How Important Is Past Analyst Forecast Accuracy?

Lawrence D. Brown

The Wall Street Journal rates analysts on the basis of past earnings forecast accuracy. These analyst ratings are important to practitioners who believe that past accuracy portends future accuracy. An alternative way to assess the likelihood of “more” or “less” accurate forecasts in the future is to model the analyst characteristics related to the accuracy of individual analysts’ earnings forecasts. No evidence yet exists, however, as to whether an analyst characteristics model is better than a past accuracy model for distinguishing more accurate from less accurate earnings forecasters. I show that a simple model of past accuracy performs as well for this purpose as a more complex model based on analyst characteristics. The findings are robust to annual and quarterly forecasts and pertain to estimation and prediction tests. The evidence suggests that practitioners’ focus on past accuracy is not misplaced: It is as important as five analyst characteristics combined.

Academic researchers have identified numerous intuitively appealing analyst characteristics related to the accuracy of analysts’ earnings forecasts. For example, Clement (1999) suggested five analyst characteristics: company-specific experience, general experience, number of companies followed, number of industries followed, and size of brokerage house.1 I refer to a type of model proposed by Clement as an ANCHAR (analyst characteristics) model.2 Foreknowledge of analyst forecast accuracy is valuable. Brown and Mohammad (2001) showed that, as of a fiscal quarter-end, if a weighted consensus estimate exceeds a simple consensus estimate by at least 5 percent, then “knowing” individual analyst forecast accuracy with respect to the upcoming quarterly earnings number and buying stocks based on that knowledge provided an average market-adjusted return of 4.12 percent during the 12 years of 1987–1998.

Of course, no model can identify the most accurate analyst with certainty, so imperfect models are used for this purpose. Practitioners often rely on the accuracy of past analyst earnings forecasts to predict future analyst earnings forecast accuracy. As examples, the Institutional Investor All-America Research Team and the StarMine SmartEstimate are based partly on past accuracy; the ratings published annually in the Wall Street Journal are based entirely on past accuracy. Consistent with the Street wisdom that past accuracy portends future accuracy, Stickel (1992) and Sinha, Brown and Das (1997) showed that past accuracy is significantly positively correlated with current accuracy. I refer to a type of model motivated by Stickel and by Sinha et al. as a PASTACC (past accuracy) model.

No evidence exists as to whether an ANCHAR model outperforms a PASTACC model. My purpose is to compare the performance of these two models. I use quarterly and annual data, and I evaluate performance by using two tests—estimation and prediction.

Samples and Methodology

I used the Thomson Financial I/B/E/S U.S. Detail file of analyst annual and quarterly earnings forecasts for the 13-year period of 1986–1998. Similar to the extant literature, I obtained the last forecast made by each analyst during the period. Consistent with Clement, I omitted analysts in the quarterly database at its beginning (1983) because ascertaining these analysts’ experience levels is impossible. I estimated parameters of the ANCHAR model by using a control for forecast age and the five analyst characteristics Clement suggested (two proxies for experience, two for task complexity, and one for resources).3 I estimated parameters of the PASTACC model by using a control for forecast age and a single

Laurence D. Brown is J. Mack Robinson Distinguished Professor of Accountancy at J. Mack Robinson College of Business, Georgia State University, Atlanta.
factor, namely, past accuracy—that is, the immediate past value of the dependent variable. I also estimated a model that is a “superset” of these two models, namely, a model consisting of forecast age, past accuracy, and the five analyst characteristics.

The precise definitions of my variables follow, beginning with the dependent variable. For ease of exposition, I use Clement’s definitions and variable names whenever applicable.

- **PMAFE** (accuracy). The individual analyst’s forecast error that year minus the mean of the forecast errors of all analysts following the company that year scaled by the mean of the forecast errors of all analysts following the company that year. Forecast error is defined as the absolute value of the difference between I/B/E/S actual annual earnings and the last forecast made by the analyst for that year.

- **DAGE** (forecast age). Number of calendar days between the analyst’s last annual forecast and the fiscal year-end minus the average “forecast age” of all analysts following the company that year.

- **DGEXP** (general experience). Number of years (including year t) the analyst had been in the database minus the average “general experience” of all analysts following the company that year.

- **DCEXP** (company-specific experience). Number of years (including year t) the analyst had been in the database for the company minus the average “company-specific experience” of all analysts following the company that year.

- **DNICS** (company complexity). Number of companies the analyst followed during the year minus the average “company complexity” of all analysts following the company during that year.

- **DNSIC** (industry complexity). Number of two-digit SICs the analyst followed during the year minus the average “industry complexity” of all analysts following the company that year.

- **DNTOPI10** (brokerage size). Dummy variable equal to 1 if the analyst worked at a top-decile brokerage house (based on the size of the firm) and 0 otherwise minus the mean value of this variable for all analysts following the company that year.

- **LPMAFE** (past accuracy): PMAFE last year. If the analyst had no history of following the company, this observation was deleted.

I obtained the SICs (for **DNSIC**2) from Compustat, and all other data are from the I/B/E/S data supplied to academic users. My analyses of quarterly data are based on the same variable definitions as for the analyses of annual data except that **PMAFE**, **DAGE**, and **LPMAFE** are defined relative to the quarter rather than the year.

The ANCHAR model adds the five analyst characteristics proposed by Clement to the control variable, **DAGE**, as follows:

\[
PMAFE = \gamma_1 DAGE + \gamma_2 DGEXP + \gamma_3 DCEXP + \gamma_4 DNICS + \gamma_5 DNSIC2 + \gamma_6 DNTOPI10 + \mu, \tag{1}
\]

where \( \mu \) is a random variable with mean 0 and variance \( \sigma \). The PASTACC model adds the immediate past value of the dependent variable to the control variable, **DAGE**:

\[
PMAFE = \beta_1 DAGE + \beta_2 LPMAFE + \nu, \tag{2}
\]

where \( \nu \) is a random variable with mean 0 and variance \( \sigma \). Using these two models, I obtained estimation results for the 1986–98 period. Based on O’Brien (1988), Mikhail, Walther, and Willis (1997), Sinha et al., Clement, and Jacob, Lys and Neale (1999), I expected **PMAFE** to be positively related to **DAGE**. Based on Clement, I expected **PMAFE** to be negatively related to **DGEXP**, **DCEXP**, and **DNTOPI10** and to be positively related to **DNICS** and **DNSIC2**. Based on Stiebel and on Sinha et al., I expected **PMAFE** to be positively related to **LPMAFE**.

I evaluated the estimation performance of ANCHAR versus that of PASTACC by comparing the two models’ adjusted \( R^2 \)s. I considered the model providing greater explanatory power to be the better model for estimation purposes. I obtained predictive results by using the 12 years of 1987 through 1998 based on estimating parameters for each year and estimating **PMAFE** for the following year. (For example, I estimated parameters for 1990 and estimated **PMAFE** for 1991.) Based on the estimated values of **PMAFE**, I identified the two extreme deciles—analysts expected to be the most and least accurate earnings forecasters. I then determined the mean actual values of **PMAFE** in the extreme deciles to ascertain these analysts’ forecast accuracy. I evaluated comparative predictive performance of the models by determining which one did the better job of identifying actual analyst performance (accuracy) in these extreme deciles.

### Results

I report in this section summary statistics and correlation coefficients for the sample of 123,670 observations, results of the estimation procedures, and analysis of which model is the better predictor.

**Summary Statistics and Correlation Coefficients.** The annual data consist of the 123,670 observations for which all data were available.
Panel A of Table 1 contains summary statistics for the dependent variable, the five analyst characteristics of the ANCHAR model, and past accuracy, as proxied by the lagged value of the dependent variable. The summary statistics are the mean, first quartile value, median, and third quartile value. In general, the distributional properties of the data are similar to those Clement found.

Panel B of Table 1 presents Pearson correlation coefficients between \( PMAFE \), \( DAGE \), the five analyst characteristics of the ANCHAR model, and the past value of the dependent variable, \( LPMAFE \). As expected, \( PMAFE \) is negatively correlated with \( DGEXP \) and \( DCEXP \) and positively correlated with \( DAGE \) and \( DNSIC2 \).\(^8\) Also as expected, \( PMAFE \) is positively correlated with \( LPMAFE \). Indeed, \( LPMAFE \) has a higher correlation with \( PMAFE \) than do any of the five analyst characteristics. Contrary to expectations, \( PMAFE \) is negatively correlated with \( DNCOS \), but the relationship between the two is insignificant.

**Estimation Results.** Panels A and B of Table 2 present ordinary least-squares estimation results for, respectively, the ANCHAR and PASTACC models for the study period. Each panel provides annual and quarterly results. Panel A shows that \( DAGE \) is positive and significant, \( DNTOP10 \) is the most important of the five analyst characteristics (as evidenced by the absolute value of its \( t \)-statistic), and \( DCEXP \) is the second most important characteristic.\(^9\) The estimation results of the ANCHAR model are similar for both annual and quarterly data.

To ascertain whether ANCHAR provides better estimation results than PASTACC, one can compare the two models' adjusted \( R^2 \)s (Panel A versus Panel B). The mean differences between the annual adjusted \( R^2 \)s is only 0.0013 and between the quarterly adjusted \( R^2 \)s is only 0.0004. These mean differences are insignificantly different from zero. In summary, the simple model performs as well as the complex model. Thus, when added to forecast age, past accuracy alone is as important for explaining differential accuracy as the five analyst characteristics related to forecast accuracy combined.

---

### Table 1. Summary Statistics and Correlation Matrix: Annual Data, 1986–98

**A. Summary statistics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (^a)</th>
<th>Quartile 1</th>
<th>Median</th>
<th>Quartile 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( PMAFE )</td>
<td>0.000</td>
<td>-0.600</td>
<td>-0.143</td>
<td>0.286</td>
</tr>
<tr>
<td>( DAGE )</td>
<td>0.000</td>
<td>-59.360</td>
<td>-15.240</td>
<td>43.400</td>
</tr>
<tr>
<td>( DGEXP )</td>
<td>0.000</td>
<td>-1.545</td>
<td>-0.118</td>
<td>1.333</td>
</tr>
<tr>
<td>( DCEXP )</td>
<td>0.000</td>
<td>-0.929</td>
<td>0.000</td>
<td>0.667</td>
</tr>
<tr>
<td>( DNCOS )</td>
<td>0.000</td>
<td>-1.000</td>
<td>0.000</td>
<td>0.571</td>
</tr>
<tr>
<td>( DNSIC2 )</td>
<td>0.000</td>
<td>-0.200</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( DNTOP10 )</td>
<td>0.000</td>
<td>-0.389</td>
<td>0.000</td>
<td>0.400</td>
</tr>
<tr>
<td>( LPMAFE )</td>
<td>0.000</td>
<td>-0.391</td>
<td>-0.063</td>
<td>0.213</td>
</tr>
</tbody>
</table>

**B. Pearson correlations**

<table>
<thead>
<tr>
<th></th>
<th>( PMAFE )</th>
<th>( DAGE )</th>
<th>( DGEXP )</th>
<th>( DCEXP )</th>
<th>( DNCOS )</th>
<th>( DNSIC2 )</th>
<th>( DNTOP10 )</th>
<th>( LPMAFE )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( PMAFE )</td>
<td>1.000(***)</td>
<td>0.414(***)</td>
<td>-0.033(***)</td>
<td>-0.040(***)</td>
<td>-0.004</td>
<td>0.006(**)</td>
<td>-0.050(***)</td>
<td>0.077(***)</td>
</tr>
<tr>
<td>( DAGE )</td>
<td>1.000(***)</td>
<td>-0.053(***)</td>
<td>-0.051(***)</td>
<td>-0.041(***)</td>
<td>-0.011(***)</td>
<td>-0.013(***)</td>
<td>0.039(***)</td>
<td></td>
</tr>
<tr>
<td>( DGEXP )</td>
<td>1.000(***)</td>
<td>0.646(***)</td>
<td>0.159(***)</td>
<td>0.086(***)</td>
<td>0.024(***)</td>
<td>0.015(***)</td>
<td>0.005(***)</td>
<td></td>
</tr>
<tr>
<td>( DCEXP )</td>
<td>1.000(***)</td>
<td>0.061(***)</td>
<td>0.032(***)</td>
<td>0.029(***)</td>
<td>0.092(***)</td>
<td>0.032(***)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( DNCOS )</td>
<td>1.000(***)</td>
<td>0.144(***)</td>
<td>0.073(***)</td>
<td>0.031(***)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( DNSIC2 )</td>
<td>1.000(***)</td>
<td>0.032(***)</td>
<td>1.000(***)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( DNTOP10 )</td>
<td>1.000(***)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: \( N = 123,670 \).

\(^a\)Because Clement's procedure mean-adjusts all variables, all means are 0.000.

\(^*\)Significant at the 5 percent level or better in a two-tailed test.

\(^**\)Significant at the 1 percent level or better in a two-tailed test.
Table 2. Estimation Results, 1986–98
(expected signs in parentheses)

<table>
<thead>
<tr>
<th>Data Period</th>
<th>DAGE (+)</th>
<th>DGEXP (-)</th>
<th>DCEXP (-)</th>
<th>DNCOS (+)</th>
<th>DNSIC2 (+)</th>
<th>DNTOP10 (-)</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual</td>
<td>0.0043</td>
<td>0.0007</td>
<td>-0.0111</td>
<td>0.0030</td>
<td>0.0029</td>
<td>-0.0910</td>
<td>0.1740</td>
</tr>
<tr>
<td>t-Statistic</td>
<td>159.546***</td>
<td>0.512</td>
<td>-6.162***</td>
<td>4.509***</td>
<td>1.011</td>
<td>-16.633***</td>
<td></td>
</tr>
<tr>
<td>Quarterly</td>
<td>0.0034</td>
<td>0.0019</td>
<td>-0.0033</td>
<td>-0.0002</td>
<td>-0.0003</td>
<td>-0.0663</td>
<td>0.0151</td>
</tr>
<tr>
<td>t-Statistic</td>
<td>61.424***</td>
<td>2.479</td>
<td>-9.900***</td>
<td>-0.231</td>
<td>-0.135</td>
<td>-19.934***</td>
<td></td>
</tr>
</tbody>
</table>

B. PASTACC model (Equation 2)

<table>
<thead>
<tr>
<th>Data Period</th>
<th>DAGE (+)</th>
<th>LPMAFE (+)</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual</td>
<td>0.0043</td>
<td>0.0793</td>
<td>0.1753</td>
</tr>
<tr>
<td>t-Statistic</td>
<td>159.383***</td>
<td>23.701***</td>
<td></td>
</tr>
<tr>
<td>Quarterly</td>
<td>0.0033</td>
<td>0.0512</td>
<td>0.0155</td>
</tr>
<tr>
<td>t-Statistic</td>
<td>60.197***</td>
<td>25.212***</td>
<td></td>
</tr>
</tbody>
</table>

Note: For annual data, N = 123,670; for quarterly data, N = 273,643.
***Significant at the 1 percent level or better in a one-tailed test.

Predictive Results: ANCHAR versus PASTACC. To determine whether ANCHAR is a better predictive model than PASTACC, I estimated a cross-sectional regression for both models in each year from 1986 through 1998 and made predictions for the following year. For example, for the 1997 predictions, I used parameter estimates of the 1996 cross-sectional regressions.10

I examined how well each model identified analysts that were expected to be most accurate or to be least accurate and the difference between these two extremes (the spread) by ranking the predicted PMAFEs from lowest to highest and forming 10 deciles, of which the first (last) decile depicts those analysts predicted to be most (least) accurate. Next, I calculated the mean actual PMAFEs of the first and tenth deciles and calculated the difference in the mean actual PMAFE for the tenth decile minus that of the first decile (the spread).

Panel A of Table 3 reveals the performance of the models, based on both annual and quarterly data, at predicting the most accurate analysts. Analysts expected to be the most accurate by the ANCHAR model were 31.56 percent more accurate than average. Thus, if the average error for all analysts following the company in a given year was $1, analysts predicted to be best by the model had an average absolute error of 68.44 cents (see the definition of PMAFE). The PASTACC model performed slightly better when either annual or quarterly forecasts were used.

Panel B of Table 3 indicates the performance of the models at predicting the least accurate analysts. Once again, the PASTACC model performed slightly better than the ANCHAR model, and the results are robust to annual and quarterly forecasts. Given the results for annual forecasts, those analysts expected to be least accurate by the ANCHAR and PASTACC models were actually, respectively, 95.27 percent and 95.36 percent less accurate than average. If the average error for all analysts following the company in a given year was $1, the average analyst predicted to be in the worst decile by both
the ANCHAR and PASTACC models actually had (absolute) errors of about $1.95. The results were similar with quarterly forecasts.

Panel C results, calculated by subtracting the Panel A numbers from the Panel B numbers, shows how good the models were relative to each other at distinguishing the most from the least accurate earnings forecasters in this study. Based on the Panels A and B results, the spreads must favor the PASTACC model. The predictive results mirror the estimation results: ANCHAR is not a relatively better predictive model than PASTACC. When added to forecast age, the single factor, past accuracy, performs as well as that of five analyst characteristics combined.\textsuperscript{11}

**Predictive Results: COMBINED versus PASTACC.** To provide more evidence regarding the importance of past accuracy, I compared the predictive results of the PASTACC model with that of a COMBINED model, which is a union of the ANCHAR and PASTACC models. In other words, the COMBINED model equals LPMAFE plus the variables in ANCHAR.\textsuperscript{12} Results are given in the last column of Table 3. As the data in Panel A show, the COMBINED model made more accurate predictions of which analysts would be the most accurate than the PASTACC model but its predictive advantage, as a comparison of the columns shows, is trivial. The same is true in Panel B for accuracy in predicting which analysts would be the least accurate. Of course, the results for spreads provide a similar signal, namely, that the five analyst characteristics add trivially to past accuracy.\textsuperscript{13}

**Conclusions**

I compared the performance of an analyst characteristics model patterned on Clement with a past accuracy model patterned on Stickel and Sinha et al. The past accuracy model used a single variable, past accuracy, in lieu of the five analyst characteristics suggested by Clement.\textsuperscript{14} I provided both estimation and predictive evidence, and I used annual and quarterly forecasts. The results show that the past accuracy model performs as well as the analyst characteristics model. This finding is robust to both estimation and prediction methodologies and to both annual and quarterly forecast data.

These findings suggest that practitioners seeking to predict those analysts who are most or least likely to be accurate earnings forecasters can do as well by combining forecast recency with past accuracy as by combining forecast recency with the five analyst characteristics I considered: company-specific experience, general experience, number of companies followed, number of industries followed, and size of brokerage house. A model combining forecast recency, past accuracy, and the five analyst characteristics predicted slightly better than a model that omitted the analyst characteristics. Whether its slight predictive advantage justifies its additional complexity depends on the decision context, a subject beyond the scope of this research.

---

I benefited from the comments of workshop participants at the University of Cincinnati, the University of Colorado, Dartmouth College, and Georgia State University. The helpful comments of Donald Byard, Carol Frost, John Jacob, Barry Lewis, Kevin Sachs, Phil Shane, Kumar Sivakumar, Richard Willis, Kent Womack, Zhaofeng Xue, and Ping Zhao are appreciated. The research assistance of Emad Mohammad was instrumental to this study. I am also grateful to Thomson Financial I/B/E/S for providing the data used in the study.

---

**Notes**

1. See also Mikhail, Walther, and Willis (1997) and Jacob, Lys, and Neale (1999).
2. Specifically, the model distinguishes between more accurate and less accurate analyst/firm forecasters, rather than between analysts per se, because an analyst who is the most accurate earnings forecaster for one company may be the least accurate earnings forecaster for another company.
3. When estimating the ANCHAR model, I made three research design choices different from the choices Clement made. First, I included the last forecast by the analyst as long as it was made before the earnings announcement date; Clement required forecasts to have been made at least 30 days prior to the earnings announcement date. My choice increased sample size and enabled inclusion of forecasts for constructing a proxy of the market's consensus estimate immediately prior to an earnings announcement. Second, to focus on individual analysts, Clement excluded forecasts by teams, but he needed a special tape from I/B/E/S for this purpose. I included forecasts by teams because I chose to use the academic tapes generally available from I/B/E/S, which do not separately identify forecasts made by teams. Third, Clement included all companies, even those followed by a single analyst. Consistent with Jacob et al., I excluded companies followed by fewer than three analysts in a given period to mitigate the probability that my dependent variable was unrelated to my independent variables (as it is by construction if one analyst is used).
4. Each independent variable name begins with “D” to indicate it is a difference from the mean.
6. Specifically, I used all available data for a particular year to estimate the model cross-sectionally. This estimation procedure yielded average parameter estimates that I multiplied by an individual analyst's identifying characteristics (e.g., DGEXP) to estimate his/her PMAFE for the following year.

7. My procedure is similar to that used by the standardized unexpected earnings (SUE) literature, which sorts SUEs into deciles (Bernard and Thomas 1990) and formulates trading rules to "buy" ("sell") stocks in the extreme deciles of positive (negative) SUEs.

8. The larger PMAFE is, the less accurate the analyst is.

9. Because the t-value is the square root of the partial F, which is the factor's incremental explanatory power given inclusion of the other factors, it indicates the factor's relative importance.

10. The predictive evidence is based on a period one year shorter than the period for the estimation evidence because I had no predictive evidence for 1986, the first year used to estimate the models' coefficients.

11. I also took the difference between the two models' spreads each year and calculated the mean and standard deviation. Similar to the estimation results, I found mean differences for both the annual and quarterly samples to be insignificantly different from zero.

12. For simplicity, I do not present the estimation results. By construction, a model fitted to more than 100,000 observations, arising from the union of the two models, provides more explanatory power than a model embedded in it. Interestingly, the explanatory power of the COMBINED model differed only slightly, however, from that of the PASTACC model. The COMBINED model's adjusted $R^2$ was 0.1775 for annual data and 0.0173 for quarterly data.

13. I also took the difference between the COMBINED and ANCHAR models' spreads each year and calculated the mean and standard deviation. Similar to the results of the comparison of ANCHAR and PASTACC, I found the mean differences for both the annual and quarterly samples to be insignificantly different from zero.

14. I purposely selected the simplest possible past accuracy metric—on based on a single past "at bat"—to showcase its comparative importance. I obtained better results by using an average of multiple at bats.

References


