Journal of Accounting Research
Vol. 28 No. 2 Autumn 1990
Printed in U.S.A.

Predicting Individual Analyst Earnings Forecasts

SCOTT E. STICKEL*

1. *Introduction*

In this study I propose and test a model that predicts individual analyst forecasts of corporate earnings per share (*EPS*) using the change in the mean consensus forecast of other analysts since the date of the analyst's current outstanding forecast; the deviation of the analyst's current forecast from the consensus forecast; and cumulative stock returns since the date of the analyst's current forecast. I find that these three variables explain about 38% of the variability in analyst forecast revisions. While there is evidence of a relation between changes in earnings expectations and price changes, virtually all of the explanatory power of my model arises from *other* analyst forecasts.

Section 2 describes the data bases used and the sample selection process. Section 3 presents the model and method for predicting individual analyst forecasts. Section 4 reports the bias and accuracy of the predicted forecasts. Conclusions are in section 5.

2. *Data Bases and Sample Selection Process*

Individual analyst forecasts of annual *EPS* are supplied by Zacks Investment Research (Zacks). Daily returns data for firms listed on the New York Stock Exchange or American Stock Exchange are provided by the Center for Research in Security Prices (*CRSP*) at the University of Chicago.

* University of Pennsylvania. I received helpful comments from Larry Brown, Nick Gonedes, Prem Jain, Rick Lambert, Dave Larcker, Don Lewin, Jody Magliolo, participants at the 1989 International Symposium on Forecasting, and an anonymous referee. I appreciate the financial support of the KPMG Peat Marwick Foundation, Deloitte Haskins & Sells, and the Institute for Quantitative Research in Finance. I am also grateful to Zacks Investment Research, Inc. for supplying the analyst forecasts.

Individual analyst forecast revisions included in the sample meet these four criteria: (1) the forecast revision and the fiscal year-end of the firm are within 1980–85; (2) stock return data are available on the *CRSP Daily Returns File* of NYSE and ASE firms; (3) the forecast revision date is within 200 trading days of the date of the analyst's prior forecast and within the current fiscal year of the firm; (4) there are at least two analysts with an outstanding forecast for the firm on the dates of the original forecast and the revision.

Table 1 summarizes the sample selection process. Of the approximately 3,600 firms on the Zacks data base for fiscal year-ends within 1980–85, about 1,500 have revisions that meet the sample selection criteria. No industry appears to be missing from the final sample, which includes many banks and utilities as well as industrial companies. However, excluded firms are, on average, smaller than sample firms. Thus, the inferences made from the final sample may not be applicable to very small firms with analyst following.

3. *Predicting Individual Analyst Earnings Forecasts*

3.1 THE MODEL AND METHOD FOR PREDICTING FORECASTS

I use publicly available information released since the date of an analyst's current forecast to predict his next forecast. Assume the current day is day $t - 1$. Define $FRCST_{i,a,t}$ as a revised forecast of *EPS* for company i to be issued by analyst a on day t and define $FRCST_{i,a,t-v}$ as the current outstanding forecast dated v days prior to day t. A positive relation is hypothesized between each of the following three pieces of information and the change in investors' expectations of $FRCST_{i,a,t}$ between day $t - v$ and $t - 1$:

1. The change in the mean consensus forecast of other analysts

TABLE 1

Summary of Sample Selection Process for Individual Forecast Revisions

	Revisions
Total revisions on Zacks files dated within the 1980–85 calendar period and within the current fiscal year of the firm for firms with fiscal year-ends within the 1980–85 period	263,962
Revisions excluded:	
Firm not on *CRSP* file of NYSE and ASE firms	(59,810)
Date of the forecast is more than 200 trading days after the date of the analyst's prior forecast	(6,843)
Only one analyst with an outstanding forecast	(5,996)
Remaining revisions included in forecast prediction regressions (table 2)	191,313
Revisions within the 1980 calendar year	(16,613)
Remaining revisions included in measures of forecast predictability (table 3)	174,700

following firm i between days $t - v$ and $t - 1$. This change proxies for new information released after date $t - v$.

2. The difference between the mean consensus forecast of other analysts following firm i on day $t - v$ and analyst a's forecast on day $t - v$. Zacks supplies brokerage houses with a "deviation report," which officials at Zacks believe pressures analysts to issue forecasts closer to the consensus.
3. The cumulative return to firm i from days $t - v$ to $t - 1$, multiplied by the forecast by analyst a on day $t - v$. This return also proxies for new information released after date $t - v$.

To mitigate potential problems from calendar clustering, the sample is segregated into 144 subsamples on the basis of the semimonthly period in which day t falls, and tests are performed on the data by *subsample.* This design subsumes any cross-sectional temporal dependence *within* subsamples and reduces any cross-sectional temporal dependence *between* subsamples. The significance of mean results from these 144 subsamples is determined by dividing the mean by its standard error, which is the estimated standard deviation of the 144 observations divided by the square root of 144 (see Fama and MacBeth [1973]).

I estimate the following ordinary least squares regression for each of the 144 subsamples:

$$(FRCST_{i,a,t} - FRCST_{i,a,t-v}) = \beta_0 + \beta_1 (CONS_{i,t-1} - CONS_{i,t-v}) + \beta_2 (CONS_{i,t-v} - FRCST_{i,a,t-v}) + \beta_3 (FRCST_{i,a,t-v} * CR_{i,(t-v,t-1)}) + \epsilon_{i,a,t}.$$

$CONS_{i,t-1}$ is the mean consensus forecast, excluding analyst a, of *EPS* for company i on day $t - 1$ and is calculated as the equally weighted average of all other individual forecasts.[1] $CR_{i,(t-v,t-1)}$ is the cumulative stock return for firm i from day $t - v$ to day $t - 1$.

3.2 EMPIRICAL RESULTS FOR PREDICTING INDIVIDUAL ANALYST FORECASTS

Table 2 reports the mean results of the 144 regressions. The mean coefficient on each explanatory variable is significantly different from zero.[2] The mean-adjusted R-square is .38.[3] The results suggest that an individual analyst's next forecast is a positive function of all three proxy

---

[1] Because there can be up to 200 days between forecasts, cross-sectional dependence of error terms between subsamples is a concern. The sensitivity of the results to this source of dependence is examined in section 3.3.

[2] The mean Pearson correlation between the first two independent variables is −.47; between the first and third independent variables, .14; and between the last two independent variables, −.04.

[3] The regression results are sensitive to the use of weighted least squares regression. Weighting the variables by the inverse of the current outstanding forecast, the inverse of price per share, and the inverse of the cross-sectional standard deviation of forecasts results in mean-adjusted R-squares of .28, .32, and .38, respectively. A divisor less than \$.20 is arbitrarily set equal to \$.20, and the variables are truncated by setting any measure less than −3 or greater than +3 to be equal to −3 and +3, respectively.

TABLE 2

Relation Between Changes in Analyst Forecasts and Other Information

	Mean Regression Results from Temporally Independent Regressions						
	Mean Coefficients				Mean-Adjusted R-square	Total Number of Revisions	Number of Regressions
	β_0	β_1	β_2	β_3			
All revisions	−.07[b]	.90[b]	.44[b]	.11[b]	.38	191,313	144
(t-statistic)[a]	(−21.44)	(48.69)	(35.62)	(15.86)			
By analyst reputation							
"All-Americans"	−.08[b]	.88[b]	.40[b]	.11[b]	.35	63,055	144
"Non-All-Americans"	−.07[b]	.92[b]	.48[b]	.10[b]	.40	128,258	144
Two independent variables	−.06[b]	.92[b]	.45[b]		.37	191,313	144
	−.07[b]	.63[b]		.12[b]	.26	191,313	144
	−.17[b]		.10[b]	.21[b]	.04	191,313	144
Using only forecast revisions from every 15th semimonthly period and restricting v to be less than 125	−.07[b]	.86[b]	.36[b]	.09[c]	.30	10,306	10

By number of analysts following the firm							
2–7 analysts	−.08[b]	.76[b]	.38[b]	.11[b]	.27	55,410	144
8–11 analysts	−.07[b]	.99[b]	.46[b]	.10[b]	.41	41,205	144
12–17 analysts	−.06[b]	1.02[b]	.52[b]	.09[b]	.47	51,180	144
18–39 analysts	−.07[b]	.96[b]	.57[b]	.09[b]	.51	43,518	144

Mean regression results for the relation between changes in analyst a's forecast of EPS and other publicly available information released since analyst a's current forecast and before analyst a's next forecast. Regressions use a sample of individual analyst forecast revisions from the 1980–85 period. The form of the estimated relation is:

$$(FRCST_{i,a,t} - FRCST_{i,a,t-v}) = \beta_0 + \beta_1(CONS_{i,t-1} - CONS_{i,t-v}) + \beta_2(CONS_{i,t-v} - FRCST_{i,a,t-v}) + \beta_3(FRCST_{i,a,t-v} * CR_{i,(t-v,t-1)}) + \epsilon_{i,a,t}$$

where $FRCST_{i,a,t-v}$ = the current outstanding forecast of EPS for firm i by analyst a, a forecast which is dated v days prior to day t.

$FRCST_{i,a,t}$ = forecasted EPS for firm i by analyst a on day t, a day on which the forecast is revised.

$CONS_{i,t-1}$ = mean consensus estimate of EPS for firm i on day $t-1$, excluding analyst a.

$CR_{i,(t-v,t-1)}$ = cumulative common stock return for firm i from day $t-v$ to day $t-1$.

[a] The t-statistics are calculated by using the estimated standard deviation of the estimated coefficients.
[b] The estimated coefficient is significantly different from zero at less than the .01 level.
[c] The estimated coefficient is significantly different from zero at less than the .05 level.

variables. The estimated coefficients imply that a $1.00 change in the mean forecast of other analysts since the date of an analyst's current forecast changes the expectation of the analyst's next forecast by $.90; an analyst's next forecast is expected to close the deviation between the mean forecast of other analysts and the analyst's forecast by approximately 44%; and cumulative share price gains or losses of 10% since an analyst's current forecast changes the expectation of the analyst's next forecast by 1.1%. The negative intercept suggests that analysts initially overestimate earnings, at least during 1980–85, and subsequently revise those forecasts downward by $.07 per revision, ceteris paribus.[4]

Table 2 also reports a comparison of the predictability of forecasts made by analysts on the *Institutional Investor* annual "All-American Research Team" with that of other analysts. Forecast revisions by analysts who are first-, second-, and third-team "All-Americans" in any year within 1981–85 are segregated from those of other analysts, and regressions are performed. Based on paired comparisons t-tests, where differences are computed semimonthly, the mean difference in β_1 is .037 (t-statistic = 1.48), the mean difference in β_2 is .081 (t-statistic = 3.68), and the mean difference in R-square is .049 (t-statistic = 3.26). Thus, ceteris paribus, the forecasts of "All-Americans" are less likely to "follow the crowd" and are less predictable than forecasts by other analysts.[5]

3.3 SENSITIVITY ANALYSES

The estimated coefficients are not sensitive to performing one regression with all 191,313 observations, although the significance levels of the t-statistics are higher. Using a single regression, the coefficients (t-statistics) for β_0, β_1, β_2, and β_3 are $-.08$ (-75.04), .86 (326.46), .41 (190.10), and .11 (54.24), respectively.

The regression results are sensitive to the exclusion of single independent variables. The results reported on table 2 suggest that most of the explanatory power of my model arises from the change in the consensus and the deviation of the forecast from the consensus forecast. The marginal explanatory power of price changes is very small.

Table 2 also reports regression results for subsamples restricted to be 14 semimonthly periods apart and revisions that are dated within 125 trading days of the date of the prior forecast. The semimonthly periods used end on the following dates: 1/15/80, 8/31/80, 4/15/81, 11/30/81, 7/

---

[4] A similar upward bias over the same period has been documented for *Value Line Investment Survey* forecasts by Abarbanell [1989] and in my own unpublished analysis of *IBES* mean consensus forecasts.

[5] The mean and median number of analysts following (1) firms followed by "All-Americans" and (2) firms followed by others are (1) 12.2 and 11 and (2) 12.3 and 11, respectively. Thus, the differences in model fit between "All-Americans" and other analysts are not associated with the differences in model fit reported in section 3.3 for analyst following.

15/82, 2/28/83, 10/15/83, 5/31/84, 1/15/85, and 8/31/85. Restricting the sample to these ten semimonthly periods ensures that any portion of the change in *FRCST* cannot be in more than one semimonthly period. The estimated coefficients are again significantly different from zero.

The results are somewhat sensitive to grouping revisions by the number of analysts with an outstanding forecast. As reported in table 2, the mean-adjusted *R*-square for the quartile of firms with the least analyst following is .27. The mean-adjusted *R*-squares for the remaining three quartiles are .41, .47, and .51, respectively.

The mean coefficients are relatively insensitive to performing regressions on a firm-by-firm basis before averaging. Regressions are performed for firms with at least 30 forecast revisions over the 1980–85 period. This procedure allows the intercept to vary across firms (see Murphy [1985]). There are 1,047 firms meeting this requirement, for which the mean coefficients for β_0, β_1, β_2, and β_3 are −.09, .80, .54, and .08, respectively. The individual firm coefficients are unbiased, but not independent; thus, t-statistics are not calculated. The mean-adjusted *R*-square for a random sample of 50 of these firms is .39. Thus, there is no apparent advantage to allowing the intercept to vary on a firm-by-firm basis.

The mean coefficient for $CR_{i,(t-v,t-1)}$ declines somewhat when cumulative abnormal returns, measured as market model residuals (e.g., Fama [1976]) and mean-adjusted returns (see Masulis [1978]), are substituted.[6] Using market model residuals and mean-adjusted returns results in estimated coefficients (*t*-statistics) for β_3 of 0.05 (8.62) and 0.05 (9.93), respectively. However, for both definitions, the explanatory power of abnormal returns is negligible, and the mean-adjusted *R*-square is again .38.

4. The Bias and Accuracy of Predicted Individual Analyst Forecasts

This section evaluates the predictive ability of the model using the following measure:

$$\text{"Updated" } PFE_{i,a,t} = (FRCST_{i,a,t} - E_{t-1}(FRCST_{i,a,t}))/FRCST_{i,a,t},$$

where $PFE_{i,a,t}$ is defined as the percentage forecast error[7] and:

$$E_{t-1}(FRCST_{i,a,t}) = FRCST_{i,a,t-v} + \hat{\beta}_0 + \hat{\beta}_1\,(CONS_{i,t-1} - CONS_{i,t-v})$$
$$+ \hat{\beta}_2\,(CONS_{i,t-v} - FRCST_{i,a,t-v}).$$

[6] For both definitions, parameters are estimated over event days +251 to +350, with at least 30 days of returns required for sample inclusion.

[7] The results discussed below are not sensitive to scaling the forecast error by price per share or the cross-sectional standard deviation of analyst forecasts at day $t - 1$. I used the procedures described in n. 3 for mitigating the small-denominator problem.

TABLE 3

Bias and Accuracy of Predicted Individual Analyst Forecasts

Percentage Forecast Error	Distribution of Percentage Forecast Error					
	Mean	Percentiles				
		10%	25%	50%	75%	90%
Using current outstanding forecast: $(FRCST_{i,a,t} - FRCST_{i,a,t-v})/FRCST_{i,a,t}$	−.160	−.364	−.119	−.036	.020	.074
Using "updated" forecast: $(FRCST_{i,a,t} - E_{t-1}(FRCST_{i,a,t})^{*})/FRCST_{i,a,t}$	−.045	−.161	−.038	.011	.054	.122

Absolute Percentage Forecast Error	Distribution of Absolute Percentage Forecase Error					
	Mean	Percentiles				
		10%	25%	50%	75%	90%
Using current outstanding forecast: $\lvert FRCST_{i,a,t} - FRCST_{i,a,t-v}\rvert/FRCST_{i,a,t}$	.235	.015	.029	.063	.150	.429
Using "updated" forecast: $\lvert FRCST_{i,a,t} - E_{t-1}(FRCST_{i,a,t})\rvert/FRCST_{i,a,t}$	.181	.008	.020	.048	.109	.289

Distribution of measures of the bias and accuracy in predicted individual analyst forecasts. Percentage forecast error, a measure of bias, and absolute percentage forecast error, a measure of accuracy, are calculated from forecast revisions dated within the 1981–85 period. The number of revisions is 174,700.

* $E_{t-1}(FRCST_{i,a,t})$ = the expected next forecast of *EPS* for firm i by analyst a as of day $t - 1$. This expectation uses publicly available information released since day $t - v$, the date of the analyst's current forecast. See table 2 for definitions of other variables.

The parameters β_0, β_1, and β_2 are estimated using data from the prior year.[8] Because of the low marginal explanatory power of past price changes noted on table 2, β_3 is not estimated or used.

As a benchmark, I use the following measure of the predictability of individual analyst forecast revisions.

$$\text{"Naive" } PFE_{i,a,t} = (FRCST_{i,a,t} - FRCST_{i,a,t-v})/FRCST_{i,a,t}.$$

"Naive" *PFE* conditions expectations of the next forecast on only the current outstanding forecast and is analogous to a random walk model.

Table 3 reports signed percentage forecast errors (measures of bias) and unsigned (absolute) percentage forecast errors (measures of accuracy). The distribution of "updated" *PFE* is more symmetrically distributed around zero and has smaller absolute values than "naive" *PFE*.[9]

[8] The regression results of the prior section are also relatively insensitive to the forecast year. Because prior year data are used for parameter estimation, there are no 1980 "updated" forecasts. This leaves 5 years or 120 semimonthly periods.

[9] Subtracting the regression intercept from each outstanding forecast and using the resulting number as the expected forecast results in a mean percentage forecast error of −12.0% and a mean absolute forecast error of 22.2%. Thus, the improvement in predictive ability from using "updated" forecasts is not simply due to the intercept term.

I used paired comparisons t-tests to evaluate the significance of the differences in bias and accuracy. The differences in bias ("naive" PFE minus "updated" PFE) and accuracy (absolute "naive" PFE minus absolute "updated" PFE) are computed at the individual analyst level, and a mean difference is computed by semimonthly period. Significance is determined by dividing the mean of the semimonthly mean differences by its standard error. Aggregated in this manner, the mean difference in bias is -11.3% (t-statistic $= -32.17$) and the mean difference in accuracy is 5.2% (t-statistic $= 31.55$). Thus, "updated" forecasts are less biased and more accurate predictors of future forecasts than the analyst's current forecast.

5. Conclusions

My model predicts an individual analyst's next EPS forecast by updating his current forecast for subsequent information. "Updated" forecasts from this model are less biased and more accurate predictors of future forecasts than the analyst's current forecast. Possible extensions of this line of research include examining whether or not "updated" forecasts are better predictors of future reported earnings; using "updated" forecasts as measures of market expectations; and using the dispersion of "updated" forecasts as measures of earnings uncertainty.

REFERENCES

ABARBANELL, J. "Information Acquisition by Financial Analysts in Response to Security Price Changes." Ph.D. dissertation, University of Pennsylvania, 1989.

FAMA, E. F. *Foundations of Finance.* New York: Basic Books, 1976.

——, AND J. D. MACBETH. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* (May/June 1973): 607–36.

MASULIS, R. W. "The Effects of Capital Structure Change on Security Prices." Ph.D. dissertation, University of Chicago, 1978.

MURPHY, K. J. "Corporate Performance and Managerial Remuneration." *Journal of Accounting and Economics* (April 1985): 11–42.