Division of Research
School of Business Administration

January 1990

FORECAST ACCURACY OF INDIVIDUAL ANALYSTS IN NINE INDUSTRIES

Working Paper #612R

Patricia C. O'Brien
The University of Michigan

The author is grateful to Yum Khim Tan for her capable research assistance, and to semilnar participants at the seventh annual Accounting Research Workshop at Columbia University, at Harvard University, the University of Michigan, and the University of Pennsylvania for their helpful comments. Special thanks are due to Dale Berman , Vic Bernard, Linda DeAngelo, John Elliot, Karen Waldemar and an anonymous referee for helpful conversations and/or comments. The author retains responsibility for all errors and omissions. This research was completed while the author was a faculty member at MIT.

1. Introduction

The purpose of this paper is to investigate whether financial analysts with superior earnings forecasting ability can be distinguished on the basis of *ex post* forecast accuracy. I explore the question by estimating and comparing average accuracy across individuals, and by considering whether the observed distribution of analyst forecast accuracies differs from the distribution expected if their relative performances each year were purely random. Overall, I do not find systematic differences in forecast accuracy across individuals.

Financial press coverage suggests there are superior financial analysts. For example, *Institutional Investor*'s annual "All American Research Team" includes analysts rated by money managers as superior on a variety of criteria, including earnings forecasting, ability to pick stocks, and the quality of written reports. Clearly, financial analyst services other than forecast accuracy are valued by their clients. I focus on only one activity, earnings forecasting, for two reasons. First, forecast data are available, quantitative, and can be evaluated against observable earnings outcomes. Services such as insightful, well-written research reports are harder to evaluate quantitatively. Second, academic use of analyst forecasts as earnings expectations data in capital markets empirical research is now widespread,[1] although many characteristics of the underlying analyst forecast data are not well understood.

Analyst services are known to have useful *aggregate* characteristics. For example, Dimson and Marsh [1984], Elton, Gruber, and Grossman [1986] and Brown, Richardson and Trczinka [1988] find that investment strategies constructed from aggregate analyst forecasts of stock prices or returns can be used profitably by investors. Similarly, Brown and Rozeff [1978] and many others find that aggregate analyst earnings forecasts are more accurate than forecasts from time-series models. However, studies that attempt to distinguish persistent differences in stock-picking ability across individuals (Dimson and Marsh [1984], Elton, Gruber, and Grossman [1986]) find none.

In O'Brien [1985], I found no evidence that some brokerage firms consistently produced more accurate earnings forecasts. This result, however, is insufficient to address the question of individual performance, since each firm employs many analysts who may differ in ability. Preliminary data analysis on individual analyst forecasts included in O'Brien [1987] also indicated no evidence of consistent differential ability. In this paper, I extend the earlier work by (1) examining a more extensive sample of individual analyst forecasts, (2) devising non-parametric tests that demonstrate the robustness of the results, (3) acknowledging analysts' industry specialization in the sample design, and (4) exploring the effects of sample selection and forecast age on the results.

The primary use of analyst earnings forecasts in academic work is to provide a proxy for the "market" expectation of a future earnings realization. Either forecast aggregations, such as the mean or median, or the most current forecast available are often used for this purpose. All of these proxies assume analysts have approximately the same forecasting ability, so the identity of the individual is ignored in defining the consensus.

If some analysts are consistently superior (or inferior) forecasters, then knowledge of ability can be used to improve the accuracy of the consensus measure. For example, under the criterion of quadratic loss, precision-weighted averages provide optimal forecasts if analysts update simultaneously and differ in ability.[2] If analysts update at different times and do not differ in ability, then under mild assumptions the most recent forecast available may be best. However, if analysts differ systematically in forecasting ability, there will be a tradeoff between the age of the forecast and the ability of the forecaster. In either of the above contexts, improvements in forecasts are possible only if analysts have measurable and persistent differences in accuracy.

I examine a sample of forecasts for firms in nine different 2-digit SIC industries over the period 1975 to 1981. The analysis proceeds along two lines. First, I estimate a fixed effects regression model to test whether analysts are heterogeneous in forecast accuracy. Since individual analysts predict earnings for varying portfolios of firms, the model controls for firm-specific and year-specific differences in predictability. In each of the nine industries, the fixed effects model fails to reject at conventional levels the hypothesis that analysts are homogeneous, conditional on the firms and years they forecast.

Second, I rank-order analysts in quartiles each year, and compare the observed distribution of analysts' average ranks with the distribution expected if all analysts were alike and each year were an independent observation. In all nine industries, these non-parametric tests fail to reject at conventional levels the hypothesis that the observed distribution is identical to the expected distribution. While the non-parametric test is partly motivated by severe non-normality in the fixed-effects model error terms, the approach also provides an intuitively appealing way to make performance comparisons. These comparisons often depend on selecting "good" performers, ex post, from among large numbers of individuals. Ex ante, however, some positive number of analysts, which depends on the number of individuals in the game, is expected to emerge as "good" performers.[3] The tests I construct consider these expectations explicitly.

The remainder of the paper is organized as follows. In section 2, I describe the sample selection process and some characteristics of the sample of analyst forecasts. Section 3 contains the fixed effects model and the results of the parametric tests. In section 4 I describe the construction of the non-parametric tests using a single industry as an example and present results for all nine industries. Section 5 contains an analysis of the effects of the sample selection criteria on the statistical tests. In section 6, I explore whether the results differ if they are conditioned on the age of the forecast. Section 7 is a discussion of some implications of these results.

2. Description of Sample

The nine-industry sample of forecasts used in this paper is selected from a database of individual forecasts from Institutional Brokers Estimate System (hereafter, I/B/E/S), produced by Lynch, Jones and Ryan. The individual forecasts are the detail data from which monthly summaries were computed in the period July 1975 to September 1982. The summary data are sold to I/B/E/S' clients and have been analyzed by Brown, Foster and Noreen [1985]. The identities of the analysts and brokerage houses included in my database are encoded numerically, and I do not have access to the names underlying the codes.

I study just nine industries to allow examination of within-industry variation, and to reduce the problem to a manageable size. This restriction is consistent with observed behavior, since analysts tend to concentrate within industries and specialize in particular types of firms. Financial press evaluations of analysts reflect this specialization. For example, Institutional Investor ranks analysts within industries in its annual survey. Analysts' concentration within industries is evident in the sample used here: only 12 of 404 analysts included in the study appear in more than one of the nine industries.

The sample firms have December year ends, at least one forecast in the I/B/E/S database in each year 1975-81, and annual earnings available on COMPUSTAT. The December year end criterion provides a basis for estimating industry-wide (time-period-specific) events comparably across firms. For each firm and year, I select the most recent forecast from each analyst issued at least 120 trading days before the annual earnings announcement. The mid-year forecast horizon, admittedly arbitrary, provides further comparability across firms and years.[4] This selection criterion is discussed further at the end of this section.

The earnings realizations are primary earnings per share (hereafter, EPS) before extraordinary items.[5] Forecasts for fully-diluted EPS are identified in the database. I convert these forecasts to primary using the ratio of primary to fully-diluted EPS for that firm and year from COMPUSTAT. When a stock split or dividend is announced after a forecast was made, I adjust the share basis of the EPS forecast using distribution data from the CRSP Master file. Industry affiliation is defined by two-digit Standard Industrial Classification (SIC) codes, as reported in the 1982 COMPUSTAT tapes.

Some characteristics of the data available in the initial sample appear in the first row of Table 1. The second row of Table 1 indicates that approximately 5% of the forecasts are dropped because the analyst who made the forecast is not identified by a code number in the database. These deletions are primarily attributable to five brokerage houses.

For this study, I select the nine industries with the most data available. Choosing industries with the most data helps ensure a reasonable sample remains after the data sufficiency criteria are applied. Some characteristics of the data available in these industries are in the third line of Table 1. The nine industries contain nearly two-thirds of the forecasts, more than two-thirds of the analysts, and more than one-half the firms in the larger dataset.

Finally, to ensure reliable estimation for each analyst, the analysts included in the study have both ten or more forecasts available for firms in the industry and forecasts in at least three of the seven years 1975-81. Requiring a minimum number of years supports tests to distinguish between persistent superior forecast accuracy and a single chance occurrence.

Industry-by-industry characteristics of the sample are reported in Table 2. The sample includes service, financial, and manufacturing firms. While the data sufficiency requirements, at least ten forecasts appearing in at least three years, eliminate more than 75% of the analysts available in I/B/E/S, the remaining analysts produce a disproportionately large share of the forecasts in each industry.

The reduction in the number of analysts included in the sample may reflect features of the market for financial analysts and of the database. An analyst who changes brokerage houses within the sample period cannot be tracked to the new job even if he or she remains in the database, but instead is identified as a new analyst in the data codes. Thus, the three-year requirement may eliminate analysts with the greatest job-mobility from the sample. In section 5, I investigate the possibility that this aspect of sample selection may influence the results. The evidence suggests that it does not.

I fix the same minimum horizon of 120 trading days for all firms and years to provide comparability, because forecast accuracy generally improves as the horizon decreases (Brown, Foster and Noreen [1985], O'Brien [1988]). I obtain the most recent forecast from each analyst issued prior to this horizon date, by examining the analyst's forecast dates. On any fixed date, the available forecasts for any firm and year vary in how recently they were produced. Because the choice of how frequently to update EPS estimates may be related to forecasting ability, I do not a priori eliminate variation in forecast age, but later investigate the effects of this variation on the results.

Table 3 contains descriptive information about the distribution of forecast ages in the sample, by industry. Forecast age is defined as the number of trading days between the analyst's forecast date and the horizon date on which the whole set of forecasts is examined. On the arbitrary horizon date of the study, 120 trading days prior to the earnings announcement, approximately 50% of the forecasts are less than two months old in most industries. The mean age is closer to three months in most industries because of a relatively long right tail to the distribution. The results in section 6 suggest that, while the age of the forecast is an important determinant of forecast accuracy in some industries, including age in the analysis does not alter the results on heterogeneity among analysts.

3. Parametric Tests of Forecast Accuracy

The forecast accuracy metric I use to compare analysts is average absolute forecast error,[6] where absolute forecast error is defined:

$$| e_{ijt} | = | A_{jt} - F_{ijt} | \quad . \qquad (1)$$

In (1), A_{jt} denotes actual EPS for firm j in year t, and F_{ijt} denotes the forecast of A_{jt} from analyst i made at least 120 trading days prior to the annual earnings announcement.

Because analyst forecast data are extremely unbalanced,[7] I use the fixed effects model given in (2), essentially regression on a matrix of zeros and ones, to estimate average accuracy.

$$| e_{ijt} | = \mu_i + \delta_j + \gamma_t + \eta_{ijt} \qquad (2)$$

In (2), μ_i, δ_j and γ_t are analyst, firm and year effects, respectively. The estimates of μ_i are interpreted as the average accuracy for analyst i, conditional on the firms and years in his or her sample portfolio of predictions. Analysts may choose the firms for which they issue forecasts each year. As long as firms and years differ in average forecast accuracy, this endogenous selection is important for statistical inferences about relative accuracy of analysts from unbalanced data. Equation (2) controls for average firm and year effects arising from this selection.

The estimated firm and year effects measure (conditional) average accuracy along the firm and year dimensions. An analyst's average accuracy μ_i is then measured as a deviation from these conditional means. This adjustment for differences in predictability across years and across firms helps to avoid attributing differences to analysts who, by chance or design, have forecasts concentrated in certain years or certain firms.[8]

Differences across years in average accuracy will occur if the unanticipated events in the last half of the year, i.e. over the forecast horizon in this study, have a greater average impact on corporate earnings in some years than in others. Differences across firms in average accuracy will occur if the earnings of some firms are harder to predict, because of differences in such characteristics as earnings volatility, company disclosure policies and sensitivity of earnings to other observable data, such as input and output prices.[9]

Table 4 contains evidence on analyst, firm and year effects for each of the nine sample industries. Based on F-tests on groups of coefficients, both firm and year effects are significant at the .0001 level in all industries, except for year effects in the Chemicals and Allied Products industry. Here year effects contribute significant explanatory power at the .06 level. In contrast, analyst effects do not contribute significant explanatory power to the model. Equivalently, the hypothesis that analysts are homogeneous in average forecast accuracy, conditional on firm and year effects, cannot be rejected at conventional significance levels for any sample industry. Thus, there is no evidence here that analysts differ systematically in forecast accuracy.

The F-tests reported in Table 4 depend on the assumption that the error terms are normally distributed. Although the dependent variable, absolute forecast error, is clearly not normally distributed, the residuals are deviations from conditional means, and so can take both positive and negative values. The residuals from regression equation (2), however, are highly right-skewed and highly leptokurtic. In all industries, the null hypothesis that the residuals come from a normal distribution is rejected at levels smaller than .01.[10] The extremely fat tails of the residual distributions indicate too-frequent rejection. This may weaken the conclusions about firm and year effects, but would not alter the result that analysts are not statistically distinguishable in forecast accuracy. Extreme skewness, on the other hand, has ambiguous implications for inferences. In the next section, I describe the results of non-parametric tests, which support the same conclusion without the requirement of normally-distributed residuals.

4. Non-parametric Tests

The unbalanced nature of analyst forecast data creates two problems in the formulation of non-parametric tests. First, differences in predictability across firms and years make it unclear how to rank the accuracy of two forecasts from different years or different firms. Second, each firm and year has a different number of analyst forecasts, so aggregating ranks across firms for a given analyst can create uninterpretable results. For example, the rank "third most accurate" should be interpreted differently if there are 20 analysts ranked than if there are 4. I have attempted to address both these concerns in the test devised below.

If analysts are alike, and differences in observed performance are all due to non-replicable "luck" that is uncorrelated across analysts, the probability of observing a particular string of luck in a sequence of T observations is determined by a multinomial distribution. I use this idea to construct an expected distribution of analysts' performance rankings, and compare the observed to the expected distribution using a Kolmogorov-Smirnov test statistic. For expository reasons, I describe the construction of the expected distribution in detail for a single industry (SIC 26, Paper and allied products), and then discuss the results of the test for the entire nine-industry sample.

The first three columns of Table 5 display the identification code for each of the 29 analysts in the Paper industry, along with the number of years in the sample and the average rank. To compute average rank, I first eliminate firm and year effects from the absolute forecast errors by means of the following model:

$$| e_{ijt} | = \delta_j + \gamma_t + \epsilon_{ijt} \quad . \tag{3}$$

Regardless of non-normality, the regression residuals from (3) are by construction purged of firm- and year-specific average predictability. For each analyst, I compute the annual average of the residuals ϵ_{ijt} across firms, for each year the analyst is in the sample. The average residuals are interpreted as yearly estimates of average accuracy, net of firm and year effects.[11]

Next, analysts' annual average accuracies are ranked in quartiles within years. I use quartiles, rather than the complete rank order, because different numbers of analysts predict EPS each year. Since there are more than four analysts in each industry in each year, quartiles do not create uninterpretable aggregations, although information on analysts' relative positions within the quartiles is lost.

Finally, I compute an average rank for each analyst over all years the analyst is in the sample. The result is a distribution of average ranks, displayed in the third column of Table 5 for analysts in the Paper industry. I compare this empirical distribution to the distribution of average ranks that would be expected by chance.

The null hypothesis for the non-parametric tests is based on the assumptions that each analyst has probability .25 of falling in each of the four quartiles in each year, and that each year is an independent observation. An analyst with forecasts in T different years can have 3T + 1 possible average ranks, or outcomes. There are 4^T possible sequences of length T that lead to these outcomes. The probability of any particular outcome under the null hypothesis is the proportion of the 4^T possible sequences that result in that outcome. The expected frequency is the probability multiplied by the sample size N_T.[12]

I calculate expected frequencies at the (3T + 1) possible outcomes, for T = 3, 4, 5, 6, and 7. In the Paper industry sample, the corresponding numbers of analysts are N_T = 10, 10, 5, 1 and 3, respectively. I aggregate expected frequencies across years into a single distribution of expected frequencies of average ranks, conditional on the numbers of analysts in the sample. The distribution of expected frequencies is converted to a density by dividing each expected frequency by 29, the number of analysts in all years for the Paper industry.

The empirical density assigns a probability mass of (1/29) or 0.03448 to each sample observation. The difference, at each sample point, between the empirical density and the expected density under the null hypothesis, constructed above, provides the basis for the test statistic:[13]

$$KS = [MN/(M+N)]^{.5} \quad D^*_{MN} \quad . \qquad (4)$$

D^*_{MN} is defined as the maximum distance, over all sample points, between the empirical density and the expected density. N denotes the sample size, in this case 29. M denotes the number of possible outcomes, or the number of points in the expected density that could have positive probability mass. With quartile ranking and T = 3, ..., 7, M equals 55.

The absolute difference between the empirical and expected densities at each sample point appears in the last column of Table 5. The maximum absolute difference is .09, which results in a test statistic of .38. From the limiting distribution for the KS test statistic, the probability of observing a value of .38 or greater is more than .99. Therefore, the observed distribution of average ranks for analysts in the Paper industry does not differ from the distribution that would be expected if each analyst had a probability .25 of falling in each quartile in each year.

Table 6 contains results of the Kolmogorov-Smirnov test for all sample industries. In all nine industries, the distribution of analysts' forecast accuracies is indistinguishable at the 18% level from the distribution expected if all analysts were identical, with independent equal chances of falling in each of the four quartiles in each year. Thus, this non-parametric test confirms the results of the parametric test, that no strong evidence of systematic differential forecasting ability among analysts appears in this sample.

5. Investigation of the Effects of Sample Selection

It is evident from Tables 2 and 3 that the sample selection requirements, at least 10 forecasts in at least three years, dramatically reduce the number of analysts included in the study. In this section, I investigate whether sample selection drives the results by eliminating the consistent heterogeneity in the population, leaving a sample of homogeneous individuals.

By definition, the analysts excluded by the sample selection criteria have relatively sparse data, so the comparisons below concentrate on group-wide, not individual, characteristics. In what follows, I refer to two groups: the Sample, or group of analysts included in the analysis in sections 3 and 4, and the Excluded group, those eliminated by the selection criteria.

The sample selection criteria are applied consistently to all nine industries, and the apparent homogeneity of analysts in forecast accuracy appears consistently in all nine industries. Therefore, a compelling explanation that the sample selection eliminates heterogeneity requires consistent differences across industries as well. In what follows, I find some differences in particular industries, but it is difficult to draw a coherent picture to suggest that sample selection drives the results of the study.

The first test compares the Sample and Excluded groups in mean conditional forecast accuracy. The estimated equations are:

$$| e_{ijt}| = \delta_j + \gamma_t + v_{ijt} \quad , \tag{5}$$

and

$$\hat{\mu}_S = \Sigma_{i\epsilon S} \hat{v}_{ijt} \quad , \quad \hat{\mu}_E = \Sigma_{i\epsilon E} \hat{v}_{ijt} \quad . \tag{6}$$

In equation (5), δ and γ are defined to be firm and year effects, as in section 3. The symbol (^) in equation (6) denotes estimated values. Equations (5) and (6) differ from equation (2) in two ways. First, only firm and year effects are estimated in equation (5). Second, only group mean effects for the Sample and Excluded analysts, μ_S and μ_E, are estimated. The null hypothesis of the test is that the two group means are identical.

When equation (5) is estimated using the pooled Sample and Excluded forecasts, the hypothesis of homogeneity across firms or years in forecast accuracy is rejected in all nine industries, as it was for equation (2). Table 7 contains the results of a Student's-t test on the difference in average accuracy, controlling for firm and year effects, between the Sample and Excluded group means of residuals, computed as described in equation (6). The difference in mean forecast accuracy between the two groups is indistinguishable at the .05 level in eight of the nine industries. The exception is the Electric, Gas and Sanitary Services industry, SIC 49, where mean accuracy is greater in the Sample than in the Excluded group.

While most industries exhibit no difference in _mean_ accuracy between the two subsamples, sample selection may truncate accuracy in both tails of the distribution of forecast errors. For example, perhaps the analysts with the most job mobility are those with the best and worst performance. The best performers may be lured away by competitors, and the worst performers may cease to be employed as analysts. If these two effects operate simultaneously, the Sample and Excluded distributions could have the same mean, but the latter should have fatter tails.

Non-normality in the data impedes construction of a statistical test for differences in the fourth moment. Instead, I inspected the fractiles of the two distributions, Sample and Excluded, for evidence that the Sample has censored tails. The lower tail of the distribution of residuals ϵ_{ijt} from equation (5) contains observations below the conditional mean, therefore observations of the greatest accuracy. The upper tail of the distribution contains the least accurate observations. The Sample has a fatter lower tail in seven of the nine industries, and a fatter upper tail in four of the nine. This result is inconsistent with the view that the selection criteria systematically eliminate heterogeneity.

A caveat on this analysis is that the comparisons concern forecast-by-forecast distributions, not analyst-by-analyst distributions. For the two to yield the same conclusions, the best-performing analysts must produce forecasts of greater accuracy, or in far greater numbers, than the best forecasts from other analysts (and symmetrically for the worst-performing analysts). However, as mentioned above, an imperfect analysis is unavoidable: the Excluded analysts, by definition, have insufficient data for reliable individual estimation.

If sample selection eliminated heterogeneity in the sample for all industries, then one should expect similar differences between the Sample and Excluded groups in all nine industries. However, in only one industry is the average forecast accuracy different between the Sample and Excluded groups, and there is no evidence that the Excluded group systematically has fatter tails than the Sample group. Overall, the evidence does not support the hypothesis that the sample selection criteria drive the results of Sections 3 and 4.

6. The Effects of Forecast Age

Since analysts do not issue forecasts at prescribed times, the set of forecasts available at any particular point in time for a given firm and year will vary in age. Forecasts generally become more accurate as the realization draws nearer, as analysts incorporate new information into their forecasts. Therefore, a potential implication of the variation in forecast ages is that differences in forecasting ability may be confounded by differences in forecast age. In this section, I investigate whether the results are sensitive to forecast age.

Some of the variation in forecast ages may be related to forecasting ability. For example, analysts who update their forecasts more frequently will on average have forecasts that are more current than analysts who seldom review and revise their forecasts. This is the reason for not selecting an initial sample of moderately current forecasts: doing so might eliminate important heterogeneity from the sample.[14]

I explore the possible confounding effects of age on the earlier results in two ways. First, I add age categories to the regression model to see whether analysts appear heterogeneous when forecast accuracy is conditioned on the age of the forecasts. Second, I censor the sample to remove older forecasts, to look for evidence of heterogeneity among only relatively current forecasts.

I use categories to incorporate age into the regression analysis, because the precise functional form of the relation between accuracy and age is unknown. The relation described above implies only that more recent forecasts should be more accurate than less recent forecasts for the same firm and year. Categories allow for some flexibility in the form of the relation.[15] I divide the forecasts into four categories: (1) less than 22 trading days (approximately one month) old, (2) between 23 and 44 trading days old, (3) between 45 and 66 trading days old, and (4) more than 66 trading days old. The regression model I use is a modification of equation (2):

$$| e_{ijt} | = \mu_i + \delta_j + \gamma_t + \alpha_a + w_{ijt} \quad . \qquad (7)$$

In equation (7), the coefficient α_a is the average forecast accuracy in category a, and the index a ranges over the four categories described above. The other coefficients are defined as in equation (2).

Table 8 contains the results from estimating equation (7). Forecast age varies in significance from industry to industry, perhaps reflecting the difficulty of specifying the functional form of the relation between age and accuracy. However, the purpose of equation (7) is to examine whether inclusion of information about the timeliness of the forecasts reveals heterogeneity among analysts. Even in industries where forecast age contributes the most explanatory power, analysts continue to appear homogeneous, and the results are remarkably similar to those in Table 4.

I investigate the influence of forecast age further by censoring the sample to eliminate all forecasts more than 130 trading days old.[16] If the older forecasts are simply irrelevant, then discarding them is appropriate. Combined with the horizon of 120 trading days, this censoring point results in elimination of forecasts made more than about one year before the EPS announcement. When equation (2) is estimated on the censored sample the conclusions are maintained: there is significant heterogeneity in forecast accuracy among firms and among years, but not among analysts.

Because analyst forecasts are non-synchronous, more recent forecasts may incorporate more information than less recent forecasts. Therefore, heterogeneity among analysts could be confounded by differences in forecast ages when the forecasts are simply viewed as panel data. The above tests confirm that forecast age is sometimes an important determinant of accuracy, though the result varies widely across industries. However, it does not appear that this feature has confounded the result that analysts are homogeneous.

7. Discussion and Summary

In both parametric and non-parametric tests, individual analysts fail to exhibit consistent differences in forecasting ability. One implication of this result is that attempts to improve the accuracy of consensus measures by weighting forecasts based on prior precision are unlikely to succeed. However, the result also raises a question: why are so many analysts engaged in forecasting earnings?

An explanation may lie in the difference between the academic's historical perspective and the investor's interest in future events. The research question addressed by this study is motivated by a common academic use of analyst forecast data: on a given date, to determine a proxy for the market expectation of a firm's earnings. That is, academic use of forecast data often concerns properties of the point estimate of earnings, given the cross-section available. In contrast, investors' use of analyst forecasts of earnings presumably requires both accuracy and timely incorporation of new information. It may be that other analysts mimic updated point estimates announced by informed analysts. While a point estimate can be mimicked, presumably it is more difficult to mimic the timing of informed updating.

Under this description of analyst forecast updating, it is not surprising that *ex post* cross-sections of forecasts taken at an arbitrary time, such as those used in many academic studies, do not display consistent differences in analysts' abilities to make point estimates. In the scenario described above, analysts compete on timely incorporation of new information, not on accuracy *per se*. Although the timing issue does not affect the measurement of market expectations at any time from *ex post* data, the issue is relevant for studies that attempt to determine the timing of *changes* in market earnings expectations. If most analysts simply follow one or a few analysts who are informed and update their forecasts quickly in response to new information, this could confound attempts to determine when the "market" expectation of earnings changed.

REFERENCES

Agnew, C. (1985). "Bayesian Consensus Forecasts of Macroeconomic Variables." Journal of Forecasting 4: 336-376.

Ashton, A. and R. Ashton (1985). "Aggregating Subjective Forecasts: Some Empirical Results." Management Science 31: 1499-1508.

Bamber, L. (1986). "The Information Content of Annual Earnings Releases: A Trading Volume Approach." Journal of Accounting Research 24: 40-57.

Brown, L., G. Richardson, and C. Trzcinka (1988). "Asset Pricing Models, Earnings Forecasts, and Expected Returns." Working paper.

Brown, L. and M. Rozeff (1978). "The Superiority of Analysts' Forecasts as Measures of Expectations: Evidence from Earnings." Journal of Finance 33: 1-6.

Brown, P., G. Foster, and E. Noreen (1985). Security Analyst Multi-Year Earnings Forecasts and the Capital Market. American Accouniting Association, Studies in Accounting Research #21.

Clemen, R. and R. Winkler (1986). "Combining Economic Forecasts." Journal of Business and Economic Statistics 4: 39-46.

DeGroot, M. (1986). Probability and Statistics (2nd ed.). Addison-Wesley Publishing Company.

Dimson, E., and P. Marsh (1984). "An Analysis of Brokers' and Analysts' Unpublished Forecasts of UK Stock Returns." Journal of Finance 39: 1257-1292.

Elton, E., M. Gruber, and S. Grossman (1986). "Discrete Expectational Data and Portfolio Performance." Journal of Finance 41: 699-714.

Gregory, N. (1983). "Testing an Aggressive Investment Strategy Using Value Line Ranks: A Comment." Journal of Finance 38: 257.

Hoskin, R., J. Hughes, and W. Ricks (1986). "Evidence on the Incremental Information Content of Additional Firm Disclosures Made Concurrently with Earnings." Journal of Accounting Research 24: 1-32.

Kang, H. (1986). "Unstable Weights in the Combination of Forecasts." Management Science 31: 1492-1498.

Newbold, P. and C. Granger (1974). "Experience with Forecasting Univariate Time Series and the Combination of Forecasts." Journal of the Royal Statistical Society - Series A 137: 136-146.

O'Brien, P. (1985). "An Empirical Analysis of Analysts' Forecasts of Earnings Per Share." unpublished PhD dissertation, The University of Chicago.

O'Brien, P. (1987). "Individual Forecasting Ability." Managerial Finance 13: 3-9.

O'Brien, P. (1988). "Analysts' Forecasts as Earnings Expectations." Journal of Accounting and Economics 10: 53-83.

Patell, J., and M. Wolfson (1984). "The Intraday Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements." Journal of Financial Economics 13: 223-252.

Pound, J. (1987). "The Informational Effects of Takeover Bids and Managerial Resistance." Working paper, Kennedy School of Government, Harvard University.

Ricks, W. and J. Hughes (1985). "Market Reactions to a Non-Discretionary Accounting Change: The Case of Long-Term Investments." The Accounting Review 60: 33-52.

Searle, S.R. (1971). Linear Models. New York: John Wiley & Sons.

ENDNOTES

1. Some recent examples are: Patell and Wolfson [1984], Ricks and Hughes [1985], Bamber [1986], Hoskin, Hughes and Ricks [1986], and Pound [1987].

2. See, for example, Newbold and Granger [1974], Ashton and Ashton [1985], Agnew [1985], and Clemen and Winkler [1986]. For a discussion of the difficulties encountered when precisions are not stable, see Kang [1986].

3. Gregory [1983] discusses this idea in his comment on *ex post* discoveries of successful investment strategies using Value Line.

4. I have replicated the parametric test results using forecast horizons of 5, 60, 180, and 240 trading days without altering the substantive conclusions of the paper.

5. In private conversation, Mr. Dale Berman and Ms. Karen Waldemar of I/B/E/S have indicated that analysts try to predict earnings *from continuing operations*. This definition differs from primary EPS by any "above the line" non-recurring

items. I have replicated the parametric tests of this study, defining actual earnings as primary EPS before "Special Items" (COMPUSTAT annual data item 17), net of tax effects. The results are not sensitive to this adjustment. These results and others that are not reported in the tables are available from the author.

6. Average squared forecast error is another commonly-used accuracy criterion. However, in these data, use of squared forecast errors results in extremely skewed and fat-tailed residual distributions, amplifying problems of non-normality that may affect statistical inferences. This point is discussed in greater detail at the end of section 3.

7. Balanced data have equal numbers of observations in each cell of the design matrix. The fact that the population of analyst forecasts is unbalanced complicates sample selection and testing. This issue is discussed in detail in O'Brien [1987].

8. See Searle [1971], especially pp. 135-49, for a discussion of the bias that can arise in unbalanced data from aggregating across heterogeneous subgroups.

9. In modelling analysts' forecast errors, it is common to scale the errors by a

firm-specific denominator to control for differences in predictability across firms. Since equation (2) is designed explicitly to control for this heterogeneity, it is less important here to do so by transformation of the forecast errors. Moreover, these transformations may not make firms homogeneous. For example, when equation (2) is estimated with the dependent variable defined as the absolute value of (forecast error divided by the previous year's actual EPS), the hypothesis of homogeneity across firms is rejected at the .001 level or better in all nine industries. Similar results are obtained when the dependent variable is defined as absolute forecast errors scaled by a five-year average of the absolute value of EPS changes. In all cases, the major conclusion of this paper is unaltered.

10. I attempted to obtain residual distributions closer to the normal using the log and the square root transformations, both of which preserve the ordering of accuracies. In both cases, the distributions of residuals from regressions on transformed data were left-skewed and severely leptokurtic, in contrast to the untransformed residuals, which are right-skewed and severely leptokurtic. Results were not substantively altered by these transformations.

11. With unbalanced data, firm effects are not orthogonal to year effects, so both year and firm effects are estimated in equation (3) to estimate the firm effects without bias. It would not otherwise be necessary to purge year effects,

since analysts are ranked within years. Obviously, the year effect is constant within each year and so does not affect the ranking.

12. Below are example expected frequency computations for T = 3 , N_3 = 10.

average rank (a)	probability under null (b)	expected frequency (c)
1.00	0.016	0.16
1.33	0.047	0.47
1.67	0.094	0.94
2.00	0.156	1.56
2.33	0.188	1.88
2.67	0.188	1.88
3.00	0.156	1.56
3.33	0.094	0.94
3.67	0.047	0.47
4.00	0.016	0.16

(a). Analysts are ranked in quartiles each year. After three years, there are ten possible average ranks: 1.00, 1.33, ... , 4.00.

(b). The number of distinct sets of quartile ranks after three years is 4^3 = 64. Under the null hypothesis that each quartile is equally likely each year, the probability of an average rank is the proportion of the 64 sets of ranks with that average.

(c). The expected frequency is the probability times the sample size, N_3 = 10.

13. See DeGroot [1986] pp. 552-9 for a description of Kolmogorov-Smirnov tests.

14. In O'Brien [1987], I describe structural features of the dataset which could favor eliminating extremely old forecasts. This is accomplished here in tests on age-censored data, described at the end of section 6.

15. I examined two other specifications for forecast age in the model: (1) including age as a continuous variable, and (2) including the natural logarithm of age. The results vary in some particulars. For example, the industries in which age contributes significant explanatory power to the model depend upon the specification. The natural log specification provides the "best fit" in the sense that age provides significant explanatory power in the largest number of industries. However, the analyst results are unchanged in all specifications.

16. I repeated this analysis, censoring the sample successively at 100, 66, and 44 trading days, without altering the conclusions.

TABLE 1

Sample Selection Criteria and Their Effects on Sample Characteristics

	Number of forecasts	Number of Observations			
		Industries[1]	Firms	Brokers[2]	Analysts[3]
Initial sample[4]	40805	51	457	122	--
Analyst identified	38611	51	457	117	3041
Nine-Industry subsample[5]	24544	9	258	117	2104
Final sample[6]	12746	9	247	57	404

[1] Industries are defined by 2-digit SIC codes from the 1982 COMPUSTAT.

[2] "Brokers" refers to brokerage houses included in I/B/E/S.

[3] "Analysts" refers to individuals producing forecasts included in I/B/E/S.

[4] The initial sample contains the most recent forecast from each analyst in I/B/E/S made at least 120 trading days prior to the announcement of annual EPS, for firms with December year ends, with at least one forecast in each year 1975-81, and with annual EPS available from COMPUSTAT.

[5] The nine industries are those with the most forecasts from identified analysts.

[6] The final sample includes only analysts with at least ten forecasts in the industry, and with forecasts in at least three of the years 1975-81.

TABLE 2

The Effects of Imposing Data-Sufficiency Requirements within Industry for Nine Industries[1]

SIC	Industry Name		Before Selection	Final Sample[2]	% of Original in Final
26	Paper and allied products	forecasts	1871	826	44.1%
		analysts	362	29	8.0
		firms	17	15	88.2
28	Chemicals and allied products	forecasts	5359	2900	54.1%
		analysts	523	104	19.9
		firms	43	43	100.0
29	Petroleum refining and related industries	forecasts	2958	1775	60.0%
		analysts	282	50	17.7
		firms	21	20	95.2
33	Primary metal industries	forecasts	1054	524	49.7%
		analysts	153	26	17.0
		firms	12	11	91.7
35	Machinery, except electrical	forecasts	2902	1106	38.1%
		analysts	635	63	9.9
		firms	29	26	89.7
36	Electrical and electronic machinery	forecasts	2064	641	31.1%
		analysts	503	41	8.2
		firms	28	24	85.7
49	Electric, gas and sanitary services	forecasts	4664	2888	61.9%
		analysts	262	49	18.7
		firms	63	63	100.0
60	Banking	forecasts	2596	1594	61.4%
		analysts	178	41	23.0
		firms	36	36	100.0
63	Insurance	forecasts	1076	492	45.7%
		analysts	138	19	13.8
		firms	9	9	100.0

[1] The nine 2-digit SIC industries examined are those with the most forecasts available. The larger dataset from which these nine industries are chosen is described in Table 1.

[2] The data sufficiency requirements are that analysts must have at least ten forecasts in the industry, and must have forecasts in at least three of the years 1975-81.

TABLE 3

Distribution of Ages of Sample Forecasts, by Industry[1]

	Percentiles							
SIC	.1	.25	.5	.75	.9	mean	std. dev.	N
26	10	23	53	92	155	67.1	60.9	826
28	6	16	45	91	161	64.5	64.9	2900
29	8	19	44	72	116	57.0	53.8	1775
33	9	22	45	88.5	148	64.8	63.6	524
35	6	21	49	83	135.6	63.6	60.5	1106
36	6	17	41	75	126.6	56.4	57.4	641
49	13	28	59	120	211	85.2	75.1	2888
60	10	21	48	88	154	64.0	58.4	1594
63	11.3	28.3	71	114	192.4	84.5	71.1	492

[1] A forecast's age is the number of trading days between the analyst's forecast date and the minimum horizon date. The minimum horizon date in this study is 120 trading days prior to the annual EPS announcement. The age distribution is over all forecasts, by industry. N denotes the sample size in each industry.

TABLE 4

Estimation of Absolute Forecast Error Model with Analyst, Firm and Year Fixed Effects, For Nine Industries in the Period 1975 to 1981[1]

$$| e_{ijt} | = \mu_i + \delta_j + \gamma_t + \eta_{ijt}$$

	Analysts[2]		Firms[2]		Years[2]		Model		[3]
SIC	n_i	F	n_j	F	n_t	F	N	R^2	F
26	29	0.47	15	12.45	7	13.70	826	.27	5.83
28	104	0.79	43	16.99	7	2.04	2900	.32	8.73
29	50	0.91	20	12.12	7	30.80	1775	.24	7.43
33	26	0.76	11	4.22	7	7.98	524	.21	3.11
35	63	0.87	26	3.33	7	5.26	1106	.16	2.09
36	41	1.08	24	11.93	7	8.32	641	.43	6.36
49	49	1.03	63	13.99	7	4.95	2888	.29	9.81
60	41	0.49	36	5.30	7	6.02	1594	.14	3.15
63	19	0.83	9	21.34	7	6.70	492	.36	7.92

[1] The estimated equation is (2) in the text, estimated separately for each industry. The dependent variable $|e_{ijt}|$ is the absolute value of the forecast error from analyst i on firm j's earnings for year t, where the forecast is made at least 120 trading days prior to the annual earnings announcement. μ_i, δ_j, and γ_t are analyst, firm, and year fixed effects, respectively, and η_{ijt} is an error term.

[2] The numbers of analysts, firms, years, and forecasts in each industry sample are denoted by n_i, n_j, n_t and N, respectively. All F-statistics in this table have denominator degrees of freedom equal to $(N - n_i - n_j - n_t + 2)$. The numerator degrees of freedom appear below. The F-statistic for analysts tests homogeneity of average forecast accuracy across analysts, conditional on the firms and years in the sample, by industry. The F-statistics for firms and years are defined analogously. The ranges of sampling theory significance (α) levels across industries for the reported Fs are:

F-statistic for:	Numerator d.f.	Range of α levels:
analysts	$(n_i - 1)$	.34 to .99
firms	$(n_j - 1)$	< .0001
years	$(n_t - 1)$	< .0001 to .06
full model	$(n_i + n_j + n_t - 3)$	< .0001

These indicate rejection of homogeneity across firms and years, and failure to reject homogeneity across analysts at conventional significance levels.

[3] Sample size, unadjusted R^2 and the full model F-statistic are reported.

TABLE 5

Comparison of the Observed Distribution of Average Ranks with the Expected Distribution if All Analysts are Alike, with Equal Probabilities of Each Rank in Each Year

The Paper Industry (SIC 26) as an Example

Analyst id code[1]	Number of years[2]	Average rank [3]	Cumulative Density		
			Observed[4]	Expected[5]	|Difference|
12 41988	3	1.33	0.03	0.03	0.00
42 8415	3	1.33	0.07	0.03	0.04
12 31650	3	1.67	0.10	0.09	0.02
17 6361	4	1.75	0.14	0.12	0.02
45 18700	4	1.75	0.17	0.12	0.06
26 35250	3	2.00	0.21	0.26	0.05
4 9031	3	2.00	0.24	0.26	0.02
9 4790	3	2.00	0.28	0.26	0.02
23 23523	7	2.14	0.31	0.27	0.04
26 35068	4	2.25	0.34	0.35	0.00
20 44636	7	2.29	0.38	0.36	0.02
10 448	5	2.40	0.41	0.45	0.04
13 24109	5	2.40	0.45	0.45	0.01
36 1498	5	2.40	0.48	0.45	0.03
38 90000	5	2.40	0.52	0.45	0.06
48 1200	4	2.50	0.55	0.53	0.02
16 37978	7	2.57	0.59	0.55	0.04
37 44985	4	2.75	0.62	0.71	0.09
44 14497	4	2.75	0.66	0.71	0.05
53 4500	4	2.75	0.69	0.71	0.02
65 1800	4	2.75	0.72	0.71	0.02
19 41566	5	2.80	0.76	0.73	0.03
1 42804	3	3.00	0.79	0.87	0.07
46 2511	6	3.00	0.83	0.87	0.04
67 3600	4	3.00	0.86	0.87	0.00
86 650	3	3.00	0.90	0.87	0.03
9 4382	4	3.00	0.93	0.87	0.07
5 27755	3	4.00	0.97	1.00	0.03
81 4000	3	4.00	1.00	1.00	0.00

Maximum Absolute Difference: 0.09
Kolmogorov - Smirnov Statistic: 0.38

[1] "Analyst id codes" are the identifying code numbers assigned by I/B/E/S to analysts and brokerage houses for confidentiality.

[2] "Number of years" denotes the number of years that an analyst appears in the sample.

[3] Average rank is the average, across the years the analyst is in the sample, of the analyst's forecast accuracy, ranked in quartiles in each year.

[4] The observed cumulative density assigns a probability mass of 1/29, or 0.03448, to each of the 29 analysts.

[5] The expected cumulative density is based on the null hypothesis that all analysts are alike, with probability .25 of falling in each quartile in each year.

TABLE 6

Kolmogorov-Smirnov Test Statisitics for Nine Industries, Testing the Null Hypothesis that the Distribution of Average Quartile Rankings is Indistinguishable from the Expected Distribution if All Analysts are Alike, Each Quartile is Equally Likely, and Each Year is an Independent Trial

SIC	Number of analysts	D_{MN}^{*} [1]	KS [2]
26	29	0.09	0.38
28	104	0.18	1.08
29	50	0.15	0.76
33	26	0.20	0.83
35	63	0.12	0.67
36	41	0.14	0.66
49	49	0.11	0.56
60	41	0.16	0.77
63	19	0.10	0.39

[1] D_{MN}^{*} denotes the absolute value of the maximum difference between the observed and expected cumulative densities of analyst forecast accuracy. An example comparison of these densities for one industry appears in Table 5.

[2] KS denotes the Kolmogorov-Smirnov test statistic. It is computed:

$$KS = [MN/(M+N)]^{.5} D_{MN}^{*} ,$$

where N equals the number of analysts in the industry and M equals 55. Critical values for statistical tests (interpolated from DeGroot (1986) p. 555) are:

P[Type I error]	KS > critical value
.05	1.36
.10	1.23
.15	1.14
.20	1.07

TABLE 7

Student's-t Test for Differences in Average Forecast Accuracy between Sample Analysts and Analysts Excluded by the Sample Selection Criteria, For Nine Industries[1]

SIC	μ_S [2]	μ_E [3]	$t[H_0: \mu_S=\mu_E]$ [4]
26	-0.013	0.011	0.693
28	-0.007	0.009	0.878
29	-0.019	0.029	1.092
33	-0.081	0.080	1.646
35	-0.015	0.010	0.475
36	-0.019	0.009	1.280
49	-0.009	0.014	3.683
60	0.009	-0.015	-0.777
63	0.026	-0.022	-1.666

[1] Average forecast accuracy, μ_S or μ_E, is the group mean of residuals from equation (5) in the text, which purges absolute forecast errors of year and firm effects. The means are computed over the Sample and Excluded groups, respectively, as shown in equation (6) in the text. The t-statistic tests the hypothesis that the group means are equal.

[2] The Sample group is the set of analysts included in the investigation in sections 3 and 4, each of whom has at least ten forecasts in at least three years between 1975 and 1981.

[3] The Excluded group is the set of analysts in the I/B/E/S database for these nine industries that were excluded by the sample selection criteria described in note 2.

[4] The t-statistic fails to reject the null at the .05 level in all industries except for industry 49, in which the null is rejected at the .0002 level.

TABLE 8

Estimation of Absolute Forecast Error Model with Analyst, Firm and Year Fixed Effects and Age Categories, For Nine Industries in the Period 1975 to 1981[1]

$$| e_{ijt} | = \mu_i + \delta_j + \gamma_t + \alpha_a + w_{ijt}$$

	Analysts[2]		Firms[2]		Years[2]		Age[2]		Model[3]		
SIC	n_i	F	n_j	F	n_t	F	n_a	F	N	R^2	F
26	29	0.45	15	12.28	7	13.45	4	0.49	826	.27	5.50
28	104	0.78	43	17.06	7	2.04	4	1.47	2900	.33	8.59
29	50	0.96	20	12.43	7	32.76	4	8.35	1775	.26	7.56
33	26	0.74	11	4.22	7	7.93	4	0.09	524	.21	2.89
35	63	0.93	26	3.50	7	5.28	4	2.62	1106	.17	2.12
36	41	1.11	24	11.74	7	8.00	4	0.81	641	.44	6.12
49	49	1.01	63	13.94	7	4.77	4	0.23	2888	.29	9.56
60	41	0.53	36	5.32	7	5.57	4	2.28	1594	.15	3.13
63	19	0.82	9	21.50	7	6.04	4	0.78	492	.36	7.30

[1] The estimated equation is (7) in the text. It differs from equation (2), reported in Table 4, by the inclusion of categories indicating forecast age. The age categories, in trading days, are: (1) less than 22, (2) 23 to 44, (3) 45 to 66, and (4) more than 66. α_a is the coefficient on the age category. All other coefficients are defined as in equation (2) and Table 4. w_{ijt} is an error term.

[2] The F-statistics are defined and interpreted similarly to those in Table 4. The ranges of sampling theory significance (α) levels across industries for the reported Fs are:

F-statistic for:	Numerator d.f.	Range of α levels:
analysts	$(n_i - 1)$	.32 to .99
firms	$(n_j - 1)$	< .0001
years	$(n_t - 1)$	< .0001 to .06
age	$(n_a - 1)$	< .0001 to .99
full model	$(n_i + n_j + n_t + n_a - 4)$	< .0001

These indicate rejection of homogeneity across firms and years, and failure to reject homogeneity across analysts at conventional significance levels. Age categories contribute significant explanatory power at the .10 level or better for SIC 29, 35, and 60.

[3] Sample size, unadjusted R^2 and the full model F-statistic are reported.