# A Note On Earnings Forecast Source Superiority

Donald P. Pagach (E-mail: Don@ncsu.edu), North Carolina State University
Barbara A. Chaney (E-mail: Barb_Chaney@ncsu.edu), North Carolina State University
Bruce C. Branson (E-mail: Bruce_Branson@ncsu.edu) North Carolina State University

## Abstract

*We examine the forecast accuracy of Value Line analysts relative to the Brown-Rozeff (100)X(011)$_4$ ARIMA model. We find that for a surprising percentage (35-41%) of our sample of small firms that time series-based earnings per share predictions are more accurate than those obtained from The Value Line Investment Survey. Further, we document exploitable characteristics of each subgroup that are associated with forecast origin. In those instances where the seasonal, univariate earnings forecast model identified by Brown and Rozeff (1979) produces more accurate forecasts than Value Line, we find significant differences in firm size, degree of diversification, magnitudes of the autoregressive and seasonal moving-average parameters, residual standard errors, and magnitude of the Ljung-Box Q-statistic. We use probit regressions to identify ex ante those firms likely to be accurately forecast by each source. We achieve a marginal improvement in forecast accuracy, which suggests there is potential for using ex ante decision rules to improve forecast accuracy.*

## 1. Introduction

The purpose of this paper is to further investigate whether analysts' forecasts are superior to those produced by univariate statistical-based models. While an impressive array of evidence supports this claim,[1] we suggest that this finding is only true "on average" and that for specific firms, over various forecast horizons, time series-based forecast models can outperform analysts on a consistent basis. Specifically, we find that for a sample of small capitalization firms covered by *The Value Line Investment Survey* a significant minority of their quarterly earnings per share (EPS) forecasts are more accurately produced by the (100)X(011)$_4$ autoregressive-integrated-moving-average (ARIMA) model identified by Brown and Rozeff (1979), as opposed to a *Value Line* analyst.

This finding is surprising for at least three reasons. One, analysts are well compensated for their production of (purportedly) accurate EPS forecasts (among other items), and their existence implies that their output has value. Two, analysts possess an information advantage over mechanical time-series models. Univariate time-series models are functions of the historical earnings series alone; while analysts, on the other hand, are able to draw upon all available information relevant to predicting firms' EPS, one item of which is the mechanical forecast. Three, analysts have a timing advantage over mechanical models. Analysts' forecasts are determined after the previous quarters' earnings are announced (in the case of one-step-ahead predictions)--often several weeks after--while the mechanical model produces an updated forecast of next quarters' EPS immediately upon receipt of the latest realization. This information, taken together, suggests that security analysts should outperform time-series models on a regular basis. In fact, Schipper (1991, 107) states that:

---

*Readers with comments or questions are encouraged to contact the authors via email.*

*". . . a surprising degree of incompetence or an exotically perverse objective function would be required for analysts to lack the capability to equal or exceed the predictive accuracy of models based on past earnings most or all of the time."*[2]

However, for our sample of small capitalization firms, we find that at the one-quarter-ahead forecast horizon the $(100)X(011)_4$ seasonal ARIMA model is more accurate than *Value Line* analysts 35% of the time. This percentage increases as the forecast horizon lengthens from one-quarter to two-quarters-ahead (35% to 40.8%) and from two-quarters-ahead to three-quarters-ahead (40.8% to 41.2%).

Given this surprising result, we identify characteristics of occasions when the EPS of small capitalization firms are more accurately forecast by the Brown-Rozeff (BR) ARIMA model (BR firms) vis-à-vis *Value Line* analysts (hereafter, VL firms). We find that the BR firms are typically smaller in size and engage in fewer lines of business relative to the VL firms. We also find that the time series characteristics of the BR and VL firms differ markedly. BR firms possess significantly larger first-order autoregressive (AR1) parameters, significantly smaller seasonal moving-average (MA4) parameters, smaller residual standard errors, and slightly higher levels of residual autocorrelation relative to VL firms.

We attempt to exploit the differences found in the descriptive portion of our study by constructing a prediction model to indicate which forecast source is likely to yield a more accurate earnings forecast over one, two, and three-quarter-ahead forecast horizons. Using a probit regression model, we estimate the probability that a given firm's EPS will be best predicted by VL analysts over each of the three separate prediction intervals. Our results indicate that a marginal improvement in forecast accuracy is obtained by adopting our approach and we argue that this provides an opportunity for additional improvement in forecast accuracy.

Our ability to identify firms or forecast occasions for which mechanical forecasts are superior to analysts' forecasts could be valuable to the business research community. This is due to the widespread use of earnings forecasts as proxies of market expectations in capital markets research. The improvement in forecast accuracy is striking if "perfect information" were available as to which of the two forecast sources should be used. In our sample of 366 one-quarter-ahead forecasts from each source in 1988, we find that exclusive reliance on time series-based forecasts produces a median absolute percentage forecast error (APFE) of 26.7%; exclusive use of *Value Line* forecasts, a median APFE of 16.7%; and for "perfect information" forecasts using our approach to determining which forecast to use, a median APFE of only 11.5%. Similar improvements are possible for other forecast horizons and are documented in a later section.

The remainder of this paper is organized as follows: first, additional background is provided on both investigations of financial analyst forecast superiority and on attempts to improve on analyst-based forecasts by constructing combination forecast models that use information from multiple sources. Next, a description of our sample and the descriptive profiles of the two subgroups (BR and VL firms) are provided. Third, probit regression results are presented along with predictive ability tests on the 1989 holdout sample. Finally, a discussion of conclusions and limitations of our study is provided.

## 2. Background

Brown, Hagerman, Griffin, and Zmijewski (1987a) provide an extensive examination of *Value Line* earnings forecasts vis-à-vis those from three popular, cross-sectionally derived univariate time-series models.[3] They find that VL forecasts are more accurate, on average, than forecasts from each of the time-series models over a six-year (24-quarter) period, three forecast horizons (one, two, and three-quarters-ahead), and a variety of error metrics and truncation alternatives. They provide evidence that analyst forecast superiority is attributable to both a timing advantage and a contemporaneous advantage. We
investigate the degree of superiority by examining accuracy on a forecast-specific basis as opposed to a firm-specific basis.

Two additional studies; Brown, Richardson, and Schwager (1987) and Kross, Ro, and Schroeder (1990); provide evidence on the determinants of analyst forecast superiority. Brown, Richardson, and Schwager find that their superiority measure (the ratio of squared forecast errors from a seasonal random-walk time-series model and *Value Line*) is associated with firm size and *ex ante* dispersion of analyst forecasts but not with a proxy for the degree of independent information signals available to the analyst (lines of business). Kross et al. find that the "analyst advantage" is related to earnings variability, the extent of *The Wall Street Journal* coverage, and the analyst's timing advantage. We incorporate firm size, lines of business, and earnings variability as potential explanatory variables in our research design.

Bathke, Lorek, and Willinger (1989) provide indirect evidence that the predictive accuracy of the Brown-Rozeff ARIMA model is positively related to increasing magnitudes of the AR1 parameter and decreasing magnitudes of the MA4 parameter. They show that quarterly earnings of large firms in their sample are forecast more accurately by the BR ARIMA model than are the earnings of small firms. They also provide evidence that large firms possess systematically larger AR1 parameter values and systematically smaller MA4 parameter values relative to smaller firms. We use this evidence to support our inclusion of the estimated AR1 and MA4 parameters in our probit regressions. We also include the magnitude of the Ljung-Box Q-statistic as an indicator of model adequacy to identify those cases where the BR ARIMA model provides a poor descriptive fit for the modeled EPS series.

Lobo (1992) investigates whether a combination forecast model can produce more accurate forecasts of *annual* earnings than a variety of single-source forecasts. He finds that a simple equally-weighted combination forecast model constructed from both I/B/E/S forecasts and predictions from one of the three popular time-series models is more accurate than any forecast obtained from a single source. Lobo (1991) uses OLS regression to build combination models from two, three, or four separate forecast sources.[4] He finds that a combination model estimated by OLS regression with no intercept term and no restrictions on the estimated beta coefficients produces the most accurate forecasts of annual earnings. This complementarity of forecasts from competing sources has also been documented by Lee and Chen (1990) and Elgers and Murray (1992), among others.

In the predictive portion of our research we include a combination model whose weights are determined by the probability measures from the probit regressions. Specifically, if the probability that a given firm will be more accurately forecast by *Value Line* is *p*, then the weight assigned to the VL forecast is *p* and the weight assigned to the Brown-Rozeff forecast is *(1-p)*. An alternative approach to the construction of combination models is to attempt to predict *ex ante* whether a firm's earnings will be more accurately forecast by one source or another. Given the estimated probability, we choose to weight the indicated forecast by 1.0 and to ignore the other forecast source. The results of our investigations follow a brief discussion of our sample.

## 3. Methodology

### 3.1 Sample

We obtained a sample of 110 small firms that received *Value Line* coverage in 1988 and 1989. Specifically, December 31 year-end firms for which complete quarterly EPS data (adjusted for stock splits/stock dividends) were available from the first quarter 1979 to the fourth quarter 1989 were obtained from the Compustat quarterly Tertiary and Full Coverage files. We also required nonmissing data necessary to compute the market value of equity for each quarter in 1988 and 1989. These data requirements initially resulted in 778 firms for analysis. We next selected the smallest 250 firms based on year-end 1988 market values of equity. We chose to focus on small firms given the relative dearth of information on forecast accuracy of these firms relative to larger firms and given the evidence in Bamber (1986) that *Value Line* forecasts are less accurate for these firms vis-à-vis larger firms. Hence, it is precisely for these firms that an improvement in forecast accuracy would be most beneficial.

Of the 250 firms originally selected, 137 received *Value Line* coverage from mid-1987 through 1989. It was necessary for VL to have covered these firms by August of 1987 in order to procure three-quarter ahead forecasts of the first quarter of 1988. Twenty-seven of these firms had stock splits/stock dividends occur in 1988 and/or 1989. We omitted these firms from our final sample due to some ambiguity in determining how *Value Line* adjusted for these changes. Data for the number of lines of business was obtained from the *Directory of Corporate Affiliations*.

### 3.2 Discussion of variables

We attempt to identify potentially discriminating variables identified in the literature as being associated with the forecast superiority of analysts or time series-based predictions (or both). From the research discussed earlier we focus on the following independent variables: firm size (market value of equity), lines of business, the magnitude of the AR1 and MA4 parameters from the BR $(100)X(011)_4$ time-series model, the residual standard error, and the magnitude of the Ljung-Box Q-statistic.

Firm size has been identified as a discriminating characteristic in determining forecast accuracy by *Value Line* (Bamber [1986]) and by the Brown-Rozeff ARIMA model (Bathke, Lorek, and Willinger [1989]). It has also been identified as a determinant of the cross-sectional forecast superiority of VL analysts relative to time series-based forecasts (Brown, Richardson, and Schwager [1987]). Based on these findings, we conclude that firm size can potentially be used to identify those firms whose quarterly earnings will be more accurately forecast by *Value Line* analysts relative to the BR ARIMA time-series model.

The number of lines of business engaged in by the sample firm is our next independent variable. The number of lines of business is likely to be related to the underlying stability of the aggregated time-series of earnings. As firms diversify into alternative business lines, the effect on total earnings from a "shock" to a singular line of business will be mitigated. One effect of this diversification will be to "deseasonalize" the earnings series. Lorek and Bathke (1984) discuss the effects of improper application of a seasonal ARIMA model (like the Brown-Rozeff model) to nonseasonal data. Their findings indicate that forecast accuracy suffers from such misguided attempts. This suggests that *Value Line* forecasts will be more accurate relative to BR ARIMA predictions as lines of business increase.

Evidence from Bathke, Lorek, and Willinger, (1989) indirectly suggests that parameter magnitudes from the Brown-Rozeff ARIMA model are systematically associated with the accuracy of forecasts produced by that model. Specifically, their evidence shows that large firms possess larger AR1 and smaller MA4 parameters than small firms in their sample of NYSE companies. Their large firms, as a whole, also exhibit smaller mean absolute percentage forecast errors relative to their small firm strata. In related research, Branson, Lorek, and Pagach (1995) demonstrate that, in their sample, those firms covered by *Value Line* have regular and seasonal moving-average parameters--from estimated Griffin-Watts $(011)X(011)_4$ ARIMA models--that are negatively associated with forecast accuracy. We investigate whether the estimated parameters are useful in identifying candidate firms for one of our forecast origin sources and suspect, on the basis of the Bathke, Lorek, and Willinger evidence, that as AR1 magnitudes increase and MA4 parameter values decline that the Brown-Rozeff ARIMA model will displace *Value Line* as the more accurate forecast source.

We also include two additional variables obtained from the estimation of the BR ARIMA models in our preliminary analysis. First, based on the results of Kross et al. (1990), we include the residual standard error (SE) as a potential discriminator in our tests and predict that increasing residual standard errors will negatively affect BR ARIMA predictions of quarterly EPS to a greater extent than those produced by VL analysts. Second, we investigate whether Brown-Rozeff model adequacy--as proxied by the magnitude of the Ljung-Box Q-statistic (LBQ)--is related to differential forecast source superiority. If the quarterly earnings series is poorly characterized by the BR ARIMA model, we expect that *Value Line* forecasts will be superior.

## 4. Results

### 4.1 Descriptive profile

For the four quarters of 1988, we determine whether *Value Line* analysts or the Brown-Rozeff $(100)X(011)_4$ ARIMA model provides more accurate EPS forecasts over each of three forecast horizons. Because we are interested in building a prediction model we discard from our analysis those observations with absolute percentage forecast errors (APFE's) exceeding 100% from *both* sources. This results in an omission of 74 of the 440 firm-quarter observations in 1988 for one-quarter-ahead forecasts, 87 (of 440) observations for two-quarter-ahead forecasts, and 81 (of 440) observations for three-quarter-ahead forecasts. In panels A and B of Exhibit 1, we provide descriptive information on the two subgroups for the three forecast horizons.

Inspection of the descriptive results in Exhibit 1 reveals that for the one-quarter-ahead forecast horizon, when earnings are more accurately forecast by the BR ARIMA model (128 of 366 firm-quarters), the earnings are significantly better predicted. The median (mean) absolute percentage forecast errors for the BR firms are: 12.45% (19.94%) for predictions obtained from the BR ARIMA model and 32.35% (41.11%) for forecasts produced by *Value Line* analysts. Similarly, for the majority of cases (238 of 366) when EPS is more accurately predicted by VL, a sizable spread in forecast accuracy from the two competing sources is evident. Specifically, for the VL firms, median (mean) APFE's are 37.82% (46.84%) for forecasts produced by the BR ARIMA model and 11.11% (18.16%) for the *Value Line* predictions. Again, the possibility of exploiting this disparity in forecast accuracy is our primary research objective.

We test for statistically significant differences between the two subgroups in the magnitudes of the variables presented in panels A and B of Exhibit 1. For the one-step-ahead forecasts, we find (using nonparametric Wilcoxon Rank-Sums tests) the following significant differences at the indicated level of probability. For market value (MV), $p < .084$; for the first-order autoregressive parameter (AR1), $p < .006$; for the seasonal moving-average parameter (MA4), $p < .049$; for the Ljung-Box Q-statistic (LBQ), $p < .041$; for the residual standard error (SE), $p < .047$; for lines of business (LOB), $p < .129$; for the absolute percentage forecast error of Brown-Rozeff ARIMA model predictions (BRAPFE), $p < .001$; and for *Value Line* forecast errors (VLAPFE), $p < .001$ as well. The results are striking--only the difference in LOB is insignificant at conventional acceptance levels. All but the differences in MV and LOB are significant at the $p < .049$ level or smaller.

For the two-quarter-ahead forecast horizon a similar pattern is observed. For those instances in which quarterly EPS is more accurately forecast by the BR ARIMA model (144 of 353 times--40.8%), the improvement relative to *Value Line* predictions is impressive. The BR ARIMA forecasts for these cases are more than twice as accurate as VL forecasts. Median (mean) APFE's for the BR firms are 15.2% (21.8%) compared to 36.97% (44.07%) when *Value Line* is the two-quarter-ahead forecast source for these companies. Conversely, for those cases when EPS is more accurately forecast by *Value Line* analysts (209 of the 353 firm-quarters), the VL forecasts are considerably more accurate than those obtained from the Brown-Rozeff model. Median (mean) APFE's are 42.50% (48.85%) when the forecast source is the BR ARIMA model and 11.25% (19.23%) when VL analysts are predicting two-quarter-ahead earnings for these firms.

As discussed previously, we also use non-parametric tests to identify statistically significant differences in the candidate variables. For MV, $p < .097$; for AR1, $p < .005$; for MA4, $p < .036$; for LBQ, $p < .009$; for SE, $p < .003$; for LOB, $p < .001$; for BRAPFE, $p < .001$; and for VLAPFE, $p < .001$ also. Each of the variables is significantly different across the two subgroups at conventional acceptance levels. Except for MV, each variable is significant at the $p < .036$ level or better. The only departure from the results reported for the one-step-ahead predictions is the significance of the LOB variable.

**EXHIBIT 1: Descriptive Profile of Sample Firms**

**Panel A: Brown-Rozeff Firms**

| | **One-Quarter-Ahead Forecasts (n=128)** | | | | **Two-Quarter-Ahead Forecasts (n=144)** | | | | **Three-Quarter-Ahead Forecasts (n=148)** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | *Mean* | *Median* | *Min* | *Max* | *Mean* | *Median* | *Min* | *Max* | *Mean* | *Median* | *Min* | *Max* |
| MV | 1023 | 208 | 15 | 24,736 | 1658 | 221 | 16 | 24,466 | 1308 | 225 | 15 | 24,224 |
| AR1 | .539 | .565 | -.271 | .999 | .533 | .586 | -.443 | .999 | .509 | .566 | -.281 | .999 |
| MA4 | .550 | .603 | -.447 | .999 | .566 | .608 | -.877 | .999 | .553 | .609 | -.820 | .999 |
| LBQ | 8.82 | 8.05 | 0.25 | 25.62 | 9.18 | 9.14 | 0.28 | 24.14 | 9.06 | 8.58 | 0.28 | 24.14 |
| SE | .336 | .171 | .007 | 2.65 | .310 | .152 | .007 | 2.62 | .304 | .144 | .006 | 2.56 |
| LOB | 3.33 | 2 | 1 | 11 | 2.93 | 2 | 1 | 9 | 3.10 | 2 | 1 | 11 |
| BRAPFE | .199 | .125 | 0 | .917 | .218 | .152 | 0 | .962 | .239 | .167 | 0 | .933 |
| VLAPFE | .411 | .324 | .022 | 1 | .441 | .370 | .026 | 1 | .477 | .421 | .024 | 1 |

**Panel B: *Value Line* Firms**

| | **One-Quarter-Ahead Forecasts (n=238)** | | | | **Two-Quarter-Ahead Forecasts (n=209)** | | | | **Three-Quarter-Ahead Forecasts (n=211)** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Variable* | *Mean* | *Median* | *Min* | *Max* | *Mean* | *Median* | *Min* | *Max* | *Mean* | *Median* | *Min* | *Max* |
| MV | 1500 | 264 | 16 | 26,412 | 1181 | 257 | 17 | 26,237 | 1400 | 259 | 16 | 26,122 |
| AR1 | .445 | .490 | -.443 | .999 | .446 | .463 | -.124 | .999 | .468 | .485 | -.093 | .999 |
| MA4 | .612 | .642 | -.822 | .999 | .616 | .673 | -.999 | .999 | .578 | .644 | -.999 | .999 |
| LBQ | 7.85 | 7.18 | 0.28 | 25.12 | 7.93 | 7.56 | 0.25 | 25.62 | 7.99 | 7.65 | 0.25 | 25.62 |
| SE | .379 | .225 | .007 | 2.56 | .397 | .213 | .007 | 2.47 | .397 | .208 | .007 | 2.43 |
| LOB | 3.64 | 3 | 1 | 11 | 4.03 | 3 | 1 | 11 | 3.93 | 3 | 1 | 11 |
| BRAPFE | .468 | .378 | .011 | 1 | .489 | .425 | .013 | 1 | .525 | .474 | .033 | 1 |
| VLAPFE | .182 | .111 | 0 | .966 | .192 | .113 | 0 | .971 | .206 | .148 | 0 | .905 |

The Brown-Rozeff firms are those whose earnings were more accurately forecast by the Brown-Rozeff $(100)X(011)_4$ ARIMA model in a given quarter of 1988 than by *Value Line (VL)*. The *VL* firms are those whose *VL* forecasts were superior to the BR ARIMA model forecasts. Those firms whose forecasts from both sources produced absolute percentage forecast errors exceeding 100% were omitted from the analysis. MV is the market value of equity at the beginning of each quarter (in millions of dollars). AR1 and MA4 are the first-order autoregressive and seasonal moving-average parameters respectively from the BR ARIMA model reestimated each quarter using the most recent 36 quarters of earnings preceding the forecast quarter. LBQ is the Ljung-Box Q-statistic--a measure of residual autocorrelation and a proxy for model adequacy. SE is the residual standard error from the BR ARIMA estimations. LOB is the number of lines of business of the sample firm. BRAPFE and VLAPFE are the absolute percentage forecast errors from the two forecast sources (truncated at 100%). The APFE error metrics are computed as the absolute value of: (Forecasted EPS - Actual EPS)/Actual EPS.

Exhibit 1 also contains similar statistics relevant to three-quarter-ahead earnings forecasts from the two sources. As noted earlier, an even larger percentage (41.2%) of these cases are more accurately forecast by the BR ARIMA model vis-à-vis the shorter forecast horizons. Again, the disparity in forecast accuracy across the two sub-groups is dramatic. Non-parametric tests for differences in variable magnitudes reveal statistically significant (at $p < .10$) differences for MV, LBQ, SE, LOB, BRAPFE, and VLAPFE. Interestingly, the AR1 and MA4 parameters are *not* significantly different across the two subgroups at this forecast horizon (their p-values are .128 and .142 respectively).

### 4.2 Probit regressions

We estimate the following model using probit regression for each of the three forecast horizons independently using the 1988 forecast data:

$$VL = ß_0 + ß_1 MV + ß_2 AR1 + ß_3 MA4 + ß_4 LBQ + ß_5 SE + ß_6 LOB + \varepsilon \quad (1)$$

where:

- VL = 1, if the firm is more accurately forecast by *Value Line*; 0, if more accurately forecast by the Brown-Rozeff $(100)X(011)_4$ ARIMA model,
- MV = Natural logarithm of the market value of equity computed as of the first day of each quarter,
- AR1 = First-order autoregressive parameter obtained from estimating the BR ARIMA model over the most recent 36 quarters of EPS preceding he forecast quarter,
- MA4 = Seasonal moving-average parameter obtained from the estimation of the BR ARIMA model over the same time period,
- LBQ = Magnitude of the Ljung-Box Q-statistic--a chi-squared distributed test statistic that identifies the extent of residual autocorrelation that remains in the modeled EPS series,
- SE = Residual standard error from estimation of the BR ARIMA model,
- LOB = Lines of business of the sample firm, and
- $\varepsilon$ = An independent, normally-distributed error term.

### 4.3 Discussion of Probit regression results

The probit regression results are presented in Exhibit 2. For the one-quarter-ahead forecast horizon, the coefficient estimates for MV, AR1, and MA4 are significant and consistent with the observed sample differences presented in Exhibit 1. Collinearity among the independent variables prevents all of the observed univariate relationships to manifest in a multivariate setting. Only those variables with incremental ability to explain variation in the dependent variable are significant in the probit regressions. The significance of the MV, AR1, and MA4 variables for one and two-quarter-ahead forecast horizons is evidence that the systematic differences across subgroups can potentially be exploited in predicting the more accurate forecast source. The results of out-of-sample predictive ability tests follow in the next section. It is interesting to note that as the forecast horizon lengthens, the significance levels of these variables decreases. Finally, inspection of the three-quarter-ahead probit regression results reveals that only lines of business (and, marginally, market value) is a significant explanatory variable for determining the probability of *Value Line* forecast superiority.

Horizons one, two, and three refer to the one, two, and three-quarter-ahead forecast data respectively. n is the number of firm-quarter observations in 1988 where the predictions from the two forecast sources did not *both* produce an absolute percentage forecast error exceeding 100%. VL equals 1 if the firm's quarterly earnings are more accurately forecast by *Value Line*; 0, if more accurately forecast by the Brown-Rozeff $(100)X(011)_4$ ARIMA model. MV is the natural logarithm of the market value of equity computed as of the first day of each quarter. AR1 is the first-order autoregressive parameter obtained from estimating the BR ARIMA model over the most recent 36 quarters of earnings preceding the forecast quarter. MA4 is the seasonal moving-average parameter obtained from the estimation of the BR ARIMA model over the same time period. LBQ is the magnitude of the Ljung-Box Q-statistic--a chi-squared distributed test statistic that identifies the extent of residual autocorrelation that remains in the modeled earnings series. SE is the residual standard error from estimation of the BR ARIMA model. LOB is the number of lines of business of the firm.

**EXHIBIT 2: Probit Regression Results for One, Two, and Three-Quarter-Ahead Forecast Horizons**

Model: $VL = ß_0 + ß_1 MV + ß_2 AR1 + ß_3 MA4 + ß_4 LBQ + ß_5 SE + ß_6 LOB + \square$

| Horizon One | Horizon Two | Horizon Three |
|---|---|---|

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (n=366) | | | | (n=353) | | | | (n=359) | | |
| Variable | | Estimate | | t-stat | | Estimate | | t-stat | Estimate | | t-stat | |
| Intercept | | -.108 | | -0.29 | | -.371 | | -0.98 | -.321 | | -0.89 | |
| MV | | .117 | | 2.32 | | .093 | | 1.85 | .090 | | 1.84 | |
| AR1 | | -.717 | | -2.97 | | -.612 | | -2.63 | .256 | | -1.11 | |
| MA4 | | .538 | | 2.37 | | .364 | | 1.62 | 127 | | 0.65 | |
| LBQ | | .018 | | -1.23 | | -.020 | | -1.41 | -.020 | | -1.43 | |
| SE | | -.142 | | -0.88 | | -.014 | | -0.08 | .146 | | 0.88 | |
| LOB | | .025 | | 0.91 | | .102 | | 3.47 | .063 | | 2.27 | |
| **Diagnostics:** | | | | | | | | | | | | |
| Likelihood Ratio: | | | | 19.801 | | | 28.731 | | | | 15.764 | |
| | (sig.) | | | (.0030) | | | (.0001) | | | | (.0151) | |
| | | | ***Predicted*** | | | ***Predicted*** | | | ***Predicted*** | | | |
| | **Actual** | **Total** | 0 | 1 | **Actual** | **Total** | 0 | 1 | **Actual** | **Total** | 0 | 1 |
| | **Total** | 366 | 25 | 341 | **Total** | 353 | 92 | 261 | **Total** | 359 | 65 | 294 |
| | 0 | 128 | 12 | 116 | 0 | 144 | 49 | 95 | 0 | 148 | 31 | 117 |
| | | 1 | 238 | 13 | 225 | 1 | 209 | 43 | 166 | 1 | 211 | 34 |
| | 177 | | | | | | | | | | | |

**4.4 Predictive ability tests**

Our goal is to develop a prediction model useful in determining which of two forecast sources is more likely to produce an accurate EPS forecast over three separate forecast horizons. The model is developed from firm characteristics that we have shown to be systematically different across two groups of firms whose earnings are more accurately forecast by one source or the other. The true test of the capability of producing an improvement in forecast accuracy is by conducting out-of-sample predictive ability tests comparing the accuracy of earnings forecasts obtained from competing methods or origins. In our tests we compute absolute percentage forecast errors of predicted 1989 quarterly EPS for our *full* sample of 110 firms. That is, unlike our previous analyses where we omitted those firms poorly forecast by both models, we include all firms in our sample in conducting the predictive ability comparisons.

We compare forecasts obtained in five different ways. First, we use forecasts produced by the BR ARIMA model for every firm, at each quarter, over each forecast horizon. Second, we collected *Value Line* analysts' forecasts of EPS for each company, for all four quarters, over each of the three forecast horizons. Third, we used the estimated parameters from the appropriate probit regression along with 1989 values of the firm-specific independent variables to determine the probability that the more accurate forecast was likely to be produced by *Value Line* or by the BR ARIMA model. If the estimated probability was .50 or higher we used the VL forecast. If the estimated probability was less than .50, we used the time series-based EPS prediction. Fourth, in an *ad hoc* procedure, we conjectured that a combination model (à la Lobo [1991,1992]) with weights determined by the probability estimates obtained from the probit regressions was a reasonable counterpoint to the 100% versus 0% weighting implicit in method three above. Specifically, if the estimated probability of *Value Line* forecast superiority was, say, .70; we weighted the VL forecast by .70 and the BR ARIMA model forecast by .30. Finally, in our fifth method, we apply a simple decision rule that operates as follows: if the previous quarter was more accurately forecast by VL analysts, then we use *Value Line* as the forecast source for the current quarter. Likewise, if the BR ARIMA model produced a more accurate earnings forecast in the preceding quarter, we use a BR ARIMA model forecast for the current quarter. The results of our predictive ability tests are presented in Exhibit 3 and discussed below

.

**EXHIBIT 3: Predictive Ability Tests of 1989 Quarterly Earnings per Share**

| | One-Quarter-Ahead Forecasts | | | | | Two-Quarter-Ahead Forecasts | | | | | Three-Quarter-Ahead Forecasts | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Forecast Origin | Qtr 1 | Qtr 2 | Qtr 3 | Qtr 4 | 1989 | Qtr 1 | Qtr 2 | Qtr 3 | Qtr 4 | 1989 | Qtr 1 | Qtr 2 | Qtr 3 | Qtr 4 | 1989 |
| BR ARIMA Model | .296 | .333 | .383 | .510 | .392 | .379 | .381 | .376 | .543 | .402 | .364 | .340 | .449 | .570 | .417 |
| | (.480) | (.449) | (.453) | (.548) | (.482) | (.475) | (.491) | (.480) | (.552) | (.500) | (.476) | (.460) | (.510) | (.566) | (.503) |
| *Value Line* | .220 | .250 | .175 | .254 | .211 | .247 | .254 | .291 | .348 | .286 | .288 | .286 | .308 | .437 | .341 |
| | (.345) | (.377) | (.323) | (.418) | (.366) | (.404) | (.405) | (.418) | (.486) | (.428) | (.440) | (.442) | (.431) | (.512) | (.456) |
| Probit-Predicted | .220 | .255 | .167 | .259 | .209 | .254 | .249 | .286 | .355 | .286 | .328 | .305 | .287 | .461 | .349 |
| | (.373) | (.380) | (.316) | (.425) | (.373) | (.414) | (.398) | (.415) | (.497) | (.431) | (.456) | (.424) | (.437) | (.518) | (.459) |
| Probit-Weighted | .224 | .250 | .209 | .329 | .258 | .304 | .313 | .278 | .414 | .312 | .364 | .299 | .329 | .450 | .356 |
| | (.399) | (.384) | (.360) | (.467) | (.403) | (.424) | (.442) | (.436) | (.495) | (.449) | (.441) | (.443) | (.462) | (.528) | (.468) |
| Last-Quarter-Source | .219 | .266 | .175 | .254 | .234 | .276 | .261 | .248 | .355 | .292 | .294 | .254 | .342 | .393 | .311 |
| | (.366) | (.387) | (.350) | (.429) | (.383) | (.419) | (.399) | (.423) | (.496) | (.434) | (.442) | (.410) | (.444) | (.498) | (.448) |

The reported numbers are median absolute percentage forecast errors (APFE) computed as: APFE = |(Forecasted EPS - Actual EPS)/Actual EPS|. Mean APFE's (truncated at 100%) are reported in ( ). Number of observations for each quarter equals 110 (440 for 1989). The forecast origins are defined as: BR ARIMA model: one, two, or three-quarter-ahead forecasts produced by estimating the Brown-Rozeff $(100)X(011)_4$ ARIMA model over the most recent 36 quarters of EPS prior to the forecast quarter. The model parameters are estimated on a firm-specific basis and reestimated each quarter. *Value Line*: one, two, or three-quarter-ahead forecasts obtained from the relevant issue of *The Value Line Investment Survey*. Probit-Predicted: the firm-specific forecast from either the BR ARIMA model or *Value Line* is used depending upon the magnitude of the estimated probability that *Value Line* forecasts are more accurate than those produced by the BR ARIMA model. The probability ($p$) is a function of the model variables and the estimated parameters for the relevant forecast horizon (see exhibit 2). Probit-Weighted: forecasts from both sources are used and weighted according to the estimated probability $p$ referred to above as follows: ($p$ x VL forecast + (1-$p$) x BR ARIMA forecast). Last-Quarter-Source: a simple decision rule that uses the forecast source that was observed to be more accurate in the immediately preceding quarter. For example, if in quarter 2, the BR ARIMA model produced an x-step-ahead (x=1,2,3) forecast that was observed *ex post* to have been more accurate than that from *Value Line*, then the BR ARIMA model forecast is used as the forecast source for quarter 3.

In Exhibit 3, the predictive ability results for the three forecast horizons using each of the five sources/methods outlined above are provided. A modest improvement in median forecast accuracy was obtained by applying the probit-predicted model relative to *Value Line* forecasts alone. The limited improvement is no doubt due to the predictive performance of our probit model in identifying EPS series likely to be amenable to BR ARIMA modeling. For 1989 one-quarter-ahead forecasts of all 110 firms (440 firm-quarter observations), the BR ARIMA model was actually more accurate on 131 occasions. However, the probit model only predicted that 36 of the 440 observations would be more accurately forecast by the BR ARIMA model. Unfortunately, of these, only 12 were actually more accurate. Conversely, the model predicted that 404 observations would be more accurately forecast by VL--only 285 of these predictions were correct. The overall classification accuracy, then, was 297/440 or 67.50%. The inadequacy of the probit model's performance is underscored by the fact that, had we merely "predicted" *Value Line* forecasts to be superior for every observation, we would have been correct 309 of 440 times (70.23% of the time).

The predictive ability of the combination model, while better than time series-based forecasts, was disappointing. Given the results in Lobo (1991, 1992), which examine annual EPS forecasts, we anticipated that the combination approach would prove fruitful in predicting quarterly EPS. However, only in quarter two is this model competitive with the other candidates. As an additional performance check, we replicated Lobo's combination models on our sample and performed identical predictive ability tests as those presented in Exhibit 3.[5] With respect to the equally-weighted model (the Lobo1 model) over a one-quarter-ahead forecast horizon, only the BR ARIMA predictions were poorer. For the OLS regression-based weightings, however, the Lobo2 model was superior to both BR ARIMA model predictions (in all four quarters) and to the probit-weighted model forecasts in quarters three and four (and for the year). Still, neither model was as accurate as *Value Line* predictions alone (or the probit-predicted forecasts) in any quarter. It could be that the wide disparity in forecast accuracy from our two sources for the two subgroups of firms we examine does not lend itself well to combinatorial modeling. While untested (and certainly a possible avenue for future research), we guess that the *difference* in forecast accuracy from the two forecast sources we investigate decreases as the size of the firms subject to the examination grows larger.

The simple decision rule--use the same forecast source that produced the previous quarter's more accurate forecast--fared reasonably well. This method outperformed BR ARIMA forecasts (and predictions from the equally-weighted model) by a wide margin and outpredicted both the probit-weighted and the OLS regression-based models in three of four quarters and for the year.

The probit-predicted results notwithstanding, the potential for improvements in forecast accuracy does exist. If we could correctly classify each firm by forecast source that produced the more accurate prediction, a median absolute percentage forecast error for our 440 firm-quarter sample of 15.4% would obtain--a 5.7% real improvement over the forecasts produced by *Value Line* analysts.

We present similar evidence for two-quarter-ahead forecasts. The probit-predicted model forecasts were superior to *Value Line* predictions in two of four quarters (quarters 2 and 3) and as good for the year. The last-quarter-source decision rule also performed adequately, superior to both *Value Line* and the probit-predicted forecasts in quarter 3. The two combination models styled after Lobo (1991, 1992) performed poorly, outpredicting only the BR ARIMA model for the year. Interestingly, the model utilizing regression-based weights for determining the combination forecast performed worse than the equally-weighted scheme.

The inability of the probit model to correctly classify a majority of the EPS series amenable to ARIMA modeling was, again, its downfall. For the 440 forecasts, 164 were actually more accurately forecast by the BR ARIMA model. The probit model predicted that 119 of the 440 quarterly EPS predictions would fall into this category. Of the 119, only 49 were truly better forecast by the mechanical model vis-à-vis *Value Line*--a success rate of 41.2% (a slight improvement over the one-quarter-ahead performance). Of the 321 cases identified as likely to be more accurately forecast by VL, only 206 were so. This produces an overall success rate of 57.95% (255 of 440). We are encouraged by the fact that, despite low classification rates, the overall forecast accuracy is indistinguishable from *Value Line's*. This suggests that an opportunity for marked improvement exists by better exploiting the observed differences in the two subgroups. Had we been able to correctly classify all firms as to ideal forecast source, a median absolute percentage forecast error of 20.0% would have been produced--a substantial reduction in forecast error.

Three-quarter-ahead predictive ability test results are also presented in Exhibit 3. The most striking finding is the dominant performance of the decision-rule approach. In quarters 2 and 4 and for the year as a whole this method of producing lower median percentage forecast errors was unmistakably superior. *Value Line* analyst forecasts were slightly more accurate than those forecasts produced by the probit-predicted and the probit-weighted model. As for the two combination models we derive from Lobo's research, neither performed well. The equally-weighted mode l out predicted BR ARIMA forecasts in each quarter but the regression-weighted model was beaten by the ARIMA model predictions in quarters one and two as well as for the year as a whole.

As was the case with one and two-quarter-ahead forecast horizons, the probit model was not particularly successful in identifying EPS realizations that were actually more accurately forecast by the BR ARIMA model. Of the 440 firm-quarter forecasts, only 82 were judged likely to be more accurately forecast by the Brown-Rozeff model. Forty-one of these predictions proved correct. Unfortunately, fully 188 of the 440 were better off forecast by the mechanical time-series model. Hence, of the 358 classified as VL firms, only 211 proved to be so. The classification accuracy for the probit model, then, was 57.27% (252/440). Again, we argue that the potential for dramatic improvement in aggregate forecasting of EPS exists and that we have only begun to explore methods of tapping this potential. For three-quarter-ahead forecast horizons, "perfect insight" would have produced a median APFE for our 440 observations in 1989 of 22.0%--a 12.1% improvement over *Value Line*.

**5. Conclusion**

We examined the time-series properties and other firm-specific characteristics of a sample of small firms. We find in those quarters (in 1988/89) where these firms are more accurately forecast by the Brown-Rozeff $(100)X(011)_4$ ARIMA model relative to *Value Line* analysts that there are significant, systematic differences between these firms and those more accurately forecast by *Value Line*. Specifically, we find that those firms whose earnings are more accurately forecast by the BR ARIMA model are smaller in size and are engaged in fewer lines of business. They also possess significantly larger AR1 parameters, smaller MA4 parameters, smaller residual standard errors, and slightly higher levels of residual autocorrelation. The last finding is somewhat puzzling. We had conjectured that increased levels of autocorrelation remaining in the earnings series would be indicative of poor model fit and, thus, poor predictive ability. This appears to be another case of the descriptive-predictive paradox--the finding that models which display superior descriptive fit relative to other models are occasionally less accurate predictors of out-of-sample data.

Our attempt to exploit these observed differences by constructing a probit regression-based prediction model that generates a probability that a given firm's EPS in a given quarter will be more accurately forecast by *Value Line* yielded reasonably impressive results. However, the potential for dramatic improvement exists if a model can be developed that can more accurately classify *ex ante* those firms likely to be well-forecast by the BR ARIMA model. Had we been able to correctly classify firms, the improvement in forecast accuracy vis-à-vis *Value Line* forecasts alone is striking. One, two, and three-quarter-ahead forecasts of quarterly EPS would have improved by 5.7%, 8.6%, and 12.1% (in real terms) respectively.

Our research is based on a sample of small firms, hence, the findings may not be generalizable to larger firms. We chose to concentrate on small firms because we believe that it is for these firms that improvements in forecast accuracy will be of most benefit to the business research community. Our results are also confined to the 1988/1989 time period and, thus, are specific to this time frame. Whether our findings recur over other time horizons is a question left for further study.

**6. Suggestions for Future Research**

Our study uses two sources of earnings per share forecast—those obtained from the Value Line Investment Survey and those produced by application of the Brown-Rozeff ARIMA model. Many other sources of EPS forecasts exist and may be exploitable in the same manner in which we combined forecasts from our two sources. As Lobo (1991) demonstrates, combinations of forecasts from multiple sources can provide minimization of forecast errors when compared to those errors produced by single-source EPS forecasts. With the veritable explosion of data that has become available, it may be fruitful to examine a class of firms for which such data was not available in the time period we examined in our study. Our evidence, and that of much prior research, tells us that smaller firms are most likely to be the source of significant forecast error—both from the analyst community and from statistical forecast models.

Future researchers could also consider alternative weighting schemes for the forecast source in combination source models. We experimented with several such alternatives but by no means exhausted the potential methods by which the forecasts we used (and those additional sources that could augment those we used) could be combined. Additionally, earnings is but one variable (albeit an important one) that is the object of prediction—sales, for example, is a frequently forecasted balance used in capital markets valuation models and our approach could be easily adapted to a search for more accurate estimates of future revenue. Further, in a financial statement audit setting, a plethora of forecasts are frequently generated for comparison to reported account balances in the conduct of analytical procedures. In this context, more accurate models could provide a significant contribution to both the body of research investigating analytical procedures as well as to professional practice.

[We are indebted to the workshop participants of Louisiana State University, North Carolina State University, and the University of Texas-Arlington. We also received helpful comments from Jon Bartley, Jennifer Francis, John Hassell, and Ken S. Lorek. Data used in this study are publicly available. A list of the sample companies is available from the last author along with details of additional tests mentioned in the paper. ]

**References**

1 Bamber, L, "The Information Content of Annual Earnings Releases: A Trading Volume Approach," *Journal of Accounting Research* (Spring): 40-56, 1986.

2 Bathke, A. W., K. S. Lorek, and G. L. Willinger, "Firm Size and the Predictive Ability of Quarterly Earnings Data," *The Accounting Review* (January): 49-68, 1989.

3 Branson, B.C., D. Guffey, and D.P. Pagach, "Information Conveyed in Announcements of Analyst Coverage," *Contemporary Accounting Research* (Summer) 1998.

4 Branson, B.C., K. S. Lorek, and D. P. Pagach, "Evidence on the Superiority of Analysts' Quarterly Earnings Forecasts for Small Capitalization Firms," *Decision Sciences* (March/April) 1995.

5 Brown, L., R. Hagerman, P. Griffin, and M. Zmijewski, "Security Analyst Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings," *Journal of Accounting and Economics* (April): 61-87, 1987.

6 Brown, L. and M. Rozeff, "Univariate Time-Series Models of Quarterly Earnings per Share: A Proposed Model," *Journal of Accounting Research* (Spring): 179-189, 1979.

7 Brown, L., G. D. Richardson, and S. J. Schwager, "An Information Interpretation of Financial Analyst Superiority in Forecasting Earnings," *Journal of Accounting Research* (Spring): 49-67, 1987.

8 *Directory of Corporate Affiliations*. New York, NY: National Register Publishing. 1990.

9 Elgers, P. and D. Murray, "The Relative and Complementary Performance of Analyst and Security-Price-Based Measures of Expected Earnings," *Journal of Accounting and Economics* (June/September): 303-316, 1992.

10 Francis, J. and D. Philbrick, "Analysts' Decisions as Products of a Multitask Environment," *Journal of Accounting Research* 31(2): 216-230, 1993.

11 Foster, G., "Quarterly Accounting Data: Time-Series Properties and Predictive Ability Results," *The*

*Accounting Review* (January): 1-21, 1977.
12 Griffin, P., "The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence," *Journal of Accounting Research* (Spring): 71-83, 1977.
13 Kross, W., B. Ro, and D. Schroeder, "Earnings Expectations: The Analysts' Information Advantage," *The Accounting Review* (April): 461-476, 1990.
14 Lee, C. J. and C. Chen, "Structural Changes and the Forecasting of Quarterly Accounting Earnings in the Utility Industry," *Journal of Accounting and Economics* (July): 93-122, 1990.
15 Lobo, G. J., "Alternative Methods of Combining Security Analysts' and Statistical Forecasts of Annual Corporate Earnings," *International Journal of Forecasting* 7: 57-63, 1991.
16 ________, "Analysis and Comparison of Financial Analysts', Time Series, and Combined Forecasts of Annual Earnings," *Journal of Business Research* 24: 269-280, 1992.
17 Lorek, K. S. and A. W. Bathke, Jr., "A Time-Series Analysis of Nonseasonal Quarterly Earnings Data," *Journal of Accounting Research* (Spring): 369-379, 1984.
18 Schipper, K., "Commentary on Analysts' Forecasts," *Accounting Horizons* (December): 105-121, 1991.
19 Watts, R., "The Time-Series Behavior of Quarterly Earnings." Working Paper. University of Newcastle. 1975.

## Endnotes

1. See, for example, Brown, Hagerman, Griffin, and Zmijewski (1987a); Brown, Richardson, and Schwager (1987); and Kross, Ro, and Schroeder (1990).
2. Recent research has posited that analysts may not focus on producing accurate EPS forecasts. Francis and Philbrick (1992) develop a model in which analysts' primary objective is the maintenance of management relations *not* in accurate earnings predictions per se. In their model, analysts have incentives to provide optimistic forecasts. For those companies whose earnings are poorly forecast by *Value Line*, we find that the forecasts are, on average, overly optimistic. Specifically, we find that for one, two, and three-quarter-ahead forecasts in the 1988 estimation period where the BR ARIMA model is more accurate than those published by *Value Line*, that the median BR forecasts are essentially unbiased but the median VL forecasts are optimistic by 7.57%, 13.07%, and 20.29% for the three forecast horizons respectively. Details of these supplementary tests are available from the authors.
3. The models examined are the seasonal random-walk with drift and the ARIMA models introduced by Foster (1977); Brown and Rozeff (1979); and Griffin (1977) and Watts (1975).
4. Lobo (1991) uses the following sources for his combination models: I/B/E/S consensus annual forecasts, one, two, three, and four-quarter-ahead forecasts obtained from *Value Line*, annual earnings forecasts from the random-walk with drift time-series model, and aggregated quarterly forecasts from the ARIMA model developed by Griffin (1977) and Watts (1975).
5. Lobo (1991, 1992) develops composite forecasting models derived from predictions generated by quarterly ARIMA models and financial analysts. We generalize from Lobo's findings to construct two additional combination models as benchmarks to which we can compare the predictive ability of our models/decision rules. The findings of Lobo (1992) suggest that a simple average of the forecasts produced by a time-series model and *Value Line* or I/B/E/S analysts can outpredict forecasts from any single source. We construct a combination model for quarterly forecasts at each horizon as the average of the VL forecast and the BR ARIMA model prediction. We term this the Lobo1 model. We also develop a combination model based on Lobo's (1991) findings that OLS regression-determined weights can be used to construct a more accurate forecast model than the predictions generated by any single source. We rely on his evidence that a model estimated with no intercept term and no restrictions on the beta coefficients yields the most accurate forecasts. We estimate the following model (the Lobo2 model) for each quarter and forecast horizon in 1989:
   Actual EPS = $ß_1$ BR ARIMA Forecast + $ß_2$ *Value Line* Forecast + $\varepsilon$ (A1)

**Notes**