cdot}10^{-6}$	$9.1448{\cdot}10^{-7}$	-0.000044	$1.3458{\cdot}10^{-6}$	0.0000203	0.0005009

The top-left element, equal to 0.0000944, is the variance of ϵ_r^2. The entry below the top-left element, equal to $-1.9729{\cdot}10^{-6}$, is the covariance between the estimate of ϵ_r^2 and the product of ϵ_r and ϵ_g, and so on.

Exploiting block diagonality of the parameters of the mean and variance, and asymptotic normality of all the estimated parameters, we generate two sets of normally distributed random variables.

Each set is independent of the other, the first set of six having the covariance matrix from Table A.II with means equal to the parameter estimates listed in Table A.I, and the second set of six having the covariance matrix from Table A.III, with means equal to the equation residual covariances listed in Section A1.1. This set of 12 random variables is then used to simulate interest rates, dividend growth rates, and equity premia, subject to the consistency checks footnoted earlier.

A1.5 Allowing for Disappearing Dividends

An issue with our calibration to dividends is the impact of declining dividend payments in the US. This phenomenon is a result of a practice adopted widely beginning in the late 1970s, whereby US firms have been increasingly delivering cashflows to investors in ways not recorded as corporate dividends, such as share repurchases. Fama and French (2001) document the widespread decline of regular dividend payments starting in 1978, consistent with evidence provided by Bagwell and Shoven (1989) and others. Fama and French find evidence that the disappearance of dividends is in part due to an increase in the inflow of new listing to US stock exchanges, representing mostly young companies with the characteristics of firms that would not be expected to pay dividends, and in part due to a decline in the propensity of firms to pay dividends. Fama and French find only a small decline in the probability to pay dividends among the firms that we calibrate to, those in the S&P 500 index.

Consistent with Fama and French, we find no evidence of a break in our data on dividend growth rates. Though dividend *yields* on the S&P 500 index have dropped dramatically over time, dividend growth rates have not. The decline in yields has been a function of prices rising faster than dividends since 1978, not dividends declining in any absolute sense. From 1952 through 1978, the year Fama and French document as the year of the structural break in dividend payments, dividend growth rates among the S&P 500 firms have averaged 4.9% with an annual standard deviation of 3.9%, and from 1979 to 2000 the dividend growth rates have averaged 5.5% with an annual standard deviation of 3.8%, virtually indistinguishable from the pre-1979 period. Time series properties pre- and post-1978 are also very similar across these two periods. Consistent with this stability of dividend growth pre- and post-1978 and Bagwell and Shoven's documentation of increased share repurchases in the 1980s, *earnings growth rates* of firms in the S&P 500 index have accelerated since the 1952-1978 period, from 6.8% pre-1979 to 7.8% post-1978. Similar to the dividend growth rate data, the time-series properties of the earnings growth rate data did not change.

In order to determine the sensitivity of our experiments to mismeasurement of cashflows to investors, we consider a dividend growth rate process with a structural break 27 years into the time series to correspond to a possible break in our dividend data for the S&P 500 data after 1978. We calibrate to the S&P 500 earnings data mean growth rate increase over 1979-2000, an upward shift of 100 basis points, to proxy for the increase in total cashflows to investors. That is, we increase the growth rate of dividends by 5 basis points a year for 20 years, starting in year 27 of the simulation (corresponding to 1978 for the S&P 500 data), to increase the mean growth rate of our dividend growth series 100 basis points, mimicking the proportional increase in earnings growth rates.

Appendix 2: Further Details on the Simulations

A2.1 Fundamentals

We define P_t as a stock's beginning-of-period-t price and E_t as the expectations operator conditional on information available up to but not including the beginning of period t. The discount rate (r_t, which equals the risk-free rate plus the equity premium) is the rate investors use to discount payments received during period t (*i.e.*, from the beginning of period t to the beginning of period $t+1$). Recall that investor rationality requires that the time t market price of a stock, which will pay a dividend D_{t+1} one period later and then sell for P_{t+1}, satisfy Equation (3):

$$P_t = E_t\left\{\frac{P_{t+1} + D_{t+1}}{1 + r_t}\right\}. \tag{3}$$

Invoking the standard transversality condition that the expected present value of the stock price P_{t+i} falls to zero as i goes to infinity, and defining the growth rate of dividends during period t as $g_t \equiv (D_{t+1} - D_t)/D_t$, allows us rewrite Equation (3) as:

$$P_t = D_t E_t\left\{\sum_{i=0}^{\infty}\left(\Pi_{k=0}^{i}\left[\frac{1 + g_{t+k}}{1 + r_{t+k}}\right]\right)\right\}. \tag{5}$$

One attractive feature of expressing the present value stock price as in Equation (5), in terms of dividend growth rates and discount rates, is that this form highlights the irrelevance of inflation, at least to the extent that expected and actual inflation are the same. Notice that working with nominal growth rates and discount rates, as we do, is equivalent to working with deflated nominal rates (*i.e.*, real rates). That is, $\frac{1+([g_t-I_t]/[1+I_t])}{1+([r_t-I_t]/[1+I_t])} = \frac{(1+g_t)}{(1+r_t)}$, where I_t is inflation. Working with nominal values in our simulations removes a potential source of measurement error associated with attempts to estimate inflation.

Properties of prices and returns produced by Equation (5) depend in important ways on the modeling of the dynamics of the dividend growth, interest rate, and equity premium processes. For instance, the stock price would equal a constant multiple of the dividend level and returns would be very smooth over time if dividend growth and interest rates were set equal to constants plus independent innovations. However, using models that capture the serial dependence of dividend growth rates, interest rates, and equity premia observed in the data, as we do, would typically lead to time-varying price-dividend ratios and variable returns of the sort we observe in observed stock market data.

A2.2 Numerical Simulation

We now provide details on the numerical simulation which comprises Step 4 of the 5-step procedure outlined in Section I above. That is, we detail for the n^{th} economy the formation of the prices (P_t^n), returns (R_t^n), ex post equity premia ($\hat{\pi}^n$), *etc.* (where $n = 1, \cdots, N$ and $t = 1, \cdots, T$), given dividends, dividend growth rates, risk-free interest rates, and the equity premium of the n^{th} economy: D_t^n, g_{t-1}^n, and $r_{t-1}^n = r_{f,t-1}^n + \pi$.[20] For simplicity, we illustrate our methodology by assuming fixed parameters (no parameter uncertainty), a constant ex ante equity premium, and an AR(1) model for interest rates. Further, to illustrate the procedure required for a moving average error model, we assume a MA(1) process for dividend growth rates. Relaxing these assumptions (the assumptions to incorporate parameter uncertainty, ARMA(1,1) processes for interest rates and dividend growth rates, and a time-varying equity premium) complicates the procedure outlined below only slightly. Note that in our actual simulations we set the initial dividend growth rate and

[20]We set the number of economies, N, at 2,000. This is a sufficiently large number of replications to produce results with very small simulation error.

interest rate to their unconditional means, innovations to zero, and dividends to \$1, then simulate the economies out for 50 periods. At period 51 we start our calculation of market prices, returns, *etc.* (to avoid contaminating the simulations with the initial conditions). For simplicity, we do not include this detail in the description below but for concreteness we describe a similar prototypical simulation.

In terms of timing and information, recall that P_t^n is the stock's beginning-of-period-t price, r_t^n is the rate used to discount payments received during period t and is known at the beginning of period t, D_t^n is paid at the beginning of period t, g_t^n is defined as $(D_{t+1}^n - D_t^n)/D_t^n$ and is not known at the beginning of period t since it depends on D_{t+1}^n, and $E_t\{\cdot\}$ is the conditional expectation operator, with the conditioning information being the set of information available to investors up to but not including the beginning of period t. Finally, recall Equation (5), rewritten to correspond to the n^{th} economy:

$$P_t^n = D_t^n E_t \left\{ \sum_{i=0}^{\infty} \left(\Pi_{k=0}^{i} \left[\frac{1 + g_{t+k}^n}{1 + r_{t+k}^n} \right] \right) \right\}. \tag{17}$$

Returns are constructed as $R_t^n = (P_{t+1}^n + D_{t+1}^n - P_t^n)/P_t^n$, and $\hat{\pi}^n = \overline{R}^n - \overline{r}_f^n$ where $\overline{R}^n = \frac{1}{T}\sum_{t=1}^{T} R_t^n$ and $\overline{r}_f^n = \frac{1}{T}\sum_{t=1}^{T} r_{f,t}^n$.

Based on Equation (17), we generate prices by generating a multitude of possible streams of dividends and discount rates, present-value discounting the dividends with the discount rates, and averaging the results, *i.e.*, by conducting a Monte Carlo integration.[21] Hence we produce prices (P_t^n), returns (R_t^n), ex post equity premia ($\hat{\pi}^n$), and a myriad of other financial quantities, utilizing only dividend growth rates and discount rates. The *exact* procedure by which we conduct this numerical simulation is described below and summarized in Figure A.1. (These steps, labeled Steps 4A through 4C, collectively constitute Step 4 of the 5-step procedure outlined in Section I above.)

[21]According to Equation (17), the stream of dividends and discount rates should be infinitely long, however truncating the stream at a sufficiently distant point in time denoted I leads to a very small approximation error. We discuss this point more fully below.

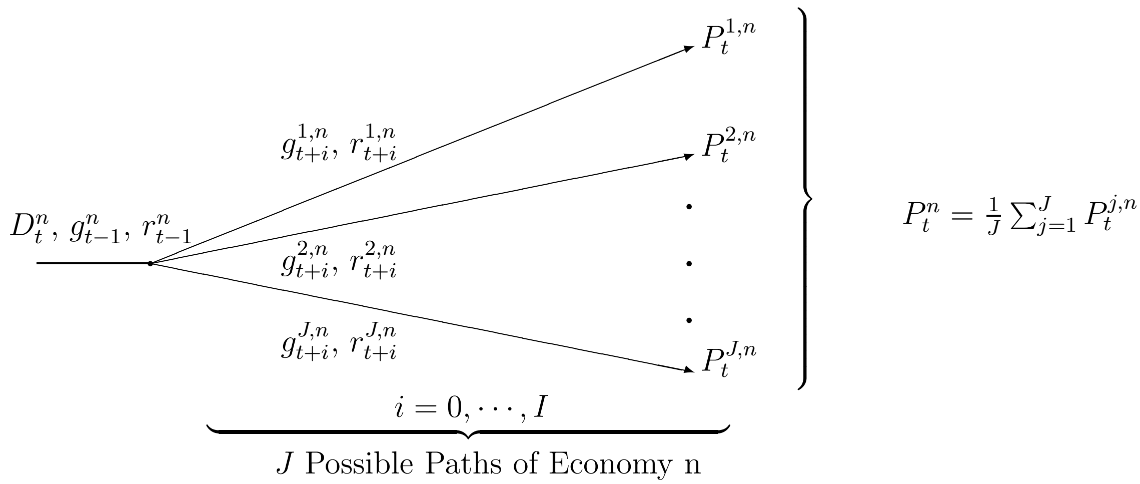


Figure A.1 *Diagram of a Simple Market Price Calculation for the t^{th} Observation of the n^{th} Economy (Steps 4A and 4B)*

Step 4A: In forming P_t^n, the most recent fundamental information available to an investor would be g_{t-1}^n, D_t^n, and r_{t-1}^n. Thus g_{t-1}^n, D_t^n, and r_{t-1}^n must be generated directly in our simulations, whereas P_t^n is calculated based on these g, D, and r. The objective of Steps 4A(i)-(iii) outlined below is to produce dividend growth and interest rates that replicate real-world dividend growth and interest rate data. That is, the simulated dividend growth and interest rates must have the same mean, variance, covariance, and autocorrelation structure as observed S&P 500 dividend growth rates and US interest rates. In terms of Figure A.1, Step 4A forms g_{t-1}^n, D_t^n, and r_{t-1}^n only.

Step 4A(i): Note that since, as described above, the logarithm of one plus the dividend growth rate is modeled as a MA(1) process, $log(1+g_t^n)$ is a function of only innovations, labeled ϵ_g^n. Note also that since the logarithm of the interest rate is modeled as an AR(1) process, $log(r_{f,t}^n)$ is a function of $log(r_{f,t-1}^n)$ and an innovation labeled ϵ_r^n. Set the initial dividend, D_1^n, equal to the total S&P 500 dividend value for 1951 (observed at the end of 1951), and the lagged innovation of the logarithm of the dividend growth rates $\epsilon_{g,0}^n$ to 0. To match the real-world interest rate data, set $log(r_{f,0}^n) = -2.90$ (the mean value of log interest rates required to produce interest rates matching the mean of observed T-bill rates). Then generate two independent standard normal random numbers, η_1^n and ν_1^n (note that the subscript on these random numbers indicates time, t), and form two correlated random variables, $\epsilon_{r,1}^n = 0.319(0.25\eta_1^n + (1-.25^2)^{.5}\nu_1^n)$ and $\epsilon_{g,1}^n = 0.0311\eta_1^n$. These are the simulated innovations to the interest rate and dividend growth rate processes, formed to have standard deviations of 0.319 and 0.0311 respectively to match the data, and to be correlated with correlation coefficient 0.25 as we find in the S&P 500 return and T-bill rate data. Next, form

$log(1+g_1^n) = 0.049+0.64\epsilon_{g,0}^n+\epsilon_{g,1}^n$ and $log(r_{f,1}^n) = -0.35+0.88log(r_{f,0}^n)+\epsilon_{r,1}^n$ to match the parameters estimated on the S&P 500 index data 1952-2004 of these models (using Full Information Maximum Likelihood).[22] Also form $D_2^n = D_1^n(1+g_1^n)$.

Step 4A(ii): Produce two correlated normal random variables, $\epsilon_{r,2}^n$ and $\epsilon_{g,2}^n$ as in Step 4A(i) above, and conditioning on $\epsilon_{g,1}^n$ and $log(r_{f,1}^n)$ from Step 4A(i) produce $log(1+g_2^n) = 0.049+0.64\epsilon_{g,1}^n+\epsilon_{g,2}^n$, $log(r_{f,2}^n) = -0.35+0.88log(r_{f,1}^n)+\epsilon_{r,2}^n$, and $D_3^n = D_2^n(1+g_2^n)$.

Step 4A(iii): Repeat Step 4A(ii) to form $log(1+g_t^n)$, $log(r_{f,t}^n)$, and D_t^n for $t = 3,4,5,\cdots,T$ and for each economy $n = 1,2,3,\cdots,N$. Then calculate the dividend growth rate g_t^n and the discount rate r_t^n (which equals $r_{f,t}^n$ plus the ex ante equity premium).

Step 4B: For each time period $t = 1,2,3,\cdots,T$ and economy $n = 1,2,3,\cdots,N$ we calculate prices, P_t^n. In order to do this we must solve for the expectation of the infinite sum of discounted future dividends conditional on time $t-1$ information for economy n. That is, we must produce a set of possible paths of dividends and interest rates that might be observed in periods $t, t+1, t+2, \cdots$ given what is known at period $t-1$ and use these to solve the expectation of Equation (17). We use the superscript j to index the possible paths of future economies that could possibly evolve from the current state of the economy. In Step 4B(iv) below, we describe how we are able to solve for the expectation of an infinite sum using a *finite* stream of future dividends.

Step 4B(i): Set $\epsilon_{g,t-1}^{j,n} = \epsilon_{g,t-1}^n$ and $log(r_{f,t-1}^{j,n}) = log(r_{f,t-1}^n)$ for $j = 1,2,3,\cdots,J$.[23] Generate two independent standard normal random numbers, $\eta_t^{j,n}$ and $\nu_t^{j,n}$, and form two correlated random variables $\epsilon_{r,t}^{j,n} = 0.319(0.25\eta_t^{j,n} + (1-.25^2)^{.5}\nu_t^{j,n})$ and $\epsilon_{g,t}^{j,n} = 0.0311\eta_t^{j,n}$ for $j = 1,2,3,\cdots,J$.[24] These

[22] Note that by construction these parameters do not match those reported for the system reported in Appendix 1 as this system does not incorporate a time-varying equity premium.

[23] We choose J to lie between 1,000 and 100,000, as needed to ensure the Monte Carlo simulation error in calculating prices and returns is controlled to be less than 0.20%. For the typical case the simulation error is far less than 0.20%. To determine the simulation error, we conducted a simulation of the simulations. Unlike some Monte Carlo experiments (such as those estimating the size of a test statistic under the null) the standard error of the simulation error for most of our estimates (returns, prices, *etc.*) are themselves analytically intractable, and must be simulated. In order to estimate the standard error of the simulation error in estimating market prices, we estimated a single market price 2,000 times, each time independent of the other, and from this set of prices computed the mean and variance of the price estimate. If the experiment had no simulation error, each of the price estimates would be identical. With the number of possible paths, J, equal to no less than 1,000 we find that the standard deviation of the simulation error is less than 0.20% of the price, which is sufficiently small as not to be a source of concern for our study. The number of simulations has to be substantially greater than 1,000 for some cases depending on the model specification and the ex ante equity premium.

[24] For our random number generation we made use of a variance reduction technique, stratified sampling. This technique has us drawing pseudo-random numbers ensuring that $q\%$ of these draws come from the q^{th} percentile, so that our sampling does not weight any grouping of random draws too heavily.

are the simulated innovations to the interest rate and dividend growth rate processes, respectively. Form $log(1+g_t^{j,n}) = 0.049 + 0.64\epsilon_{g,t-1}^{j,n} + \epsilon_{g,t}^{j,n}$ and $log(r_{f,t}^{j,n}) = -0.35 + 0.88log(r_{f,t-1}^{j,n}) + \epsilon_{r,t}^{j,n}$.

Step 4B(ii): Produce two correlated normal random variables $\epsilon_{r,t+1}^{j,n}$ and $\epsilon_{g,t+1}^{j,n}$ as in Step 4B(i) above, and conditioning on $\epsilon_{g,t}^{j,n}$ and $log(r_{f,t}^{j,n})$ from Step 4B(i) produce $log(1+g_{t+1}^{j,n}) = 0.049 + 0.64\epsilon_{g,t}^{j,n} + \epsilon_{g,t+1}^{j,n}$ and $log(r_{f,t+1}^{j,n}) = -0.35 + 0.88log(r_{f,t}^{j,n}) + \epsilon_{r,t+1}^{j,n}$ for $j = 1, 2, 3, \cdots, J$.

Step 4B(iii): Repeat Step 4B(ii) to form $log(1+g_{t+i}^{j,n})$ and $log(r_{t+i}^{j,n})$ for $i = 2, 3, 4, \cdots, I$, $j = 1, 2, 3, \cdots, J$, and economies $n = 1, 2, 3, \cdots, N$.

Step 4B(iv): The discounted present value of each of the individual J streams of dividends is now taken in accordance with Equation (17), with the j^{th} present value price noted as $P_t^{j,n}$. Finally, the price for the n^{th} economy in period t is formed: $P_t^n = \frac{1}{J}\sum_{j=1}^{J} P_t^{j,n}$.

In considering these prices, note that according to Equation (17) the stream of discount rates and dividend growth rates should be infinitely long, while in our simulations we extend the stream for only a finite number of periods, I. Since the ratio of gross dividend growth rates to gross discount rates are less than unity in steady state, the individual product elements in the infinite sum in Equation (17) eventually converge to zero as I increases. (Indeed, this convergence to zero is exactly what is required for the standard transversality condition that the expected present value of the stock price P_{t+i} falls to zero as i goes to infinity.) We therefore set I large enough in our simulations so that the truncation does not materially effect our results. We find that setting $I = 1,000$ years is sufficient in all cases we studied. That is, the discounted present value of a dividend payment received 1,000 years in the future is essentially zero. Also note that the steps above are required to produce P_t^n, D_t^n, g_t^n, and r_t^n for $n = 1, \cdots, N$ and $t = 1, \cdots, T$; the intermediate terms superscripted with a j are required only to perform the numerical integration that yields P_t^N. Note that the length of the time series T is chosen to be 53 to imitate the 53 years of annual data we have available for the S&P 500 from 1952 to 2004.

Step 4C: After performing Steps 4A(i)-(iii) and 4B(i)-(iv) for $t = 1, \cdots, T$, rolling out N independent economies for T periods, we construct the market returns for each economy, $R_t^n = (P_{t+1}^n + D_{t+1}^n - P_t^n)/P_t^n$, and the ex post equity premium that agents in the n^{th} economy would observe, $\hat{\pi}^n$, estimated from Equation (1) as the mean difference in market returns and the risk-free rate.

Table I

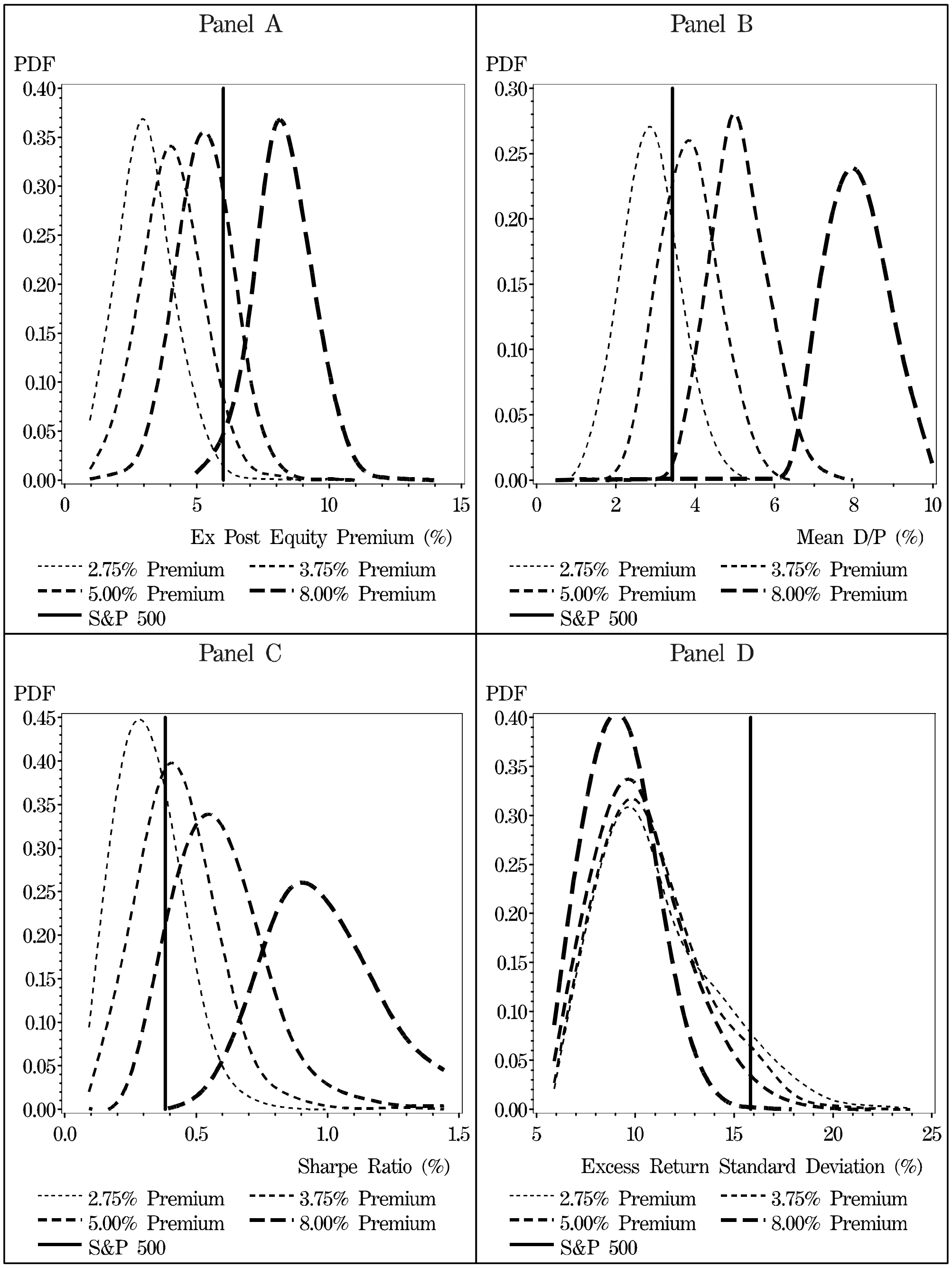
Characteristics of Simulated Models

Here we present the 12 models we consider, identifying the characteristics of their underlying data generating processes. The column titled "Processes for r, g, & π" indicates the nature of the time-series models used to generate the interest rates, dividend growth rates, and equity premium. See Appendix 1 for details on how this set of models was chosen and a description of how the equity premium series is produced. The column titled "Downward Trend in Equity Premium Process," identifies whether the ex ante equity premium trends downward over the course of the 53-year experiment, and if it does, provides the amount of the downward trend. The next column, "Structural Break in Equity Premium Process," indicates whether the model incorporates a sudden 50 basis point (bps) drop in the value of the ex ante equity premium. The column "Structural Break in Dividend Growth Process," indicates whether the model incorporates a gradual 100 basis point increase in the growth rate of the dividend growth rate. The final column indicates that all the models except Models 11 and 12 incorporate sampling variability in generating parameters. Additional model details are as follows. Parsimonious Model: interest rates follow an AR(1), dividend growth rates follow a MA(1), the equity premium follows an AR(1). Deterministic π Model: interest rates follow an AR(1), dividend growth rates follow a MA(1), the equity premium follows a deterministic downward trend with a 50 bps structural break. Best BIC Model:† interest rates follow an ARMA(1,1), dividend growth rates follow a MA(1), the equity premium follows an AR(1). Second-Best BIC Model:† interest rates follow an ARMA(1,1), dividend growth rates follow a MA(1), the equity premium follows an ARMA(1,1). Further details about each model feature are provided in Appendix 1.

Model	Processes for r, g, & π	Downward Trend in Equity Premium Process	Structural Break in Equity Premium Process	Structural Break in Dividend Growth Process	Sampling Variability in Generating Parameters
1	Parsimonious Model	No	No	No	Yes
2	Parsimonious Model with π Trend	Yes (80 bps)	No	No	Yes
3	Parsimonious Model with π Break	No	Yes (50 bps)	No	Yes
4	Parsimonious Model with Dividend Growth Trend	No	No	Yes	Yes
5	Parsimonious Model with π Trend and Dividend Growth Trend	Yes (80 bps)	No	Yes	Yes
6	Parsimonious Model with π Break, π Trend, and Dividend Growth Trend	Yes (30 bps)	Yes (50 bps)	Yes	Yes
7	Best BIC Model† with π Break, π Trend, and Dividend Growth Trend	Yes (30 bps)	Yes (50 bps)	Yes	Yes
8	Second-Best BIC Model† with π Break, π Trend, and Dividend Growth Trend	Yes (30 bps)	Yes (50 bps)	Yes	Yes
9	Parsimonious Model with π Break and π Trend	Yes (30 bps)	Yes (50 bps)	No	Yes
10	Deterministic π Model with π Break and π Trend	Yes (30 bps)	Yes (50 bps)	No	Yes
11	Parsimonious Model with Constant Parameters π Break, π Trend, and Dividend Growth Trend	Yes (30 bps)	Yes (50 bps)	Yes	No
12	Parsimonious Model with Constant Parameters	No	No	No	No

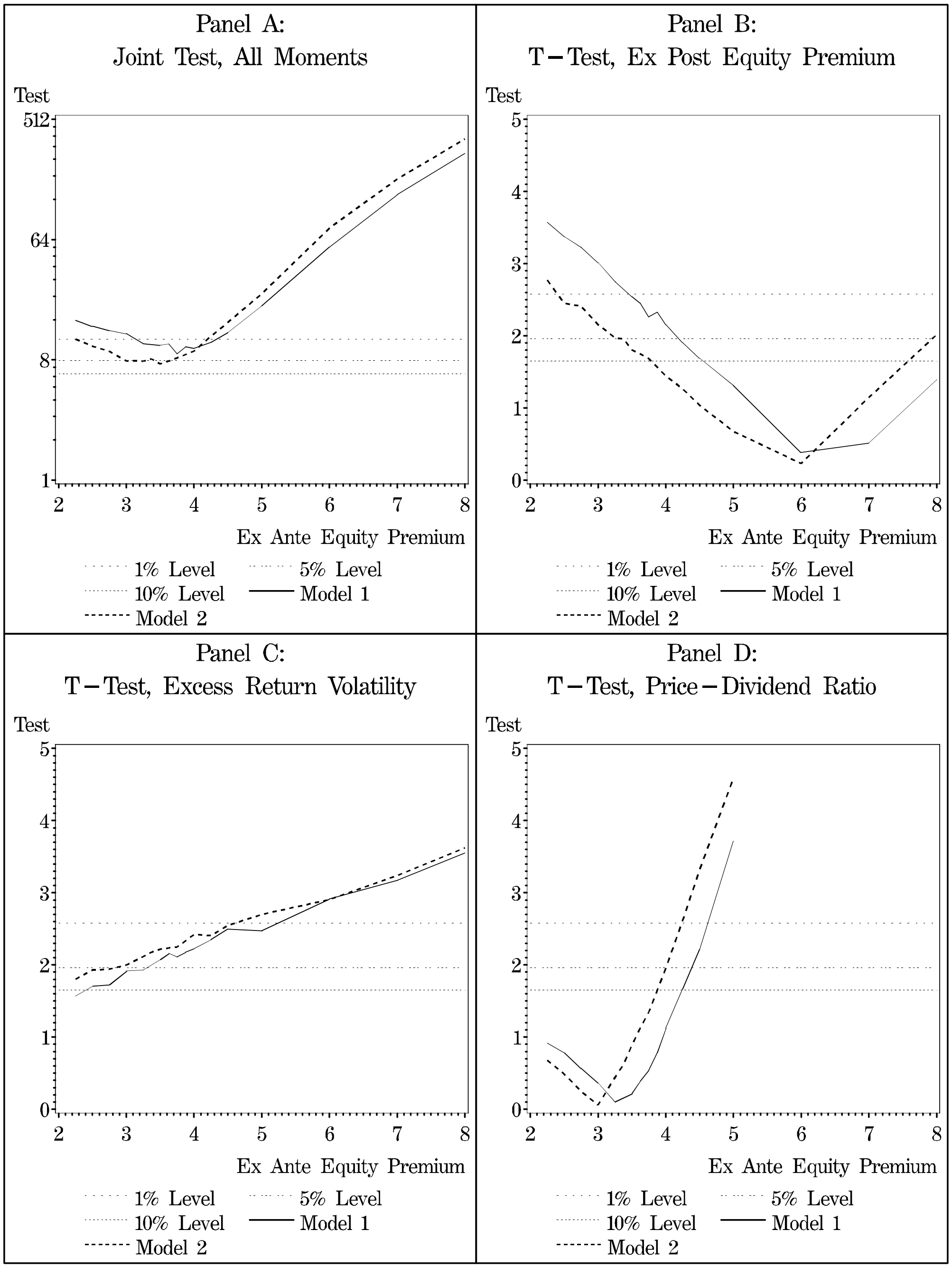
† For Models 7 and 8 we employ the Bayesian Information Criterion (BIC) to select the order of the ARMA model driving each of the interest rate, equity premium, and dividend growth rate processes. The order of each AR process and each MA process for each series is chosen over a (0, 1, 2) grid.

Figure 1: Probability Distribution Functions of Simulated Ex Post Equity Premia, Dividend Yields, Sharpe Ratios, and Return Standard Deviations



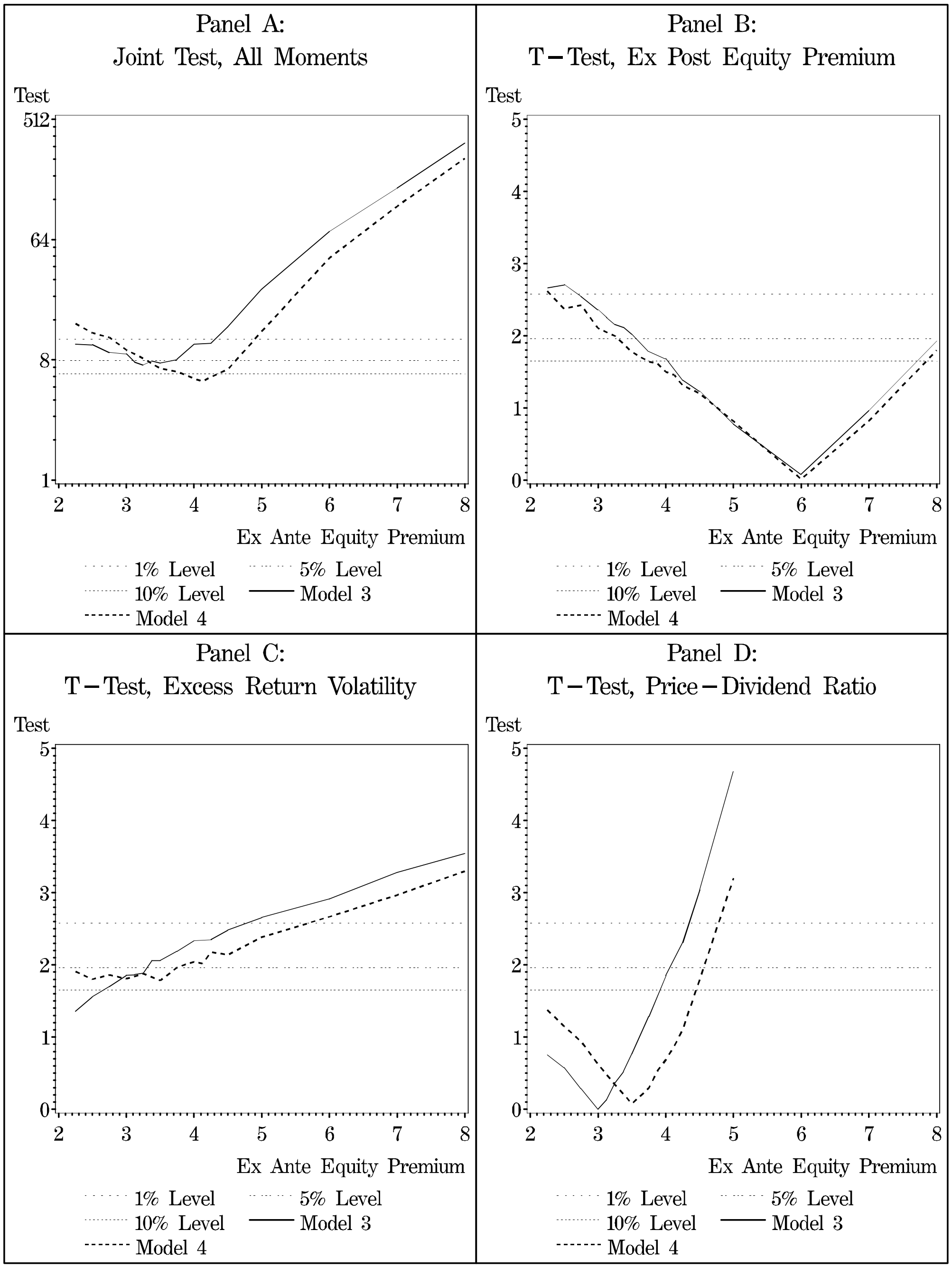
This figure contains probability distribution functions (PDFs) for various financial statistics generated in 2,000 simulated economies based on Model 1 from Table I. Each panel contains a PDF for each of four different assumed values of the ex ante equity premium: 2.75%, 3.75%, 5%, and 8%. Panel A shows the distribution of the ex post equity premium (mean return minus mean interest rate), Panel B shows the mean dividend yield distribution (dividend divided by price), Panel C shows the Sharpe ratio distribution (excess return divided by the standard deviation of the excess return), and Panel D shows the distribution of the standard deviation of excess returns. In each panel, a vertical line indicates the US data realized over 1952-2004, the value of the estimated ex post equity premium, mean dividend yield, mean Sharpe ratio, and excess return standard deviation, respectively. The simulated statistics are estimated on 53 years of generated data for each economy, mimicking the data period we used to estimate the actual US results.

Figure 2: Joint and Individual Tests Statistics for Models 1 and 2



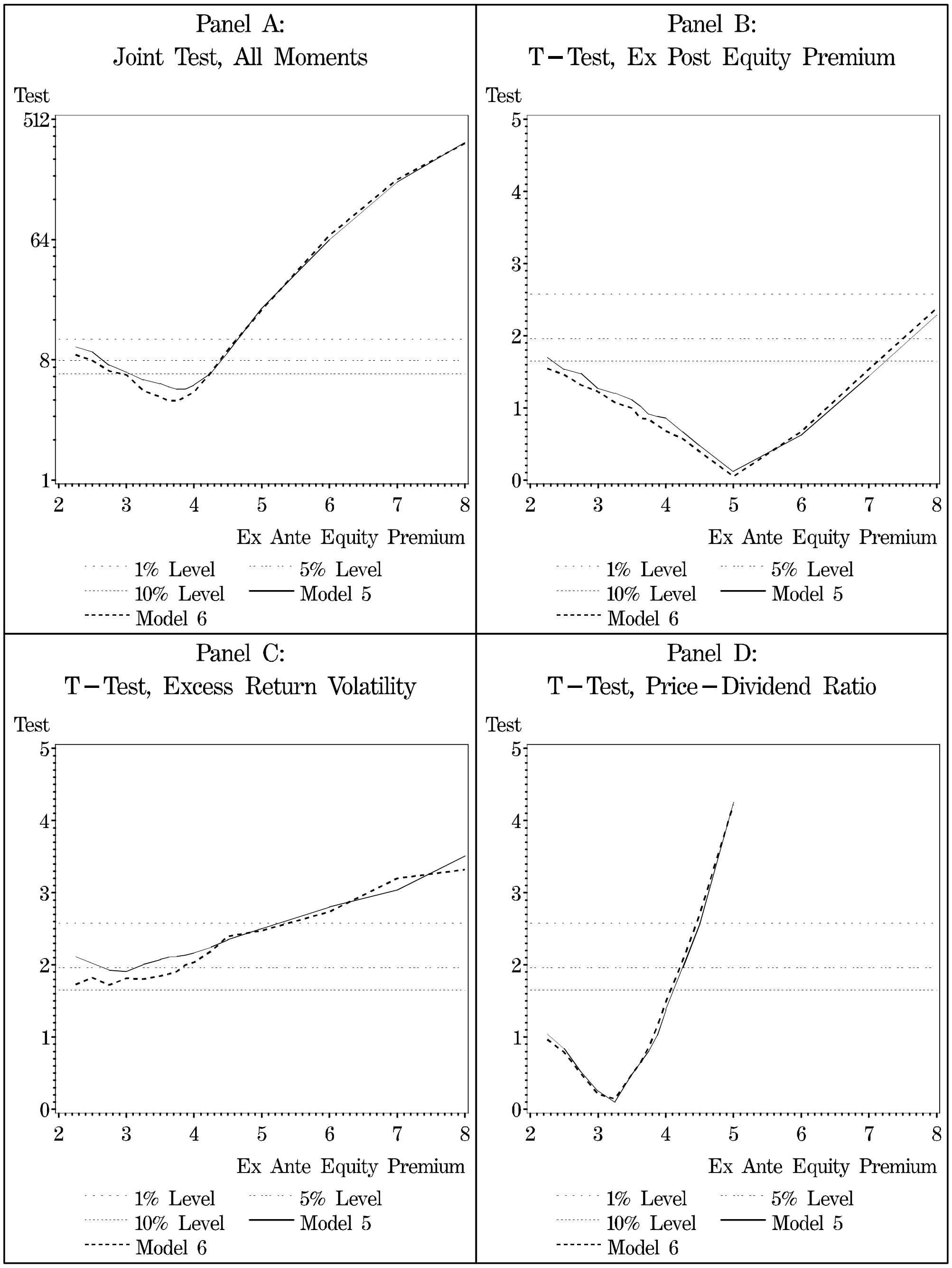
This figure contains plots of test statistics for Models 1 and 2. Panel A plots joint χ^2 tests based on a set of three variables (the ex post equity premium, the mean dividend yield, and the excess return volatility) for various ending values of the ex ante equity premium for each model. In Panel A the vertical axis is plotted on a log scale. The remaining panels contains t-test values corresponding to tests on the individual variables for each of the models: the ex post equity premium in Panel B, the excess return volatility in Panel C, and price-dividend ratio in Panel D. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Figure 3: Joint and Individual Tests Statistics for Models 3 and 4



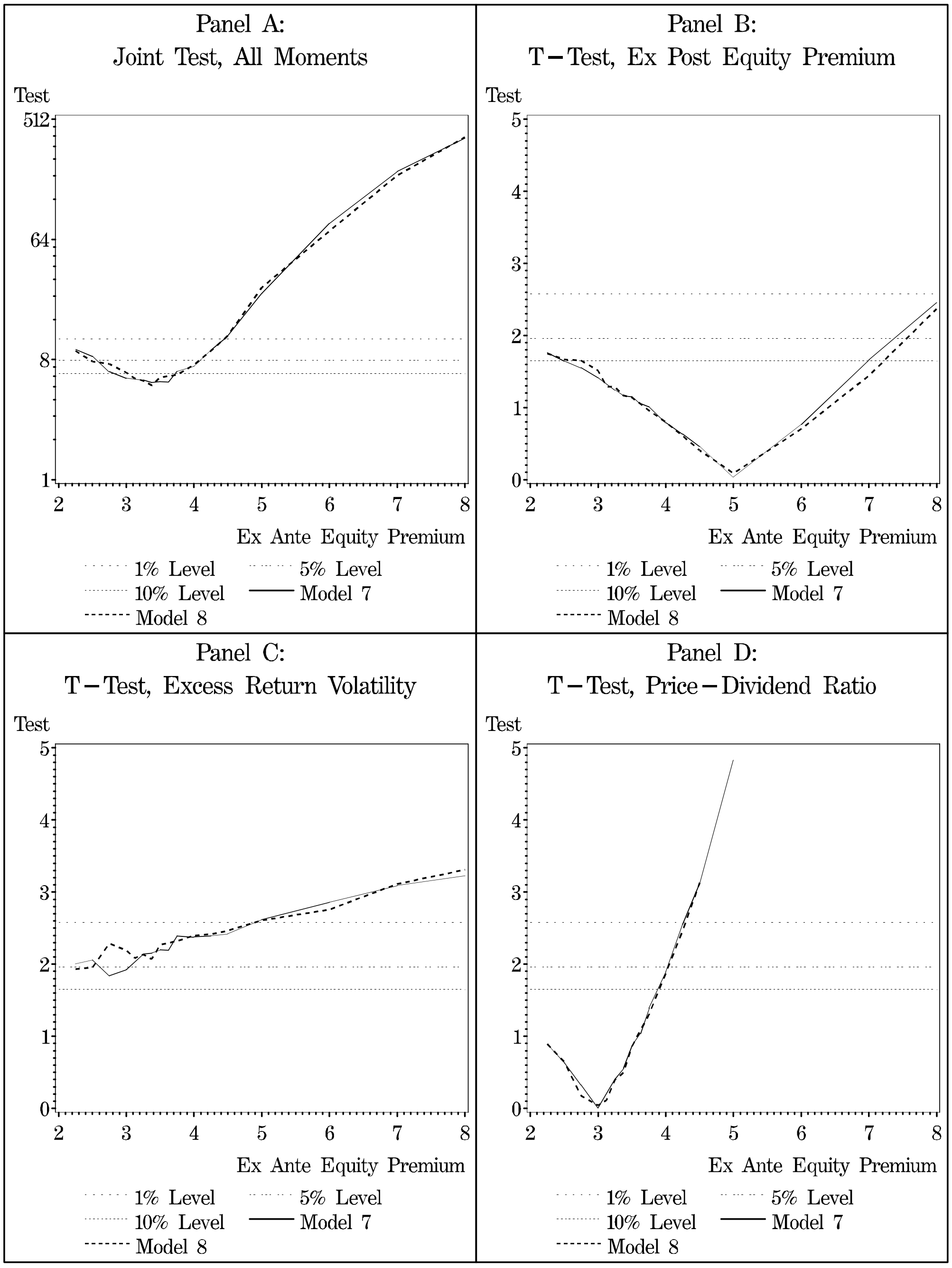
This figure contains plots of test statistics for Models 3 and 4. Panel A plots joint χ^2 tests based on a set of three variables (the ex post equity premium, the mean dividend yield, and the excess return volatility) for various ending values of the ex ante equity premium for each model. In Panel A the vertical axis is plotted on a log scale. The remaining panels contains t-test values corresponding to tests on the individual variables for each of the models: the ex post equity premium in Panel B, the excess return volatility in Panel C, and price-dividend ratio in Panel D. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Figure 4: Joint and Individual Tests Statistics for Models 5 and 6



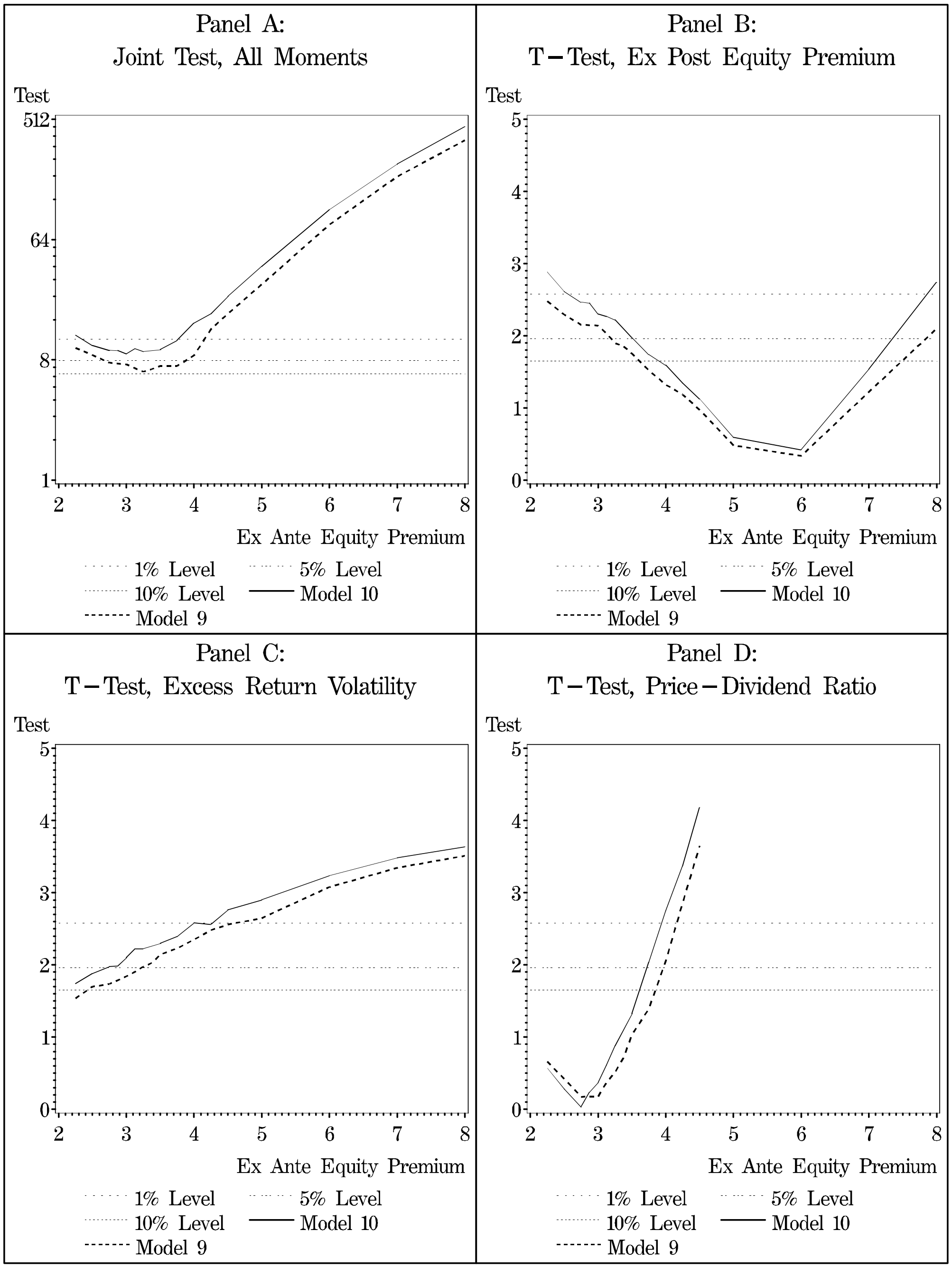
This figure contains plots of test statistics for Models 5 and 6. Panel A plots joint χ^2 tests based on a set of three variables (the ex post equity premium, the mean dividend yield, and the excess return volatility) for various ending values of the ex ante equity premium for each model. In Panel A the vertical axis is plotted on a log scale. The remaining panels contains t-test values corresponding to tests on the individual variables for each of the models: the ex post equity premium in Panel B, the excess return volatility in Panel C, and price-dividend ratio in Panel D. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Figure 5: Joint and Individual Tests Statistics for Models 7 and 8



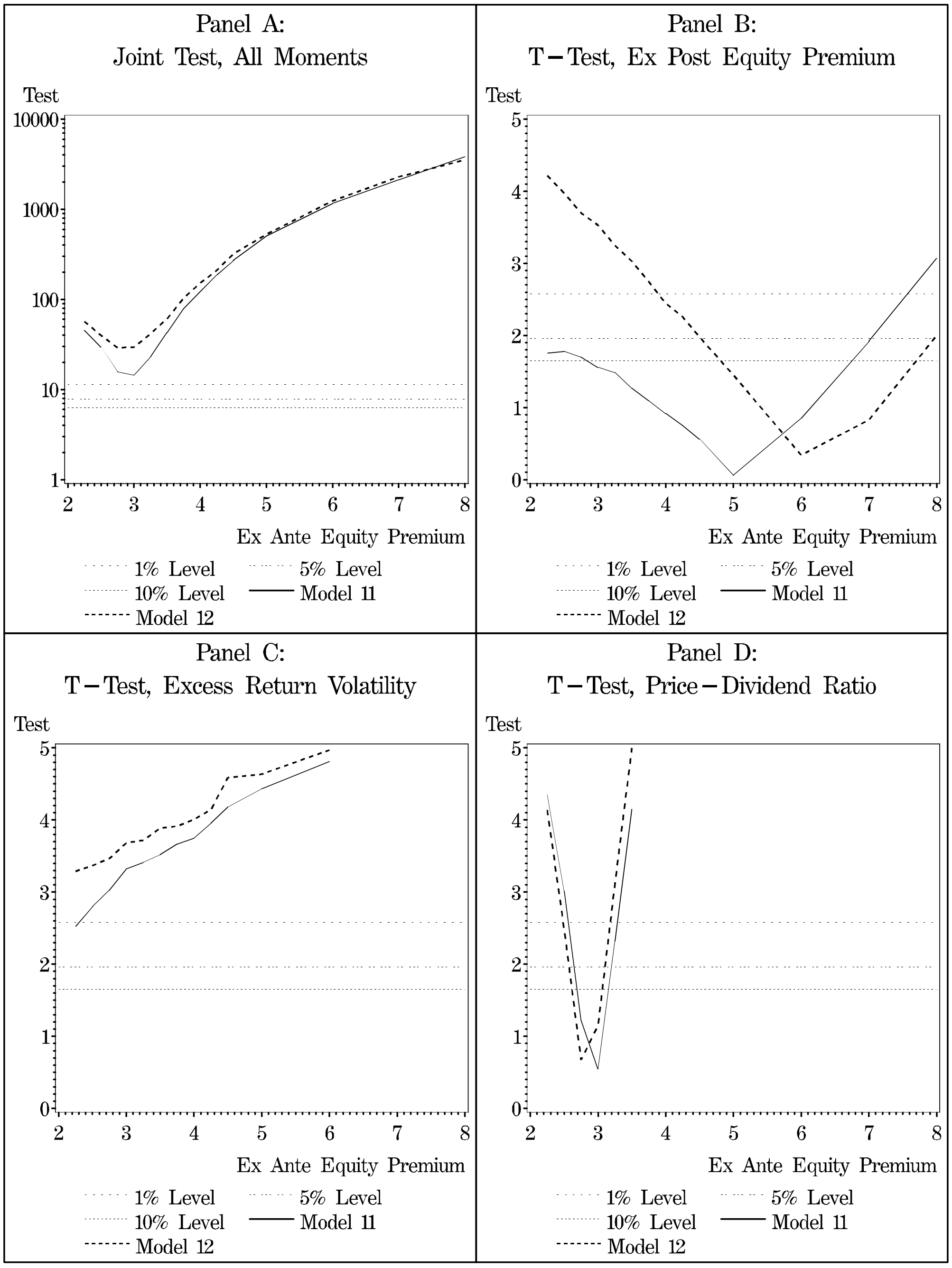
This figure contains plots of test statistics for Models 7 and 8. Panel A plots joint χ^2 tests based on a set of three variables (the ex post equity premium, the mean dividend yield, and the excess return volatility) for various ending values of the ex ante equity premium for each model. In Panel A the vertical axis is plotted on a log scale. The remaining panels contains t-test values corresponding to tests on the individual variables for each of the models: the ex post equity premium in Panel B, the excess return volatility in Panel C, and price-dividend ratio in Panel D. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Figure 6: Joint and Individual Tests Statistics for Models 9 and 10



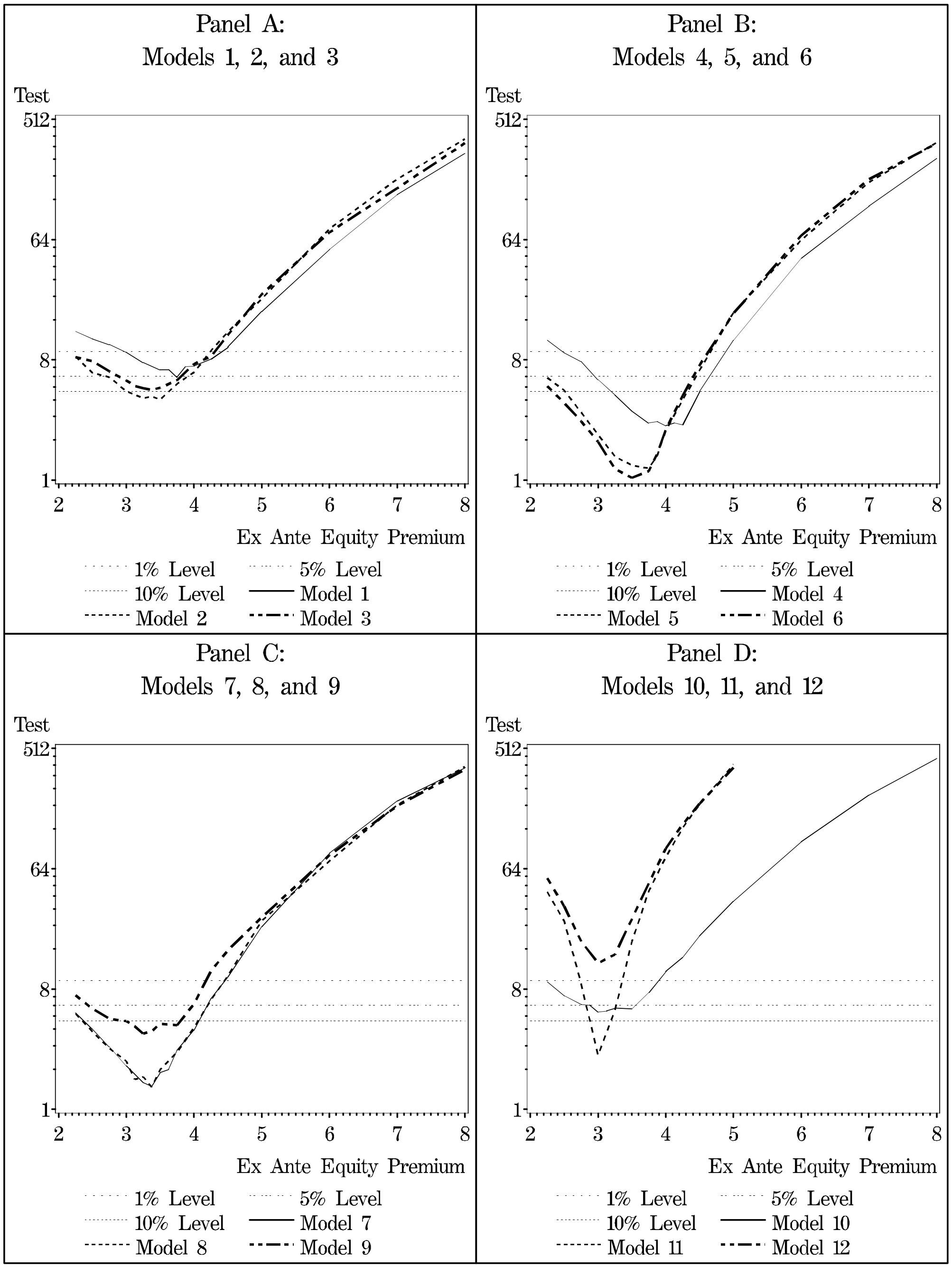
This figure contains plots of test statistics for Models 9 and 10. Panel A plots joint χ^2 tests based on a set of three variables (the ex post equity premium, the mean dividend yield, and the excess return volatility) for various ending values of the ex ante equity premium for each model. In Panel A the vertical axis is plotted on a log scale. The remaining panels contains t-test values corresponding to tests on the individual variables for each of the models: the ex post equity premium in Panel B, the excess return volatility in Panel C, and price-dividend ratio in Panel D. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Figure 7: Parameter Estimation Certainty:
Joint and Individual Tests Statistics for Models 11 and 12



This figure contains plots of test statistics for Models 11 and 12. Panel A plots joint χ^2 tests based on a set of three variables (the ex post equity premium, the mean dividend yield, and the excess return volatility) for various ending values of the ex ante equity premium for each model. In Panel A the vertical axis is plotted on a log scale. The remaining panels contains t-test values corresponding to tests on the individual variables for each of the models: the ex post equity premium in Panel B, the excess return volatility in Panel C, and price-dividend ratio in Panel D. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Figure 8: Investors' Model Uncertainty
Joint Tests Based on a Subset of Moments for Models 1-12



This figure contains plots of joint χ^2 tests based on a set of two variables, the ex post equity premium and the mean dividend yield, for various ending values of the ex ante equity premium for each model. Panel A presents the test statistics for Models 1, 2, and 3, Panel B presents the test statistics for Models 4, 5, and 6, Panel C presents the test statistics for Models 7, 8, and 9, and Panel D presents the test statistics for Models 10, 11, and 12. The vertical axis of each plot is on a log scale. In each panel the critical values of the test statistics corresponding to test significance at the 10%, 5%, and 1% levels are indicated by horizontal lines.

Are Stocks Cheap? A Review of the Evidence

Fernando Duarte and Carlo Rosa

We surveyed banks, we combed the academic literature, we asked economists at central banks. It turns out that most of their models predict that we will enjoy historically high excess returns for the S&P 500 for the next five years. But how do they reach this conclusion? Why is it that the equity premium is so high? And more importantly: Can we trust their models?

The equity risk premium is the *expected* future return of stocks minus the risk-free rate over some investment horizon. Because we don't directly observe market expectations of future returns, we need a way to figure them out indirectly. That's where the models come in. In this post, we analyze twenty-nine of the most popular and widely used models to compute the equity risk premium over the last fifty years. They include surveys, dividend-discount models, cross-sectional regressions, and time-series regressions, which together use more than thirty different variables as predictors, ranging from price-dividend ratios to inflation. Our calculations rely on real-time information to avoid any look-ahead bias. So, to compute the equity risk premium in, say, January 1970, we only use data that was available in December 1969.

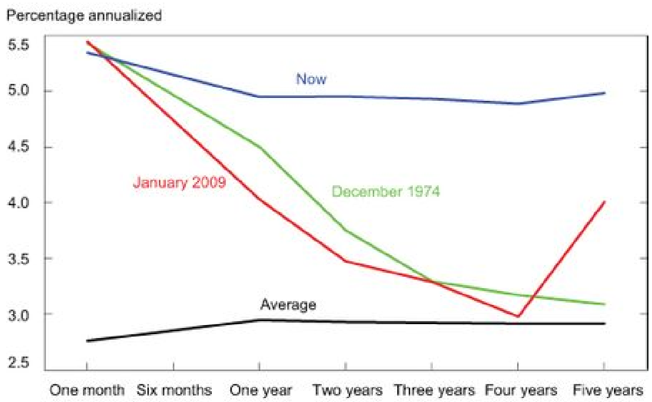
Let's now take a look at the facts. The chart below shows the weighted average of the twenty-nine models for the one-month-ahead equity risk premium, with the weights selected so that this single measure explains as much of the variability across models as possible (for the geeks: it is the first principal component). The value of 5.4 percent for December 2012 is about as high as it's ever been. The previous two peaks correspond to November 1974 and January 2009. Those were dicey times. By the end of 1974, we had just experienced the collapse of the Bretton Woods system and had a terrible case of stagflation. January 2009 is fresher in our memory. Following the collapse of Lehman Brothers and the upheaval in financial markets, the economy had just shed almost 600,000 jobs in one month and was in its deepest recession since the 1930s. It is difficult to argue that we're living in rosy times, but we are surely in better shape now than then.

Today's equity premium has reached a historic high

Sources: Authors' calculations; Barclays; Deutsche Bank; Duke/CFO Business Outlook survey; Federal Reserve Board; Federal Reserve Bank of New York; Goldman Sachs; J.P. Morgan; Nomura; the Center for Research in Security Prices; Federal Reserve Economic Data; Thomson Reuters; the websites of NYU's Aswath Damodaran; Dartmouth's Kenneth French, University of Lausanne's Amit Goyal, University of California at Berkeley's Martin Lettau, Yale's Robert Shiller.

The next chart shows a comparison between those two episodes and today. For 1974 and 2009, the green and red lines show that the equity risk premium was high at the one-month horizon, but was decreasing at longer and longer horizons. Market expectations were that at a four-year horizon the equity risk premium would return to its usual level (the black line displays the average levels over the last fifty years). In contrast, the blue line shows that the equity risk premium today is high irrespective of investment horizon.

The equity premium is elevated at all horizons

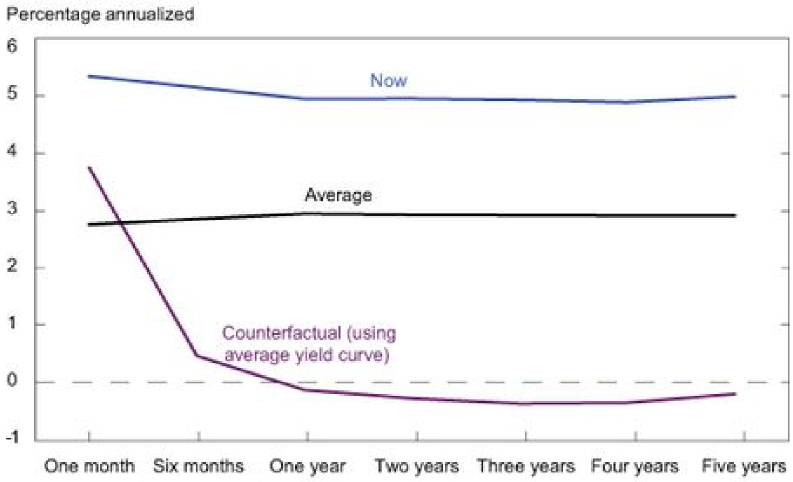


Sources: Authors' calculations; Barclays; Deutsche Bank; Duke/CFO Business Outlook survey; Federal Reserve Board; Federal Reserve Bank of New York; Goldman Sachs; J.P. Morgan; Nomura; the Center for Research in Security Prices; Federal Reserve Economic Data; Thomson Reuters; the websites of NYU's Aswath Damodaran; Dartmouth's Kenneth French, University of Lausanne's Amit Goyal, University of California at Berkeley's Martin Lettau, Yale's Robert Shiller.

Why is the equity premium so high right now? And why is it high at all horizons? There are two possible reasons: low discount rates (that is, low Treasury yields) and/or high current or future expected dividends. We can figure out which factor is more important by comparing the twenty-nine models with one another. This strategy works because some models emphasize changes in dividends, while others emphasize changes in risk-free rates. We find that the equity risk premium is high mainly due to exceptionally low Treasury yields at all foreseeable horizons. In contrast, the current level of dividends is roughly at its historical average and future dividends are expected to grow only modestly above average in the coming years.

In the next chart we show, in an admittedly crude way, the impact that low Treasury yields have on the equity risk premium. The blue and black lines reproduce the lines from the previous chart: the blue is today's equity risk premium at different horizons and the black is the average over the last fifty years. The new purple line is a counterfactual: it shows what the equity premium would be today if nominal Treasury yields were at their average historical levels instead of their current low levels. The figure makes clear that exceptionally low yields are more than enough to justify a risk premium that is highly elevated by historical standards.

The equity premium is high because Treasury yields are low

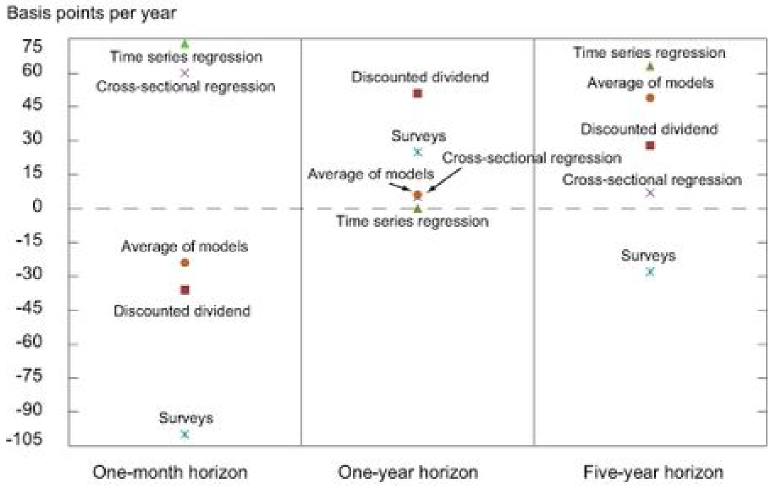


Sources: Authors' calculations; Barclays; Deutsche Bank; Duke/CFO Business Outlook survey; Federal Reserve Board; Federal Reserve Bank of New York; Goldman Sachs; J.P. Morgan; Nomura; the Center for Research in Security Prices; Federal Reserve Economic Data; Thomson Reuters; the websites of NYU's Aswath Damodaran; Dartmouth's Kenneth French, University of Lausanne's Amit Goyal, University of California at Berkeley's Martin Lettau, Yale's Robert Shiller.

But none of this analysis matters if excess returns are unpredictable because the equity risk premium is all about expected returns. So…*are* returns predictable? The jury is still out on this one, and the debate among academics and practitioners is alive and well. The simplest predictive method is to assume that future returns will be equal to the average of all past returns. It turns out that it is remarkably tricky to improve upon this simple method. However, with so many models at hand, we couldn't help but ask if any of them can, in fact, do better.

The table below gives the extra returns that investors could have earned by using the models instead of the historical mean to predict future returns. For investment horizons of one month, one year, and five years, we pick the best model in each of the four classes we consider together with the weighted average of all twenty-nine models. We compute these numbers by assuming that investors can allocate their wealth in stocks or bonds, and that they are not too risk-averse (for the geeks again, we solved a Merton portfolio problem in real time assuming that the coefficient of relative risk aversion is equal to one). The table shows positive extra returns for most of the models, especially at long horizons.

Model performance is varied, but better at longer horizons



Sources: Authors' calculations; Barclays; Deutsche Bank; Duke/CFO Business Outlook survey; Federal Reserve Board; Federal Reserve Bank of New York; Goldman Sachs; J.P. Morgan; Nomura; the Center for Research in Security Prices; Federal Reserve Economic Data; Thomson Reuters; the websites of NYU's Aswath Damodaran; Dartmouth's Kenneth French, University of Lausanne's Amit Goyal, University of California at Berkeley's Martin Lettau, Yale's Robert Shiller.

Notes: We tested twenty-nine models in four classes (surveys, dividend-discount models, cross-sectional regressions, and time series regressions) over three investment horizons. In this chart, we plot the single best-performing model in each category. We also show how the optimal weighted average (the first principal component) of all models performs.

At face value, this result means that the models are actually helpful in forecasting returns. However, we should keep in mind some of the limitations of our analysis. First, we have not shown confidence intervals or error bars. In practice, those are quite large, so even if we could have earned extra returns by using the models, it may have been solely due to luck. Second, we have selected models that have performed well in the past, so there is some selection bias. And of course, past performance is no guarantee of future performance.

Disclaimer

The views expressed in this post are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

The Equity Risk Premium: A Consensus of Models

By Fernando Duarte and Carlo Rosa

Abstract

We estimate the equity risk premium by combining information from twenty models. Our main finding is that there is broad agreement across models that the equity premium reached historical heights in July 2013 even when the models are substantially different from each other and use more than one hundred different economic variables. Our preferred estimator places the one-year-ahead equity premium in July 2013 at 14.5 percent, the highest level in fifty years and well above the 10.5 percent that was reached during the financial crisis in 2009. The models also show broad agreement that the term structure of equity risk premia is high and flat: expected excess returns at all foreseeable horizons are just as high as at the one-year horizon. A high equity premium that is not expected to mean-revert in the near future is an unprecedented phenomenon. Because expected dividend growth has not been above average in 2013, we conclude the high equity premium is mostly due to unusually low discount rates at all horizons.

1. Introduction

The equity risk premium —the expected return of stocks in excess of the risk-free rate— is a fundamental quantity in all of asset pricing, both for theoretical and practical reasons. It is a key measure of aggregate risk-aversion and an important determinant of the cost of capital for corporations, saving decisions of individuals and budgeting plans for governments. Recently, the equity risk premium (ERP) has also returned to the forefront of policymaking as a leading indicator of the evolution of the economy, a potential explanation for the jobless recovery and a gauge of financial stability. As an indicator of future activity, a high ERP at short horizons tends to be followed by higher GDP growth, higher inflation and lower unemployment, thus informing both fiscal and monetary decisions. Bloom (2009) and new research by Duarte, Kogan and Livdan (2013) point to large effects of the ERP on real aggregate investment, a component that has been lagging in the present recovery compared to policymakers' forecasts in the current cycle and actual performance in past cycles. As a potential explanation of the jobless recovery, Hall (2013) and Kuehn, Petrosky-Nadeau and Zhang (2012) have proposed that increased risk-aversion has prevented firms from hiring as much as would be expected in today's macroeconomic environment. From the perspective of financial stability, the so-called "great rotation" from bonds to stocks could be exacerbated in speed and magnitude if the ERP is persistently high. A sudden flow of money out of the bond market into stocks could spell large capital losses for fixed income investors, including the Federal Reserve. Low returns in other asset classes could provide incentives for investors to engage in potentially unsafe "reach for yield" either through excessive use of leverage or through other forms of risk-taking. The ERP is also important from the perspective of unconventional monetary policy: a high ERP may make the portfolio channel of Large Scale Asset Purchases more effective because it further increases the demand for risky assets.

In this article, we estimate the ERP by combining information from twenty models that are prominently used by practitioners and featured in the academic literature. Our main finding is that there is broad agreement across models that the ERP has reached historical heights even when the models are substantially different from each other and use more than one hundred different economic variables. Our preferred estimator places the one-year-ahead ERP in July 2013 at 14.5 percent, the highest level in fifty years and well above the 10.5 percent that was reached during the financial crisis in 2009. The models also show broad agreement that expected excess returns at all foreseeable horizons are just as high as at the one-year horizon. A high equity premium that is not expected to mean-revert in the near future is an unprecedented phenomenon.

In addition to estimating the level of the ERP, it is useful for policymakers and other economic agents to know why the ERP is high. We conclude the ERP is high at all foreseeable horizons because Treasury yields are unusually low at all maturities. In other words, the term structure of equity premia is high and flat because the term structure of interest rates is low and flat. Current and expected future dividend and earnings growth play only a minor role. A high ERP caused by low bond yields indicates that a stock market correction is likely to occur only when bond yields start to rise. Additionally, a bond-driven ERP makes it more unlikely that irrational exuberance can take hold in equity markets, especially at times of increasing expectations for a steepening of the yield curve. Another implication of a bond-driven ERP is that we should no longer rely on traditional indicators of the ERP like the price-dividend or price-earnings ratios, which all but ignore the term structure of risk-free rates.

As a second contribution, we evaluate the performance of different ERP models. Statements about the implications of a high ERP are valid only to the extent that expected returns predict future realized returns. For the models we consider, predictability is weak but present. We first categorize the twenty models we study into five groups: predictors that use historical mean returns, dividend-discount models, cross-sectional regressions, time-series regressions and surveys. To assess whether models can indeed predict returns, we regress realized excess returns on the corresponding ERP given by the models. We then use the out-of-sample R-squared for these regressions as a measure of success. We find that dividend-discount models perform best at short horizons, while cross-sectional regressions perform best at longer horizons. Combining all models into a single principal component — our preferred measure— reduces noise. A mean-variance investor with unit risk aversion using the principal component as an investment signal would have earned 15 percentage points more over the last fifty years (30 basis points per year) than if she had assumed expected returns are equal to past mean returns.

2. The Equity Risk Premium: Definition

Conceptually, the ERP is the compensation investors require to make them indifferent between holding the risky market portfolio and a risk-free bond. Because this compensation depends on the future performance of stocks, the ERP incorporates expectations of future stock market returns, which are not directly observable. At the end of the day, any model of the ERP is a model of investor expectations. Additionally, it is not clear what truly constitutes the market return and the risk-free rate in the real world. In practice, the most common measures of market returns are given by broad stock market indices, like the S&P 500 or the Dow Jones Industrial Average, but those indices do not include the whole universe of traded stocks and miss several other components of wealth. Even if we included all traded stocks, we still have several choices to make, such as whether to use value or equal-weighted indices, or whether to exclude penny stocks or rarely traded stocks. A similar problem arises with the risk-free rate. While we almost always use Treasury yields as measures of risk-free rates, they are not completely riskless since nominal Treasuries are exposed to inflation and liquidity risks. In this paper, we follow common practice and always use the S&P 500 as a measure of stock market prices and either nominal or real Treasury yields as risk-free rates. The models we consider differ only in how expectations are computed.

While implementing the concept of the ERP has pitfalls, we can precisely define the ERP mathematically. First, we decompose stock returns into an expected component and an unpredictable random component:

$$R_{t+k} = E_t[R_{t+k}] + error_{t+k} \tag{1}$$

In equation (1), R_{t+k} are net *realized* returns between t and $t+k$, $E_t[R_{t+k}]$ are the returns that were expected from t to $t+k$ using information available at time t and $error_{t+k}$ is a mean-zero random variable that is unknown at time t but is realized at $t+k$. The ERP at time t for horizon k is defined as

$$ERP_t(k) = E_t[R_{t+k}] - R^f_{t+k} \tag{2}$$

where R^f_{t+k} is the net risk-free rate for investing from t to $t+k$ (which, being risk-free, is known at time t).

This definition shows three important aspects of the ERP. First, because the unexpected component $error_{t+k}$ is stochastic and orthogonal to expected returns, the ERP is always less volatile than realized excess returns. Therefore, while realized stock returns are very volatile compared to bonds, we expect good ERP estimates to be somewhat smoother. Second, the ERP itself is a random variable, since expectations can change through time when new information arrives. Third, the ERP has an investment horizon k embedded in it, since we can consider expected excess returns over, say, one month, one year or five years from today. If we fix t, and let k vary, we trace the *term structure* of the equity risk premium.

3. Data

In constructing all estimates of the ERP we use over one hundred variables. The sources and definitions are standard. The nominal and real price, earnings and dividends for the S&P 500 are from Shiller. Inflation, the "cyclically adjusted price-earnings ratio" and the ten-year nominal treasury yield are also from Shiller. Expected earnings per share are mean analyst forecasts from Thomson Reuters I/B/E/S. Nominal bond yields for all maturities except 10-years[1] and all TIPS yields are from the Federal Reserve Board. Fama-French and momentum factors and portfolios are from Professor French's website. Corporate bond spreads and the NBER recession indicator are from the St. Louis Federal Reserve (FRED). Book value per share for the S&P 500 is from Compustat. Debt issuance and equity issuance are from Jeffrey Wurgler's website. Consumption to wealth ratio measured by *cay* is from Martin Lettau's website (Ludvigson and Lettau, 2001). ERP estimates from CFOs are from the Duke CFO survey. The sentiment measure of Baker and Wurgler is from Jeffrey Wurgler's website. Professor Damodaran's estimates of the ERP are from his website. All variables are monthly from January 1960 to July 2013, except for *cay* and CFO surveys, which are quarterly, and book value per share and Damodaran's ERP estimates, which are annual. Other variables are constructed using the variables mentioned before. A detailed description is in Appendix A.

4. Models of the Equity Risk Premium

We classify models of the ERP into five categories and discuss their advantages and disadvantages. We also describe in detail the models we use within each category and how to obtain a term-structure of the ERP for each one. Of course, there are many more models of the ERP than the ones we consider. We selected which models to include in our study based on the recent academic literature and widespread use by practitioners. All models are constructed in real time[2], so that an investor who lived through the

[1] Except for the 10-year yield, which, as described above, is from Shiller. We use Shiller's 10-year yield for ease of comparability with the existing literature. Results are virtually unchanged if we use all yields, including the 10-year yield, from the Federal Reserve Board.

[2] The one exception is Adrian, Crump and Moench's (2013) cross-sectional model, which is constructed using full-sample regression estimates. Our out-of-sample predictability results are essentially unchanged if we omit this model from the analysis.

sample would have been able to construct the measures at each point in time using available information only. This helps avoid look-ahead bias and makes the out-of-sample evaluation of the models meaningful.

4.1 Historical mean

The easiest approach to estimating the ERP is to assume it is equal to the historical mean of realized market returns in excess of the contemporaneous risk-free rate. The main choice is how far into the past to go when computing the historical mean. This model is very simple and, as we show in Section 8, quite difficult to improve upon when considering out-of-sample performance measures. The main drawbacks are that it is purely backward looking, and assumes that the future will behave like the past, i.e. it assumes the conditional mean of excess returns is not time-varying, giving very little time-variation in the ERP.

To trace the term structure of the ERP using the historical mean method, we simply use returns computed over different horizons and the corresponding maturity risk-free rate before taking the mean.

Model 1: We compute the historical mean going as far back into the past as the data allows.

Model 2: Same as Model 1 but we compute the mean using the previous 5-years of data only (i.e. we use a backward looking 5-year rolling window).

4.2 Dividend discount models (DDM)

All DDM start with the basic intuition that the value of a stock is determined by no more and no less than the cash flows it produces for its shareholders (Gordon 1962). Today's stock price should then be the sum of all expected future dividends, discounted at an appropriate rate to take into account their riskiness and the time value of money. The formula that reflects this intuition is

$$P_t = E_t \sum_{k=0}^{\infty} \frac{D_{t+k}}{\rho_{t+k}} \tag{3}$$

where E_t is the conditional expectations operator, P_t is the current price of the stock, D_t is the current level of dividends, D_{t+k} is the level of dividends k periods from now, and ρ_{t+k} is the discount rate for time $t+k$. The discount rate can be decomposed into

$$\rho_{t+k} = 1 + R^f_{t+k} + ERP_t(k) \tag{4}$$

When using a DDM, we refer to $ERP_t(k)$ as the *implied* ERP, since we plug in observed or estimated values for the price, dividends and the risk-free rates, and figure out what value of $ERP_t(k)$ makes the right-hand side equal to the left-hand side in equation (3). In this framework, the risk-free rate captures the discounting associated with the time value of money and the ERP captures the discounting associated with the riskiness of the dividends.

DDM are forward looking and are consistent with no arbitrage. In fact, equation (3) is an equilibrium condition that must hold in any bubble-free economy with no arbitrage. Another advantage of DDM is that they are easy to implement. A drawback of DDM is that the results are sensitive to how we measure expectations of future dividends. In addition, ignoring the bubble term, i.e. assuming that

$$\lim_{k \to \infty} E_t \left[\frac{D_{t+k}}{\rho_{t+k}} \right] = 0$$

may impute a higher ERP whenever a bubble is present but not considered in the model.

Even though DDM do not require the term structure of the ERP to be flat, in practice all DDM assume that $ERP_t(k) = ERP_t(j)$ for all k and j. With a single ERP measure for all horizons, equation (3) pins down the ERP completely, while if we had different ERP estimates for different horizons, equation (3) would become a single equation in several unknowns and the ERP would not be identified.

Model 3: The simplest DDM assumes a constant growth rate of dividends and a flat yield curve in addition to a flat term structure of the ERP (Gordon 1962). Under these assumptions, equation (3) becomes

$$P_t = \sum_{k=0}^{\infty} \frac{D_t (1+g)^k}{(1 + R_t^f + ERP_t)^k} = \frac{D_t}{R_t^f + ERP_t - g}$$

Solving for the ERP gives

$$ERP_t = \frac{D_t}{P_t} - (R_t^f - g) \quad (5)$$

Note that even though the term-structure of the ERP is assumed to be flat, this model does not assume a constant ERP or a constant risk-free rate. In practice, there are several ways to operationalize equation (5). Model 3, called[3] the "Fed Model", uses the nominal ten-year Treasury yield as an estimate of $R_t^f - g$ and current earnings E_t as a proxy for current dividends D_t.

Model 4: The "Shiller model". Same as Model 3 but uses Shiller's cyclically adjusted price-earnings ratio (CAPE) as a proxy for the price-dividend ratio. CAPE is the current price of the S&P 500 divided by a trailing twelve month average of earnings.

Model 5: Same as Model 3, but uses the real ten-year Treasury yield as an estimate of $R_t^f - g$ (computed as the ten-year nominal Treasury rate minus the ten year breakeven inflation implied by TIPS). There are two typical justifications for this choice. First, in the long run, the growth rate of dividends g should be at least approximately equal to breakeven inflation. Second, the dividend-price ratio D_t/P_t, being the ratio of two nominal variables, is a real variable. Thus, it should be compared to the real risk-free rate and not to the nominal one as in Model 3.

Model 6: Same as Model 3, but uses one-year ahead expected earnings as a proxy for dividends. The usual justification is that including future expectations should better capture the forward-looking nature of the DDM.

Model 7: A variation in the assumptions in Models 3, 4 and 5: it uses the nominal ten-year Treasury yield and one-year ahead expected earnings.

[3] The name "Fed Model" was coined by Ed Yardeni, at Deutsche Morgan Grenfell, in reference to a report issued in 1997 by the Federal Reserve that used the model. However, the Federal Reserve has never endorsed this model. See Asness (2003) for a critical view of the "Fed Model".

Model 8: A two-stage DMM from Panigirtzoglou and Loeys (2005) where the first stage corresponds to the first five years, and the second stage corresponds to years 6 and onwards. In this case, formula (3) becomes[4]

$$P_t = \frac{D_t[1 + g_{LR} + 5(g_{SR} - g_{LR})]}{R_t^f + ERP_t - g_{LR}}$$

where R_t^f is the ten year nominal Treasury yield; D_t is estimated by the current (observable) level of earnings-per share multiplied by a payout ratio assumed to be 50%; g_{LR} is the long-run estimate for earnings growth and assumed to be 2.2 percent; g_{SR} is the estimated growth rate of earnings over the first five years, which is estimated by using the fitted values in a regression of average realized earnings growth over the last five years on its lag and lagged earnings-price ratio. The main advantage of having two stages instead of a single one (as in Models 3 through 7) is that it allows for changes in the growth rate of dividends, a useful feature when growth rates are far away from their long-run level.

Model 9: A multi-stage DDM constructed by Damodaran (2012). We simplify equation (3) by assuming there are 6 stages. Each of the first five stages corresponds to each of the first five years, while the last stage corresponds to years six and onwards. Dividends are assumed to grow at a rate g_t for each of the first five stages, and then at a rate equal to the ten year nominal Treasury yield for the final stage. The discount rate is assumed to be constant over different horizons, so that $\rho_{t+k} = \rho_t^k$. With these assumptions, equation (3) becomes

$$P_t = \sum_{k=1}^{5} \frac{D_t(1+g_t)^k}{\rho_t^k} + \frac{D_{t+6}(1+g_t)^6}{(\rho_t - R_t^f)\rho_t^5} \tag{6}$$

where $\rho_t = 1 + R_t^f + ERP_t$ and R_t^f is the ten year nominal Treasury yield. Given P_t, D_t, R_t^f and g_t, equation (6) determines a unique ERP_t.

Model 10: Is the same as Model 9 –and also proposed by Damodaran (2012)— but includes stock buybacks in cash flows. The idea is that investors care about total cash flows, not just dividends, and that buybacks are significant enough to affect measures of the ERP. In practice, we use free-cash-flow-to-equity as a proxy for dividends plus stock buybacks. Damodaran (2012) estimates that buybacks can increase the ERP by one to four percentage points per year.

4.3 Cross-sectional regressions

This method exploits the variation in returns and exposures to the S&P 500 of different assets to infer the ERP.[5] Intuitively, this method finds the ERP by answering the following question: what is the level of the ERP that makes expected returns of a variety of stocks consistent with their exposure to the S&P 500? Because we need to explain the relationship between returns and exposures for multiple assets with a single value for the ERP (and perhaps a small number of other controls), this model imposes tight restrictions on the estimation of the ERP.

The first step is to find the exposures of assets to the S&P 500 by estimating an equation of the following form:

[4] For a derivation, see Fuller and Hsia (1984).
[5] See Polk, Thompson and Vuolteenaho (2006) and Adrian, Crump and Moench (2012) for a detailed description of this method.

$$R_{t+k}^{i} - R_{t+k}^{f} = \alpha_{t+k}^{i} \times state\ variables_{t+k} + \beta_{t+k}^{i} \times risk\ factors_{t+k} + idiosyncratic\ risk_{t+k}^{i} \quad (7)$$

In equation (7), R_{t+k}^{i} is the realized return on a stock or portfolio from time t to $t+k$. $State\ variables_{t+k}$ are any economic indicators that help identify changes in the investment opportunity set (possibly including a constant). $Risk\ factors_{t+k}$ are any measures of systematic contemporaneous co-variation in returns across all stocks or portfolios. Finally, $idiosyncratic\ risk_{t+k}^{i}$ is the component of returns that is particular to each individual stock or portfolio that is not explained by $State\ variables_{t+k}$ or $Risk\ factors_{t+k}$. Examples of state variables are inflation, unemployment, the term spread, the yield spread between Aaa and Baa bonds and the S&P 500's dividend-to-price ratio. It is crucial that we include the excess return on the S&P 500 as a risk-factor in the estimation so that we can infer the ERP. Other risk-factors usually used are the Fama-French (1992) factors and the momentum factor of Carhart (1997). The value of α_{t+k}^{i} gives the strength of asset-specific return predictability and β_{t+k}^{i} gives the asset-specific risk exposures we are trying to estimate. For the cross-section of assets, we can use the whole universe of traded stocks, a subset of them, or portfolios of stocks grouped, for example, by industry, size, book-to-market or recent performance.

The second step is to find the ERP associated with the S&P 500 by estimating the cross-sectional equation

$$R_{t+k}^{i} - R_{t+k}^{f} = ERP_t(k) \times \hat{\beta}_{t+k}^{i} \quad (8)$$

where $\hat{\beta}_{t+k}^{i}$ are the values found when estimating equation (7). Equation (8) attempts to find the single number $ERP_t(k)$ (or vector of numbers, if we have more than one risk factor) that makes exposures β_{t+k}^{i} consistent with realized excess returns of all stocks or portfolios considered. The term structure of the ERP is obtained by computing returns over different horizons on the left hand side of equations (7) and (8).

One advantage of the cross-sectional regression method is that it uses more asset prices than other models, which provide more independent information about the ERP. Cross-sectional regressions also have sound theoretical foundations, since they are one way to implement Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM). Finally, this method nests many of the other models considered. The two main drawbacks of this method are that results are dependent on what portfolios, state variables and risk factors are used and that it is not easy to implement.

Model 11: The most widely used cross-sectional model is the Fama-French model (Fama and French 1992). The only state variable is a constant, and there are three risk factors: the returns on the market portfolio, a size portfolio and a book-to-market portfolio. Equation (7) is estimated by running rolling OLS regressions over the previous five years, and equation (8) is estimated by OLS without a constant[6].

Model 12: Same as Model 11, but includes momentum as an additional risk factor (Carhart 1997).

Model 13: Same as Model 12, but also includes inflation as a risk factor, which has been shown to account for a substantial part of the equity premium beyond the four factors of Carhart's model (Duarte 2013). Additionally, the time-varying coefficients α_{t+k}^{i} and β_{t+k}^{i} are estimated with the non-parametric kernel estimator of Ang and Kristensen (2012).

Model 14: This model is from Adrian, Crump, and Moench (2012). The state variables are the dividend yield, the default spread, and the risk free rate, which are commonly thought to capture changes in the

[6] Using OLS *with* a constant is an equally valid procedure; whether to include a constant depends on the familiar tradeoff between efficiency and robustness (Cochrane 2001).

investment opportunity set. The inclusion of these state variables allows the model to capture dynamics of the pricing kernel not captured by Models 11 through 13. The risk free rate is the one-month Treasury bill rate; the dividend yield is for the S&P500; and the default spread is calculated as the difference between Moody's seasoned Baa corporate bond yield and the 20-year Treasury bond yield at constant maturity. The market is the single risk factor. The model is estimated using a three step regression approach. First, the market return is orthogonalized with respect to the state variables and the residual of that regression is the considered the risk factor. Then each stock or portfolio's excess return is regressed on the lagged state variables and the risk factor to obtain the coefficients α^{i}_{t+k} and β^{i}_{t+k}. Finally, the ERP is obtained by estimating equation (8) using OLS.

4.4 Time-series regressions

This method uses the relationship between economic variables and stock returns to estimate the ERP. The idea is to run a linear regression of realized excess returns on lagged "fundamentals":

$$R_{t+k} - R^{f}_{t+k} = a + b \times Fundamental_t + error_t \tag{9}$$

Once estimates $\hat{a}$ and $\hat{b}$ for a and b are obtained, the ERP is obtained by ignoring the error term:

$$ERP_t(k) = \hat{a} + \hat{b} \times Fundamental_t \tag{10}$$

In other words, we estimate only the forecastable or expected component of excess returns. This method attempts to implement equations (1) and (2) as directly as possible in equations (9) and (10), with the assumption that "fundamentals" are the right sources of information to look at when computing expected returns and that the conditional expectation is a linear function.

The use of time-series regression requires minimal assumptions; there is no concept of equilibrium and no absence of arbitrage necessary for the method to be valid. In addition, implementation is quite simple, since it involves running univariate OLS regressions. The challenge of this method is to select the variables to include in the right-hand side of equation (9), since results can change substantially depending on what fundamental variables are used. In addition, including more than a single variable gives poor out-of-sample predictions even if economic theory may suggest a role for many variables to be used as predictors[7]. Finally, time-series regressions ignore information in the cross-section of stock returns.

The term structure of the ERP, as equations (9) and (10) suggest, is easily obtained in this method by simply running the predictive regressions with excess returns computed over different horizons.

Model 15: This model uses the dividend-price ratio as the only predictive variable. The key rationale is that the dividend-price ratio is first-order stationary so that it should eventually return to its long-run mean[8]: Values of the dividend-price ratio above its mean should forecast either low returns or high dividends going forward (and vice-versa for low values). Empirically, a high dividend-price ratio forecasts higher returns, not lower dividends, so the price-dividend ratio contains information about the ERP (Cochrane 2011).

Model 16: Same as Model 15, but uses the twelve predictive variables proposed by Goyal and Welch (2008). We use each variable independently and all of them together. At each point in time, we select

[7] Goyal and Welch (2008).

[8] See Lettau and Van Nieuwerburgh (2008) for an argument against first-order stationarity and its implications for predictability of returns.

the specification that performs the best out-of-sample (see Section 7 for a detailed description of how we do this) and use that specification for the next period. In the following period, we repeat the procedure; it is possible that this method uses different predictors depending on which one is performing best at each point in time.

Model 17: Same as Model 16, but imposes two restrictions on the estimation. First, the coefficient b in equation (9) is replaced by zero if it has the "wrong" theoretical sign. For example, if the price-dividend ratio has a negative coefficient, then we replace b by zero. Second, we replace the estimate of the ERP by zero if the estimation otherwise finds a negative ERP. These two restrictions are imposed one at a time and then together, and considered for the same twelve predictive variables considered in Model 16. The best specification at time *t* is used for prediction of *t+k* returns, so specifications can be changing over time. This model is advocated by Campbell and Thompson (2008), who argue that the restrictions, being based on theory, should improve estimation efficiency compared to unrestricted estimation.

Model 18: Uses as predictors the price-dividend ratio adjusted by the growth rate of earnings RE_t, dividends RD_t or stock prices RP_t:

$$RE_t = \frac{D_t}{P_{t-1}} + \left(\frac{E_t}{D_{t-1}}\right)\frac{CPI_{t-1}}{CPI_t} - 1 \tag{11}$$

$$RD_t = \frac{D_t}{P_{t-1}} + \left(\frac{D_t}{D_{t-1}}\right)\frac{CPI_{t-1}}{CPI_t} - 1 \tag{12}$$

$$RP_t = \frac{D_t}{P_{t-1}} + \left(\frac{P_t}{P_{t-1}}\right)\frac{CPI_{t-1}}{CPI_t} - 1 \tag{13}$$

where D_t are dividends, E_t are earnings, P_{t-1} is the lagged price of the S&P 500 and CPI_t is the consumer price index. We also consider the three measures constructed above, but subtracting the ten year nominal Treasury yield from each of them. The idea behind these measures is to impose –rather than assume— that stationary variables must eventually return to their long-run mean. As in models 16 and 17, at time t we use the predictor that has the best out-of-sample performance until $t-1$, which leads to different measures being used at different points in time. This model was proposed by Fama and French (2002) who argue that stock returns have been too high compared to dividend or earnings growth, and therefore must have been in part due to luck (positive shocks). A way to account for this sample-specific realization is to "correct" the dividend-price ratio as in equations (11), (12) and (13).

Model 19: The predictor is Baker and Wurgler's (2007) sentiment measure. The measure is constructed by finding the most predictive linear combination of five variables: the closed-end fund discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium. Baker and Wurgler (2007) have a more detailed explanation.

4.5 Surveys

The survey approach consists in asking economic agents what they think the ERP is. Surveys incorporate the views of many people, some of which are very sophisticated and/or make real investment decisions based on the level of the ERP. Surveys should also be good forecasters of the ERP because in principle stock prices are determined by supply and demand of investors such as the ones taking the surveys. On the other hand, Greenwood and Shleifer (2012) document that investor expectations of future stock market returns are positively correlated with past stock returns and with the current level of the stock market, but strongly *negatively* correlated with model-based expected returns and future realized stock market returns. Other studies such as Easton and Sommers (2007) also argue that survey measures of the ERP can be systematically biased.

The term structure of the ERP can only be obtained from surveys to the extent that questions are asked about the ERP at different horizons. To the authors' knowledge, the only consistent survey with a long enough time-series for analysis that asks about point estimates of the ERP at different horizons is the Duke CFO survey by Graham and Harvey (2012), which we use for the next model.

Model 20: Chief financial officers (CFOs) are asked about the one and ten-year-ahead ERP. A typical question in the survey is the following:

> *On November 19, 2007 the annual yield on 10-yr treasury bonds was 4.1 percent. Please complete the following: Over the next year, I expect the average annual S&P 500 return will be:*

The survey has grown over time and now has around 600 respondents. We take the mean of all responses as our measure of the ERP[9]. We construct the term-structure of the ERP by linearly interpolating the one and ten-year-ahead ERP estimates given by respondents. For this model, we do not construct ERP measures for horizons shorter than a year.

5. The Equity Risk Premium has Reached a Historic High

We summarize the behavior of the twenty models we consider by their first principal component. Let X be the matrix containing the demeaned ERP estimates at a monthly horizon from the different models we consider, with columns corresponding to models and rows corresponding to observation periods. The matrix is a 643-by-20 matrix, since we have 643 monthly observations and 20 models. The first principal component is the eigenvector of the variance-covariance matrix of X associated with the largest eigenvalue. Because X was demeaned, this principal component has mean zero. We take as our preferred ERP estimate the sum of the first principal component and the unconditional mean of ERP estimates across all models (i.e. the average of all elements of X). We repeat this process using ERP estimates at different horizons to obtain a single ERP time-series for each horizon. We call these estimates our *preferred measures*. The share of the variance explained by these measures ranges between 81% and 94%, suggesting that they are good summary statistics for the behavior of the models.

One challenge that arises in computing the principal component is that the matrix X has missing observations, either because some models can only be obtained at frequencies lower than monthly or because the necessary data is not available for all time periods (Appendix A contains a detailed description of when this happens). To overcome this challenge, we use an iterative linear projection method[10]. On the first iteration, we make a guess for the principal component and regress the non-missing elements of each row of X on the guess and a constant. We then find the first principal component of the variance-covariance matrix of the fitted values of these regressions, and use it as the guess for the next iteration. The process ends when the norm of the difference between consecutive estimates is small enough.

Figure 1 displays our preferred measures for the one-month and one-year-ahead ERP in blue and red, respectively. Recessions are indicated by shaded bars. The correlation between the two measures is

[9] Taking the median does not substantially alter results.

[10] We thank Richard Crump for suggesting this method and providing code for its implementation.

86%, but we do see that the one-month ahead ERP is sometimes above and sometimes below the one-year-ahead ERP, indicating that the slope of the term structure of the ERP is time-varying. As expected, the ERP measures tend to peak during financial turmoil, recessions and periods of low real GDP growth or high inflation. The ERP tends to bottom out after periods of sustained bullish stock markets and high real GDP growth.

The one year ahead ERP is at 14.5 percent in July of 2013, the highest it has even been. The one month ahead ERP is at 11.5 percent in July 2013, nearing the record levels obtained in February 2009, July 2012 and the early 1980s, but still below the peak of 15 percent in September of 1974.

The current high levels of the ERP are unusual in that we are not currently in a recession and we have just experienced an extended period of high stock returns, with 60 percent returns since July 2010 and almost 20 percent since the beginning of 2013. During previous periods, the ERP has always decreased during periods of sustained high realized returns. This is also the only period in which the ERP is elevated and the one-year ahead ERP is significantly higher than the one month ahead ERP.

Figure 2 displays in red the standard deviation of one-year-ahead ERP estimates across models for each time period. The standard deviation has been steadily decreasing since 2000 except for a few months during the financial crisis and has reached an all-time low in the last three months. A low standard deviation can be interpreted as models displaying a high degree of agreement – in this case, agreement that the ERP is high. Figure 2 also shows the reason for the recent increase in agreement and in the ERP by plotting the 25th and 75th percentiles of the distribution of models in blue and green, respectively. The interquartile range –the difference between the 25th and 75th percentiles— has compressed, mostly because the models in the bottom of the distribution have had higher ERP estimates since 2010. It is also interesting to note that the 75th percentile has remained fairly constant over the last 10 years, and is actually somewhat below its long-run mean.

6. The Term Structure of Equity Risk Premia

In Section 4, we described how each of the different models can trace out a term structure of the ERP – what expected excess returns are over different time horizons[11]. Figure 3 plots our preferred ERP measures as a function of investment horizon (rather than time) for some selected dates. The black line shows the average of the term structure across all periods. It is slightly upward sloping, with a short-term ERP at just over 6% and a three-year ERP at almost 7%. We selected the other dates because they are typical dates for when the ERP was unusually high or unusually low at the one-month horizon. We see that the ERP is strongly mean-reverting, with the term structure sloping downward for high one-month ERP periods, and sloping upward for low one-month ERP periods. In contrast, the ERP in July 2013 is upward sloping, something that has never happened in periods of elevated ERP.

7. Why is the Equity Risk Premium High?

The last two sections showed evidence that the ERP is high at all horizons, and that this is an unusual occurrence given the current economic and financial environment. There are two reasons why the ERP can be high: low discount rates and high current or expected future cash flows.

---

[11] For other ways to estimate the term structure of the ERP using equilibrium models or derivatives, see Ait-Sahalia, Karaman and Mancini (2012), Ang and Ulrich (2012), Berg (2013), Boguth, Carlson, Fisher and Simutin (2012), Croce, Lettau and Ludvigson (2012), Lemke and Werner (2009), Lettau and Wachter (2011), Muir (2013), among others.

Figure 4 shows that earnings are likely not the reason why the ERP is high. The blue line shows the realized monthly growth rates of real earnings for the S&P500 expressed in annualized percentage points. Since 2010, earnings growth has been declining, hovering around zero for the last few months of the sample. It currently stands at 2.5%, which is near its long-run average. Perhaps more importantly for the equity premium, the expectations of future earnings growth since 2010 have also been moderate to low. The red line in Figure 4 shows the year-on-year change in the mean expectation of one-year-ahead earnings per share for the S&P500. Similarly to realized earnings growth, earnings per share have been declining over the last three years, making expected earnings growth an unlikely reason for why the ERP is near its all-time high.

Nominal and real bond yields, on the other hand, have been exceptionally low since the end of the financial crisis. Figure 5 displays the term structure of the ERP under two counterfactual scenarios, in addition to the mean and current term structures already displayed in Figure 4. In the first counterfactual scenario, we leave expected stock returns unmodified but change the risk-free rates from the current values of nominal bond yields to the average nominal bond yields over 1960-2013. In other words, we replace R^f_{t+k} in equation (2) by the mean of R^f_{t+k} over t. The result of this counterfactual is shown in Figure 5 in orange. Using average levels of bond yields brings the whole term structure of the ERP much closer to its mean level (the black line), especially at short horizons. This shows that a "normalization" of bond yields, everything else being equal, would bring the ERP down substantially. In our second counterfactual exercise, we do not keep expected stock returns unchanged, but instead estimate the following regression:

$$\Delta ERP_t(k) = a(k) + b(k) \times \Delta y^1_t + c(k) \times \Delta y^5_t + d(k) \times \Delta y^{10}_t + error_t(k) \tag{14}$$

where y^1_t, y^5_t and y^{10}_t are nominal yields for one, five and ten-year constant maturity bonds, and Δ is the first-difference operator, i.e. $\Delta x_t = x_t - x_{t-1}$ for any variable x. Equation (14) can be thought of as regressing $\Delta ERP_t(k)$ on basic level, slope and curvature factors of the nominal yield curve since these factors are linear combinations of the three bond yields y^1_t, y^5_t and y^{10}_t. We chose to run regression (14) in differences to avoid spurious regression bias, since bond yields and the ERP are persistent variables[12]. We then add to the current term structure of the ERP the fitted values of regression (14) that result from plugging in the values of $\Delta y^1_t, \Delta y^5_t$ and Δy^{10}_t that would bring bond yields from their current levels to their historical levels:

$$\Delta y^j_t = \bar{y}^j - y^j_{July\ 2013}$$

for $j = 1, 5$ and 10 years, where $\bar{y}^j$ is the time-average of y^j_t. The resulting counterfactual term structure of the ERP is shown in green in Figure 5. Unlike the case in which expected returns were held constant, this counterfactual assumes that expected returns respond to changes in yields in the same way that they have responded in the past. The resulting counterfactual term structure of the ERP is now flat and substantially below its average value. This means that if yields increased to their average levels and expected returns reacted to this increase as they have in the past, the ERP would decrease below its average levels at all horizons. This exercise shows that the current environment of exceptionally low bond yields is capable, quantitatively speaking, of causing an ERP as high as we are currently observing.

[12] An augmented Dickey-Fuller test fails to reject a unit root in the EPR at the 5% level using any number of lags between 5 and 18 (the maximum lag chosen by the Schwert criterion). The tau test statistic using 15 lags (the optimal number of lags obtained by the Ng-Perron procedure) is -1.96, which is smaller than the critical value of -2.83. A similar analysis for bond yields also fails to reject the null hypothesis of a unit root.

8. Are excess returns predictable?

In this section, we analyze how ERP models perform when trying to predict the realized equity premium. There is substantial debate in the academic literature[13] on whether any model can explain or predict the ERP better than the historical mean of realized excess returns. For this reason, we choose performance metrics that are relative to the historical mean. The historical mean itself is only a weak explanatory and predictive variable. Its correlation with the realized ERP is about 6 percent, and the R^2 in one to sixty month predictive regressions is less than 1 percent.

The measure we use for how well a model predicts the ERP is the out-of-sample R^2, popularized by Campbell and Thompson (2008):

$$R_{OOS}^2 = 100 \times \left(1 - \frac{\sum_{t=1}^{T}(R_t^e - Model\ ERP_t)^2}{\sum_{t=1}^{T}(R_t^e - Historical\ mean_t)^2}\right) \tag{15}$$

Here R_t^e are the realized excess returns at time t, $Model\ ERP_t$ is the time t real-time estimate of the ERP given by some model and $Historical\ mean_t$ is the real-time mean of the ERP since the beginning of the sample. Because $Model\ ERP_t$ is computed using information available at $t-1$ but R_t^e is only realized at time t, we interpret R_{OOS}^2 as a measure of how well the model predicts the ERP compared to the historical mean.

The out-of sample R^2 in equation (15) ranges from minus infinity to +100 percent. If the R^2 is 0, the historical mean and the model in question perform equally well, whereas a positive R^2 implies that the model outperforms the historical mean. Note that this measure is not a traditional R^2, so its units cannot be interpreted as the percentage of the variance of the dependent variable explained by the model. However, the R^2 numbers have an intuitive economic interpretation. A mean-variance investor with coefficient of relative risk aversion equal to γ, when using the $Model\ ERP_t$ as the measure of expected excess returns, will earn returns over the whole sample in excess of those predicted by the historical mean equal to R_{OOS}^2/γ.[14] For example, if $\gamma = 2$, and $R_{OOS}^2 = 10$ percent, then the investor can, though better predictability, earn extra returns of 5 percent over the T periods considered. Of course, if R_{OOS}^2 is negative, using the model would lead to returns that are lower than those obtained by using the historical mean as the sole estimate for the ERP.

Figures 6 and 7 show the results. In Figure 6, we show the R_{OOS}^2 for the models that perform the best within each category. For DDM, the Shiller model (model 4) outperformed all other DDM at all horizons. For cross-sectional regressions, the model by Adrian, Crump and Moench (model 14) did best at the one month to two year horizons, but was outperformed by Fama-French and momentum (model 12) at the three and four year horizons, and by the cross-sectional regression with inflation (model 13) at the five year horizon. For time-series regressions, the results were mixed: The Goyal and Welch predictors (model 16) were the best at the one month and three year horizons; The predictors in Fama and French (model 18) were the best at the four and five year horizons; The dividend-price ratio (model 15) was the best at the other horizons. For surveys, all results correspond to CFO surveys (model 20), which is the only survey we analyze.

The main conclusion is that the R-squares are small, which means that none of the models drastically outperform the historical mean out of sample. For example, cross-sectional regressions (the green line with crosses) starts at almost 12%, which means that the mean-variance investor with coefficient of

[13] See, for example, Goyal and Welch (2008), Lettau and Van Nieuwerburgh (2008), Campbell and Thompson (2008), Cochrane (2011), Fama and French (2002) and references therein.
[14] For a derivation of this fact, see Campbell and Thompson (2008).

relative risk aversion equal to one would have made an extra 12% over the last 53 years compared to just using the historical mean as a measure of the ERP. This amounts to about 23 basis points per year. In addition, although surveys are clearly inferior predictors, all other predictors are comparable at all horizons.

Figure 7 displays the same analysis but using principal components instead of the individual models for each category. The principal components were computed in real time following the procedure explained in Section 4. The predictability of the principal component need not be better or worse than the best model in each category. One the one hand, the principal component may reduce noise and aggregate useful information from the many models. On the other hand, it puts some weight on models that have worse predictability than the best models. Figure 7 shows that, as a group, dividend discount models perform substantially better than other models at short horizons but are worst at long horizons, while cross-sectional regressions perform best at long horizons but are worst at short horizons. The principal component of all models, in the solid purple line, performs well across all horizons and is always close to the model with the best predictability. The good performance of the principal component reinforces its usefulness as a summary statistic.

9. Conclusion

Estimates for the ERP as high as we have found should give policymakers pause. We have argued that it is unusual for the ERP to be at its present level in the current stage of the business cycle, especially when expectations are that it will continue to rise over the next three years. Because the ERP is a key input in many important decisions of economic agents, an unusually high ERP can herald unusual behavior. Our analysis provides evidence that is consistent with a bond-driven ERP: expected excess stock returns are high not because stocks are expected to have high returns, but because bond yields are exceptionally low. In such an environment, we should expect monetary policy –both conventional and unconventional—to have a large impact on asset prices and hence the real economy.

Our study of the ERP has many limitations. The main one is that stocks returns are very difficult to predict, if they are predictable at all. We have shown how to improve upon current estimates by using principal components yet still found weak evidence in favor of predictability, at least at horizons shorter than five years. Any conclusions that rely on ERP estimates must be weighted by how strongly it predicts future returns. Another limitation is that even though we have conducted all of our out-of-sample tests in real time (using information available at time t for $t+1$ estimates), some of the models we use had not been yet proposed for many periods of our sample, so there is some selection and forward looking biases. Finally, we have not focused on the possibility that a bubble –rational or irrational— could be a further driver of the recent high realized and expected returns, a topic we consider outside of the scope of the broadly used models we consider.

References

Adrian, T., Crump, R., Moench E., 2012. Efficient regression-based estimation of dynamic asset pricing models. Federal Reserve Bank of New York Staff Reports, Number 439, May 2011.

Ait-Sahalia, Y., Karaman, M., Mancini, L., 2012. The term structure of variance swaps, risk premia and the expectation hypothesis. Risk Premia and the Expectation Hypothesis (August 24, 2012).

Ang, A., Kristensen, D., 2012. Testing conditional factor models. NBER Working Paper 17561.

Ang, A., Ulrich, M., 2012. Nominal Bonds, Real Bonds, and Equity.

Asness, C., 2003. Fight the Fed model. Journal of Portfolio Management (Fall).11-24.

Berg, T., 2013. The Term Structure of Risk Premia: Evidence from CDS Spreads.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica, . 77 (3), 623-685.

Boguth, O., Carlson, M., Fisher, A. J., Simutin, M., 2012. Leverage and the limits of arbitrage pricing: Implications for dividend strips and the term structure of equity risk premia. manuscript, Arizona State University.

Campbell, J.Y., Thompson, S., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21 (4), 1509-1531.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Cochrane, J. H., 2001. Asset pricing. Princeton University Press, New Jersey.

Cochrane, J.H., 2011. Presidential address: discount rates. Journal of Finance 66 (4), 1047-1108.

Croce, M., Lettau, M., Ludvigson, S., 2012. Investor information, long-run risk, and the term structure of equity. Unpublished paper, New York University.[40, 49].

Duarte, F., 2013. Inflation and the cross-section of stock returns. Federal Reserve Bank of New York Staff Reports, Number 621.

Duarte, F., Kogan, L., Livdan, D., 2013. Aggregate Investment and Stock Returns. Unpublished paper, M.I.T.

Damodaran, A., 2012. Equity Risk Premiums (ERP): Determinants, Estimation and Implications – The 2012 Edition. Working paper New York University.

Easton, P.D., Sommers, G.A., 2007. Effect of analysts' optimism on estimates of the expected return implied by earnings forecasts. Journal of Accounting Research 45, 983-1016.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47 (2), 427-465.

Fama, E.F., French, K.R., 2002. The equity premium, Journal of Finance, 57 (2), 637-659

Fuller, R., Hsia, C., 1984. A Simplified Common Stock Valuation Model, Financial Analysts Journal, September-October, 1984.

Gordon, M. J., 1962. The investment, financing, and valuation of the corporation. Greenwood Press.

Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium predictions. Review of Financial Studies 21 (4), 1455-1508.

Graham, J., Harvey, C., 2012. The Equity Risk Premium in 2012. Working paper, Duke University.

Greenwood, R., Shleifer, A., 2012. Expectations of returns and expected returns. Working paper, Harvard University.

Gürkaynak, R. S., Sack, B., & Wright, J. H., 2007. The US Treasury yield curve: 1961 to the present. Journal of Monetary Economics, 54(8), 2291-2304.

Hall, R. E., 2013. The Routes into and out of the Zero Lower Bound. Working paper, Stanford University.

Kuehn, L. A., Petrosky-Nadeau, N., & Zhang, L., 2012. An equilibrium asset pricing model with labor market search (No. w17742). National Bureau of Economic Research

Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the return predictability evidence. The Review of Financial Studies 21 (4), 1607-1652.

Lettau, M., Wachter, J. A., 2011. The term structures of equity and interest rates. Journal of Financial Economics, 101(1), 90-113.

Ludvigson, S.C., Lettau, M., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56 (3), 815-849.

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41 (5), 867-887.

Muir, T., 2013. Financial Crises, Risk Premia, and the Term Structure of Risky Assets.

Panigirtzoglou, N., Loeys, J., 2005. A fair value model for US bonds, credit and equities. JP Morgan, June 2005.

Polk, C., Thompson, S., Vuolteenaho, T., 2006. Cross-sectional forecasts of the equity premium. Journal of Financial Economics 81, 101-141.

Appendix A: Data Variables

Variable	Source	Original Frequency	First period	Last Period
Sentiment measure from "Investor Sentiment in the Stock Market," Journal of Economic Perspectives, 2007	Jeffrey Wurgler	monthly	07/01/65	12/01/10
Equity issuance	Jeffrey Wurgler	monthly	01/01/58	12/01/10
Debt issuance	Jeffrey Wurgler	monthly	01/01/58	12/01/10
CAY	Martin Lettau	quarterly	03/01/52	09/01/12
One-year ahead ERP from CFO survey	Duke CFO Survey	quarterly	06/06/00	06/05/13
Ten-year ahead ERP from CFO survey	Duke CFO Survey	quarterly	06/06/00	06/05/13
Book value per share	Compustat	annual	12/31/77	12/31/12
S&P 500 closing price	Compustat	monthly	02/28/62	06/30/13
ERP from a dividend discount model	Damodaran	annual	12/01/61	12/01/12
ERP from a dividend discount model using free cash flow	Damodaran	annual	12/01/61	12/01/12
Size and book-to-market sorted portfolios	Fama-French	monthly	07/01/26	06/01/13
Realized excess returns for the market	Fama-French	monthly	07/01/26	06/01/13
Size factor	Fama-French	monthly	07/01/26	06/01/13
Book-to-market factor	Fama-French	monthly	07/01/26	06/01/13
Risk free rate	Fama-French	monthly	07/01/26	06/01/13
Baa minus Aaa bond yield spread	FRED	monthly	01/01/19	07/01/13
NBER recession indicator	FRED	monthly	01/02/00	06/01/13
Momentum portfolios	Fama-French	monthly	01/01/27	12/01/12
Momentum factor	Fama-French	monthly	01/01/27	06/01/13
ERP as constructed in Adrian, Crump and Moench (2013)	NY Fed	monthly	01/01/63	07/01/13
Nominal price for the S&P 500	Shiller	monthly	01/02/00	07/01/13
Nominal dividends for the S&P 500	Shiller	monthly	01/02/00	06/01/13
Nominal earnings for the S&P 500	Shiller	monthly	01/02/00	03/01/13
Consumer Price Index	Shiller	monthly	01/02/00	07/01/13
10 year nominal treasury yield	Shiller	monthly	01/02/00	07/01/13
Real price for the S&P 500	Shiller	monthly	01/02/00	07/01/13
Real dividends for the S&P 500	Shiller	monthly	01/02/00	06/01/13
Real earnings for the S&P 500	Shiller	monthly	01/02/00	03/01/13
Cyclycally Adjusted Price-Earnings ratio	Shiller	monthly	01/02/00	07/01/13
Realized Earnings per Share for the S&P 500	Thomson Reuters I/B/E/S	annual	12/31/81	12/31/12
Mean analyst forecast for Earnings per share for the S&P500	Thomson Reuters I/B/E/S	monthly	01/14/82	04/18/13
Yield on TIPS	Fed Board	monthly	01/01/03	07/01/13
Nominal yields	Fed Board (Gurkaynak, Sack and Wright)	daily	06/14/61	08/12/13

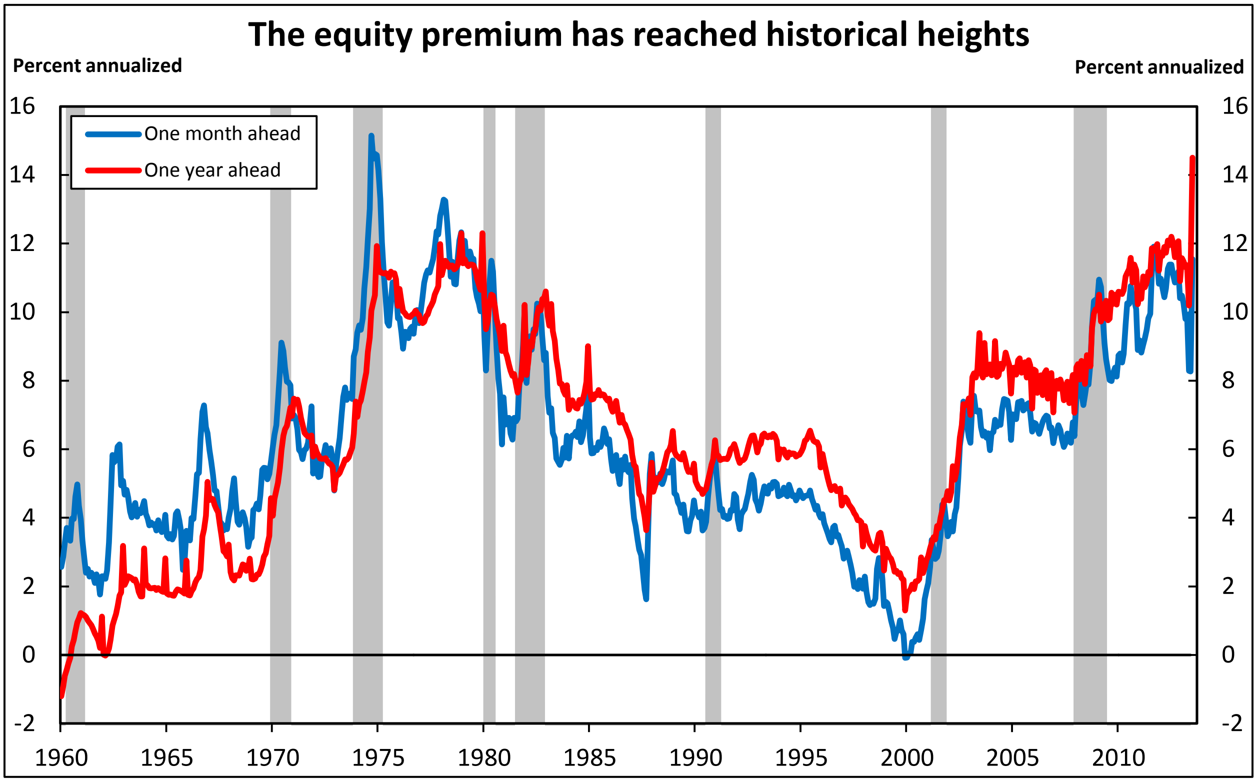


Figure 1 The equity risk premium (expected excess returns) over a one year ahead and one month ahead horizons are the first principal components of 20 models of the equity premium. The models include time-series and cross-sectional regressions, dividend discount models and surveys. Shaded bars are NBER recessions.

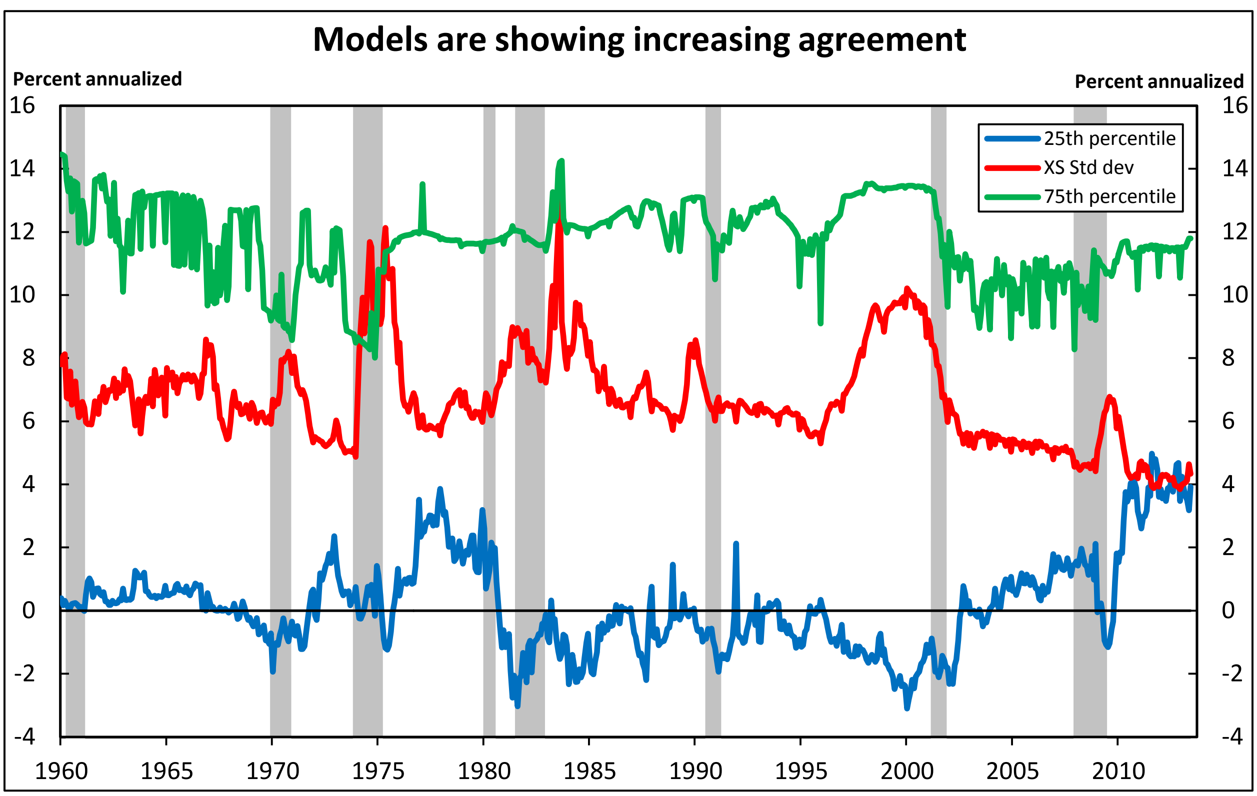


Figure 2 The cross-sectional standard deviation (labeled "XS Std dev", in red) computes, at each time period, the standard deviation of the 20 equity risk premium estimates given by the different models. The 25th and 75th percentiles (in blue and green, respectively) give the corresponding quartile of the 20 estimates for each time period. Shaded bars are NBER recessions.

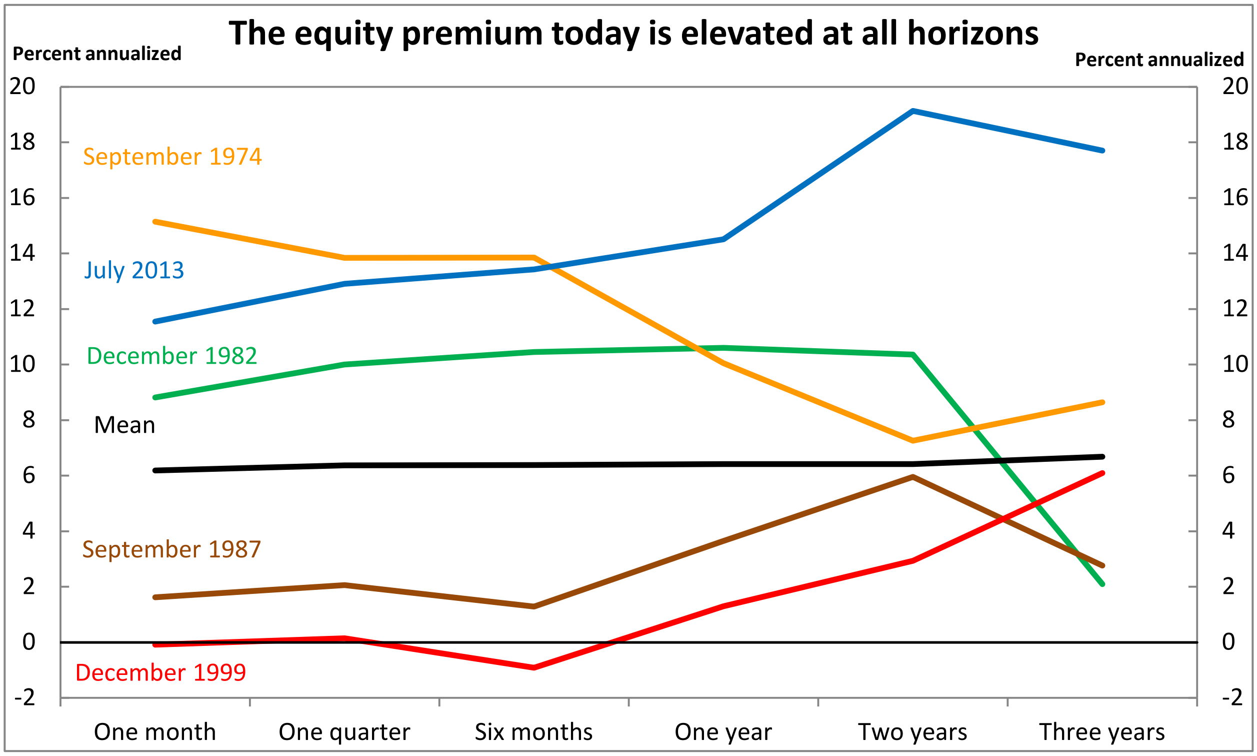


Figure 3 The equity risk premium at different horizons are the first principal component of 20 estimates of expected excess returns at different horizons. The estimates are obtained from cross-sectional and time-series regressions, dividend discount models and surveys. The black line (labeled "Mean") shows the mean of expected excess returns at different horizons over the sample 1960-2013. The most recent estimates of the term structure of the equity risk premium (labeled "July 2013" in blue), does not show mean reversion, unlike other periods when the equity risk premium was substantially above or below its mean at the one-month horizon.

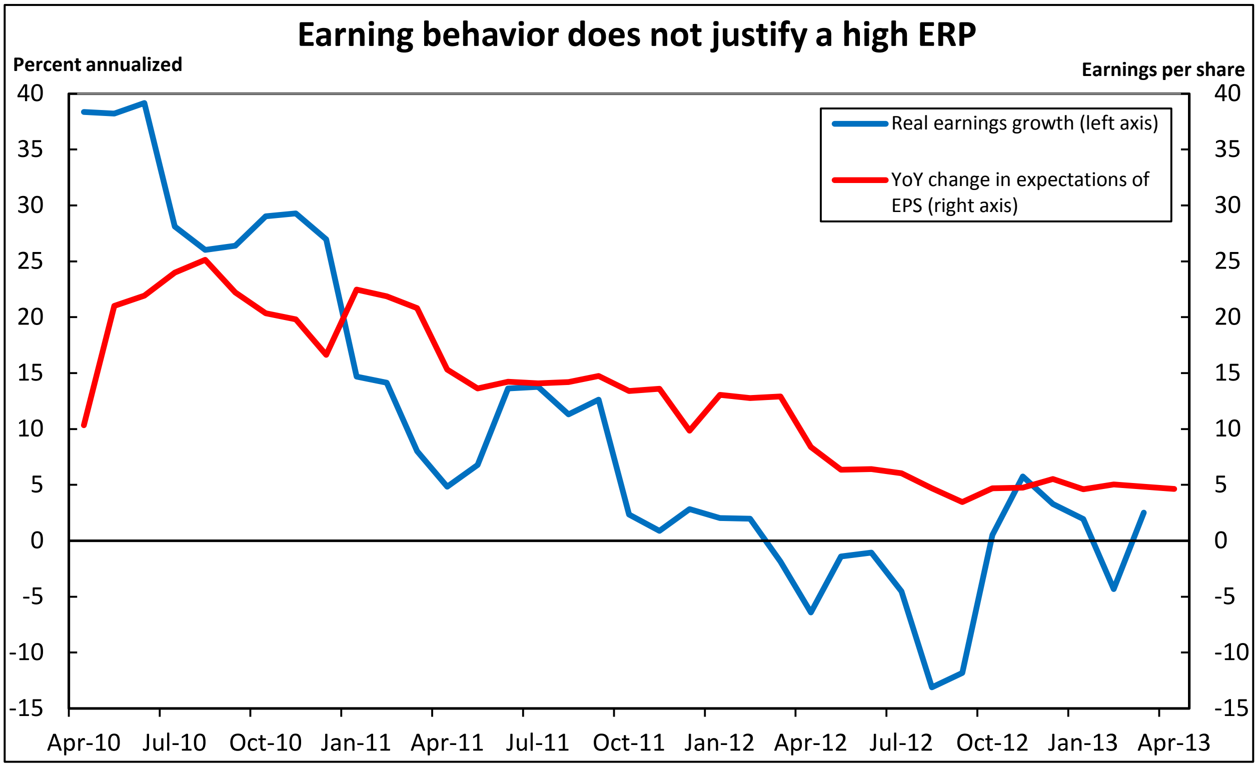


Figure 4 The blue line shows the monthly growth rate of real S&P 500 earnings, annualized and in percentage points. The red line shows the year-on-year change in the mean expectation of one-year ahead earnings per share for the S&P 500 from analysts from Thomson Reuters I/B/E/S.

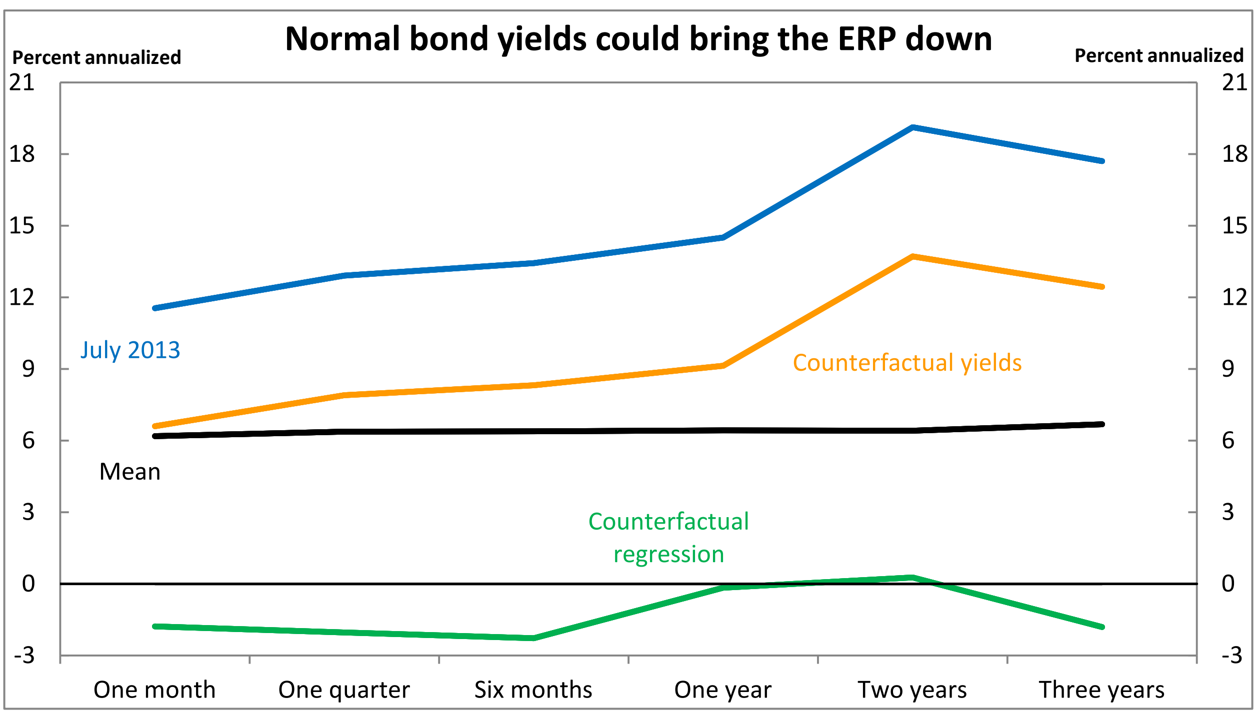


Figure 5 The black line (labeled "Mean") shows the mean term structure of the equity risk premium over the sample 1960-2013. The blue line (labeled "July 2013") shows the most recent estimates. The orange line (labeled "Counterfactual yields") shows what the term structure of the equity risk premium would be in July 2013 if instead of subtracting today's yield curve from expected returns we subtracted the average yield curve for the period 1960-2013. The green line shows an estimate of what the term structure of the equity premium would be if yields rose to their average historical levels and expected stock returns co-moved with yields with the same correlation as during 1960-2013.

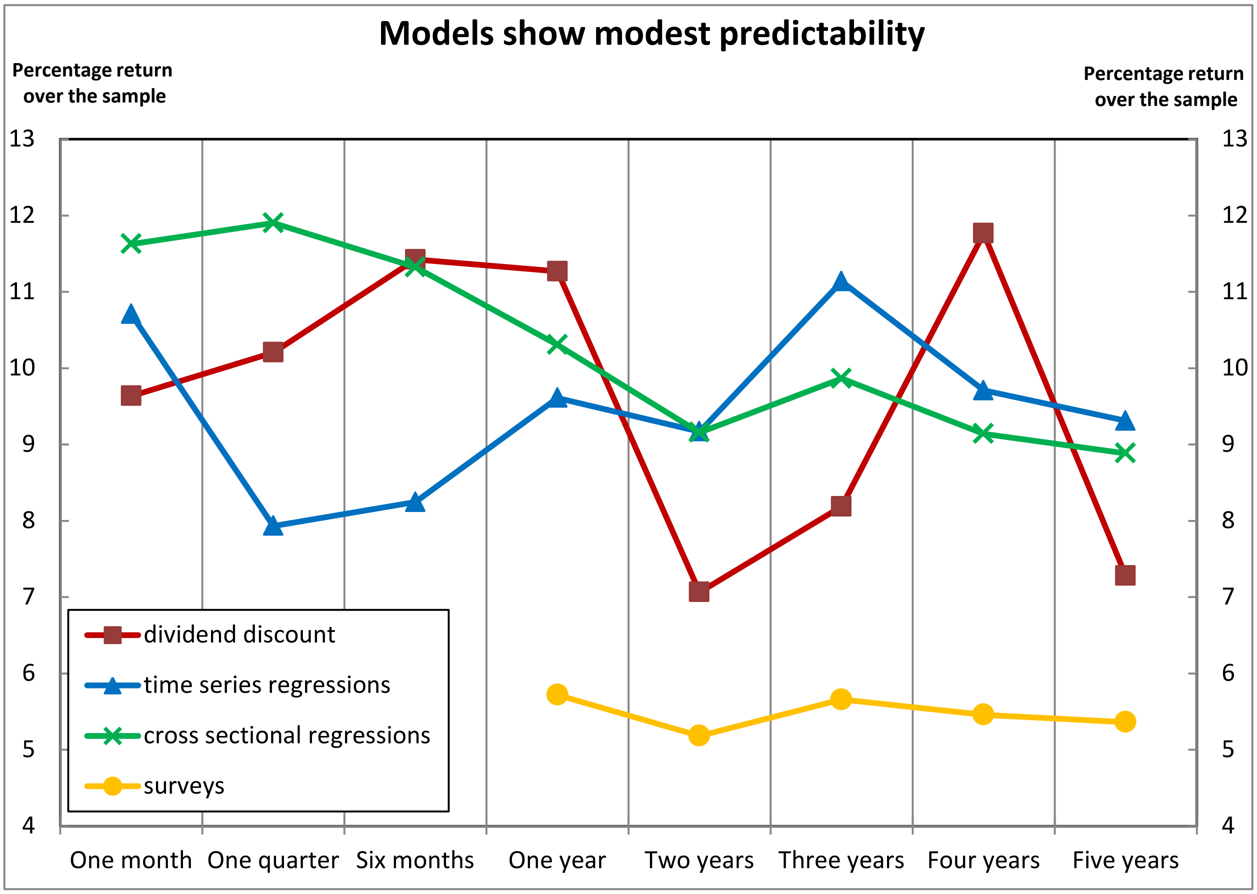


Figure 6 Each data point corresponds to the returns that a mean-variance investor with unit coefficient of relative risk aversion would have earned over the period 1960-2013 if she had used one of the equity risk premium models over and above the returns she would have made if she had assumed that expected excess returns are equal to their historical mean at each point in time. The x-axis shows the investment horizon of the investor (how often the portfolio is rebalanced and hence how far ahead excess returns must be forecast).

For each class of model (dividend discount, time-series regressions, cross-sectional regressions, surveys) we report the model that had the best predictability. For dividend discount models, the Shiller model (model 4) outperformed all other discount models at all horizons. For cross-sectional regressions, the model by Adrian, Crump and Moench (model 14) did best at the one month to two year horizon, but was outperformed by Fama-French and momentum (model 12) at the three and four year horizons, and by the cross-section with inflation (model 13) at the five year horizon. For time-series regressions, the results were mixed: The Goyal and Welch predictors (model 16) were the best at a one month and three year horizons; The predictors in Fama and French (model 18) were the best at the four and five year horizons; The dividend-price ratio (model 15) was the best at the other horizons. For surveys, all results correspond to CFO surveys (model 20), which is the only survey we analyze.

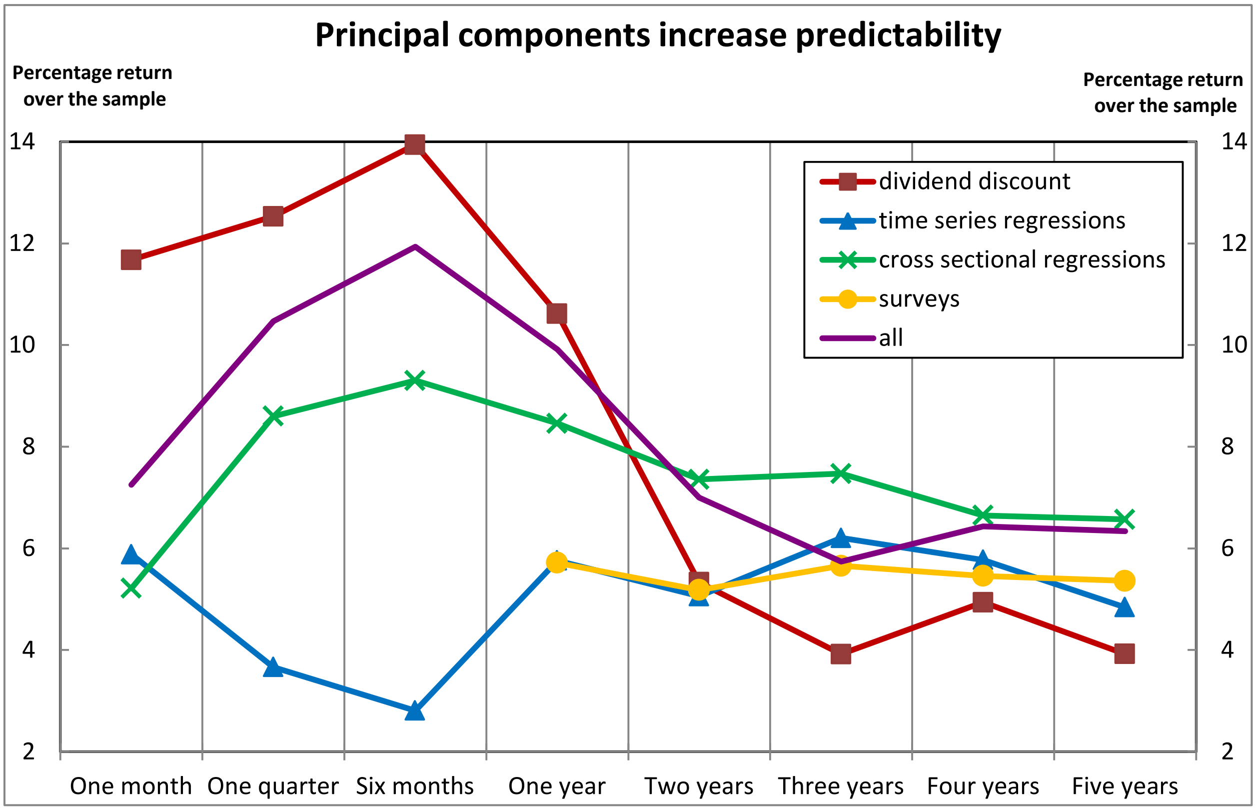


Figure 7 Each data point corresponds to the returns that a mean-variance investor with unit coefficient of relative risk aversion would have earned over the period 1960-2013 if she had used the first principal component of all models within a certain class (dividend discount, time-series regressions, cross-sectional regressions, surveys) over and above the returns she would have made if she had assumed that expected excess returns are equal to their historical mean at each point in time. The x-axis shows the investment horizon of the investor (how often the portfolio is rebalanced and hence how far ahead excess returns must be forecast). The line labeled "all" corresponds to the principal component of all models (our preferred measure).

DUFF & PHELPS

Table: Equity Risk Premium & Risk-Free Rates

March 6, 2014

Duff & Phelps Recommended Equity Risk Premium (ERP) and Corresponding Risk-Free Rates (R_f); January 2008–Present

For additional information, please visit www.duffandphelps.com/CostofCapital

	Duff & Phelps Recommended ERP	Risk-Free Rate
Current ERP Guidance February 28, 2013 – UNTIL FURTHER NOTICE	5.0%	4.0% Normalized 20-year Treasury yield *
Year-end 2013 Guidance December 31, 2013	5.0%	4.0% Normalized 20-year Treasury yield *
January 1, 2013 – February 27, 2013	5.5%	4.0% Normalized 20-year Treasury yield *
Year-end 2012 Guidance December 31, 2012	5.5%	4.0% Normalized 20-year Treasury yield *
Change in ERP Guidance January 15, 2012 – February 27, 2013	5.5%	4.0% Normalized 20-year Treasury yield *
Change in ERP Guidance September 30, 2011 – January 14, 2012	6.0%	4.0% Normalized 20-year Treasury yield *
July 1 2011 – September 29, 2011	5.5%	4.0% Normalized 20-year Treasury yield *
June 1, 2011 – June 30, 2011	5.5%	Spot 20-year Treasury Yield
May 1, 2011 – May 31, 2011	5.5%	4.0% Normalized 20-year Treasury yield *
December 1, 2010 – April 30, 2011	5.5%	Spot 20-year Treasury Yield
June 1, 2010 – November 30, 2010	5.5%	4.0% Normalized 20-year Treasury yield *
Change in ERP Guidance December 1, 2009 – May 31, 2010	5.5%	Spot 20-year Treasury Yield
June 1, 2009 – November 30, 2009	6.0%	Spot 20-year Treasury Yield
November 1, 2008 – May 31, 2009	6.0%	4.5% Normalized 20-year Treasury yield *
Change in ERP Guidance October 27, 2008 – October 31, 2008	6.0%	Spot 20-year Treasury Yield
January 1, 2008 – October 26, 2008	5.0%	Spot 20-year Treasury Yield

* Normalized in this context means that in months where the risk-free rate is deemed to be abnormally low, a proxy for a longer-term sustainable risk-free rate is used.