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

1. Footnotes appear at end of article.

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.

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)*[b,c]	*Alpha (3)*[c,b]	*Alpha (12)*[c,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)