Neural Network Earnings per Share Forecasting Models: A Comparative Analysis of Alternative Methods

Wei Zhang  
School of Business, Clarkson University, Potsdam, NY 13699, e-mail: weizhang@clarkson.edu

Qing Cao  
University of Missouri-Kansas City, Henry W. Bloch School of Business and Public Administration, 5100 Rockhill Rd., Kansas City, MO 64110-2499, e-mail: caoq@umkc.edu

Marc J. Schniederjans†  
University of Nebraska-Lincoln, College of Business Administration, Lincoln, NE 68588-0491, e-mail: Mschniederjans1@unl.edu

ABSTRACT

In this paper, we present a comparative analysis of the forecasting accuracy of univariate and multivariate linear models that incorporate fundamental accounting variables (i.e., inventory, accounts receivable, and so on) with the forecast accuracy of neural network models. Unique to this study is the focus of our comparison on the multivariate models to examine whether the neural network models incorporating the fundamental accounting variables can generate more accurate forecasts of future earnings than the models assuming a linear combination of these same variables. We investigate four types of models: univariate-linear, multivariate-linear, univariate-neural network, and multivariate-neural network using a sample of 283 firms spanning 41 industries. This study shows that the application of the neural network approach incorporating fundamental accounting variables results in forecasts that are more accurate than linear forecasting models. The results also reveal limitations of the forecasting capacity of investors in the security market when compared to neural network models.

Subject Areas: Methodological Areas: Artificial Neural Network, Comparative Analysis, and Forecasting Methods; Functional Areas: Accounting and Finance.

INTRODUCTION

Forecasting earnings per share (EPS) is an important task for both outside investors and internal managers. Outside the firms, investors use these forecasts as a basis to form profitable investment portfolios. Inside the firms, managers use these forecasts
for a host of critically important decisions including operational budgeting, capital investment, and other resource allocation decisions. Accuracy in these forecasts, therefore, is essential for both optimum portfolio management in capital markets and optimum resource allocation within a firm (e.g., Jarrett, 1990; Elgers, Lo, & Murray, 1995).

Selecting a forecasting methodology is, in itself, a major decision for investors and for managers. Earlier research on the accuracy of alternative forecasting methods has been in one of the two categories: (1) comparing different statistical or mechanical forecasting methods, and (2) comparing statistical models with judgmental analysts’ forecasts (see Lobo & Nair, 1990, and Williams, 1995, for a review on these two genres of research). Building on these two lines of research, later literature suggests that a combination of the statistical forecasts and financial analyst forecasts will produce more accurate results (Lobo & Nair, 1990; Conroy & Harris, 1987; Guerard, 1987). In this paper, we will focus on the first genre.

The considerable amount of research on statistical forecasting models in the literature has created the need for varied means of categorization to identify and select models (for a comprehensive review on the mechanical forecasting methods used in different domains, an appendix with a taxonomy of this literature is available from the authors). One means of basic categorization of the statistical forecasting methods used to aid in model selection is to identify models as either being linear or nonlinear. Another categorization of statistical forecasting models is to refer to them as being either “univariate” (i.e., using one independent variable in the model) or “multivariate” (i.e., more than one independent variable). In the context of EPS forecasts, to date, the majority of the statistical methods used in forecasting EPS are linear (for a classification of mechanical EPS forecasting methods, an appendix is available from the authors). For example, the family of time-series forecast models widely used in forecasting EPS belongs to the linear forecasting category. More recently, researchers have begun considering the nonlinear nature of the financial data and have incorporated nonlinearity into their forecasting models. For example, Callen, Kwan, Yip, and Yuan (1996) noted that quarterly EPS is nonlinear financial data, and therefore a neural network forecasting approach is appropriate. Moreover, in EPS forecasting literature, more research has been conducted in the univariate category than in the multivariate category. In the business-context of forecasting EPS, Abarbanell and Bushee (1997) refer to these variables as “fundamental accounting variables” and include such items as accounts receivables, inventory, and capital expenditures. Their research on EPS has shown that multivariate models are more efficient than univariate ones in a linear context.

The purpose of this paper is to examine the accuracy of a nonlinear approach (i.e., neural network or NN) in forecasting quarterly EPS. In order to do this, we will make forecast comparisons between NN methods and a variety of ARIMA (AutoRegressive Integrated Moving Average) statistical linear models, utilizing both univariate and multivariate fundamental accounting variables (Box & Jenkins, 1977). In the business-context of forecasting EPS, a set of accounting variables have been identified as useful in forecasting future earnings either by statistical methods (Ou & Penman, 1989) or by a thorough search of financial press reports (Lev & Thiagarajan, 1993). These variables, such as accounts receivables, inventory,
and capital expenditures, are referred to as “fundamental accounting variables.” Because these variables can signal the security values and future earnings of the firm (Abarbanell & Bushee, 1997; Beneish, Lee, & Tarpley, 2001), we include these fundamental variables in our multivariate forecasting models. In this paper, we further examine how the market responds to the forecast error for each model to determine whether investors follow the most accurate forecasting method.

Our paper contributes to the neural network literature by documenting evidence that neural networks produce more accurate forecasts of quarterly EPS than do linear models, and this is especially the case when fundamental accounting variables are incorporated. The NN has been successfully used in prediction or forecasting studies in all functional areas of business, including accounting (Lenard, Alam, & Madey, 1995), economics (Hu, Zhang, Jiang, & Patuwo, 1999), finance (Etheridge, Sriram, & Hsu, 2000; Bruce & Michael, 1998), management information systems (Zhu, Premkumar, Zhang, & Chu, 2001), marketing (Papatla, Zahedi, & Zekic-Susac, 2002), and production management (Kaparthi & Suresh, 1994). In one comparative analysis study after another (e.g., Desai & Bharati, 1998; Bhattacharyya & Pendharkar, 1998; Jiang, Zhong, & Klein, 2000), NN consistently outperformed or is more accurate at predicting or forecasting than other more traditional quantitative methods. Unfortunately, there have been some exceptions reported. Tam and Kiang (1992), Ainslie and Dreze (1996), and West, Brockett, and Golden (1997) all found that NN did not do well when compared to traditional statistical methods. One of these exceptions is a study by Callen et al. (1996) where the neural network approach did not produce superior quarterly EPS forecasts. They felt that in the business-context of forecasting EPS, the performance of the neural network model was less accurate than other ARIMA time-series models of forecasting. In our paper, the neural network method we employ takes into account the impact of fundamental accounting variables on future EPS, yielding evidence supporting the superiority of the neural network approach in forecasting quarterly EPS. Our paper therefore highlights the importance that NN as a forecasting method should incorporate broader sets of information.

Our paper contributes to the EPS forecasting literature by addressing the nonlinear manner by which the value-relevant accounting information is incorporated into future quarterly EPS. Abarbanell and Bushee (1997) suggest a nonlinear relationship between the future EPS and some fundamental financial signals, and we test this nonlinearity by employing NN. Additionally, while most studies on EPS forecasts consider nonlinearity and multivariate models separately, we simultaneously consider these two aspects. We present evidence that more accurate forecasts of quarterly EPS can be obtained from the multivariate nonlinear forecast model than from the models considering nonlinearity and the multivariate method separately. Furthermore, the result that the multivariate NN model outperforms all other methods is robust across quarters, years, and industries.

The paper, however, also points out that the security market does not seem to use the most accurate forecast method in setting its earnings expectations. This provides an exploitable opportunity for forming profitable investment strategies. The findings of our paper can also aid the earnings forecasts of small firms which financial analysts do not tend to cover (Branson, Lorek, & Pagach, 1995) and contribute to a composite forecast methodology suggested by the literature (Conroy...
If the superiority of NN method vis-à-vis financial analysts’ judgmental forecasts can be documented by future research, then our findings will have implications for improving the quality of the earnings forecast service provided by the financial analyst profession.

PRIOR LITERATURE AND HYPOTHESES

Neural Network Applications in Forecasting Financial Data

Due to the variety of forecasting methodologies, different methods are usually examined using a “comparative analysis” to show which is more accurate, given a particular set of methods and a particular business-context (e.g., forecasting EPS, stock prices, and so on). Seeking to achieve greater levels of accuracy forecasters are always looking for new quantitative forecasting methods to compare to those they use. One methodology that has received much attention is artificial “neural networks” (NN). For a review of the basics of NN see Chauvin and Rumelhart (1995).

According to Kim, Kim, and Lee (2003), a new trend of combining multiple classifiers has emerged to improve classification results in data mining. For instance, Spangler, May, and Vargas (1999) created an audit matrix methodology for evaluating the performance of three popular data-mining techniques in multiple classification problems, namely, linear discriminant analysis, neural networks, and decision tree induction. The results of their study show that neural networks give the best overall results for the largest multiple classification cases. In their comparative study, Zhu et al. (2001) built intrusion detection systems using various data mining methods for network administrators to deal with network security attacks. Results show that data mining methods and data proportion have a significant impact on classification accuracy and NN-based intrusion detection systems outperform the inductive learning method in the classification. By capturing nonlinear dependencies between the combined classifiers and individuals, Zhu, Beling, and Overstreet (2002) proposed a Bayesian framework (a type of neural networks for classification) for constructing combinations of classifier outputs. Empirical results from the combination of credit scores generated from four different scoring models show that the NN based (Bayesian) multiple classifiers improve overall classification results. Kim et al. (2003) presented a methodology for predicting customers’ purchasing behaviors in an e-commerce setting by combining multiple classifiers. The NN-based method shows better performance than individual classifiers and other known combining methods. In summary, studies in data mining have shown that NN-based models are viable choices in constructing ensembles to combine multiple classifiers to deal with forecasting problems and these comparative studies also show that NN-based models outperform traditional multiple classification methodologies.

The neural network model has been used for a wide range of forecasting tasks, and the literature in this area is growing (see Zhang et al., 1998). In financial data forecasting, examples of the application of the neural network approach include (but are not limited to) the research conducted by Hutchinson, Lo, and Poggio...
While some studies found that the neural network outperforms the traditional statistical methods, others provide evidence that they did not. In the Callen et al. (1996) study, a comparison between the forecasting ability of a family of univariate linear time-series ARIMA models and a neural network model was examined. From a sample of 296 NYSE firms, they used EPS data spanning a period of 89 quarters. A rolling window of 40 quarters was used to identify the forecast models, and 49 forecasted EPS were compared with their actual EPS. They found that the linear Brown-Rozeff and Griffin-Watts univariate time-series models generated more accurate EPS forecasts than neural network models. They argue their findings indicate that the performance of the neural network model may be business-context sensitive, and there is no clear evidence that the neural network approach dominates the standard statistical methods of ARIMA in forecasting quarterly EPS.

The results of Callen et al. (1996) pose a potential forecasting limitation of NN in supporting decisions based on EPS forecasts and, as such, warrant further investigation and clarification. Specifically, Callen et al. (1996) state that univariate neural network models are not necessarily superior to linear ARIMA models in the context of forecasting quarterly EPS, even when the quarterly EPS data is financial, seasonal, and nonlinear. However, we are interested in finding out whether Callen et al.’s (1996) assertion still holds if we include into the forecasting models fundamental accounting variables. Given prior research’s findings that fundamental accounting variables have predictive power with respect to earnings and that these accounting variables may be related to future earnings in a nonlinear manner, we expect that Callen et al.’s (1996) assertion regarding the lack of superiority of the NN method will not hold for the multivariate models.

In our paper, we incorporate fundamental accounting variables into the EPS forecasting models. We construct multivariate time-series models, thus placing our forecast models in a richer context than the prior studies. While there is still debate surrounding the uncertainty and theoretical foundation of the neural network, Gorr (1994) points out that it would be appropriate to use the neural network approach for multivariate time-series forecasting. Thus, we expect that by using the neural network method that takes into account the implications of fundamental accounting variables for future EPS, our research is likely to produce different results on the accuracy of the neural network approach in forecasting quarterly EPS than that documented by Callen et al. (1996).

FUNDAMENTAL ACCOUNTING VARIABLE ANALYSIS AND THE NONLINEARITY OF ACCOUNTING INFORMATION

Penman (1992, p. 465) defines fundamental accounting variable analysis or simply “fundamental analysis” as a process involving “the determination of the value of securities from available information, with a particular emphasis on accounting information.” Research conducted by Lev and Thiagarajan (1993) documents that a list of fundamental accounting variables, claimed by financial analysts as value-relevant in security valuation, possesses incremental explanatory power with
Neural Network Earnings per Share Forecasting Models

respects to excess returns. Using a similar set of fundamental accounting variables but adopting a different view on the role of fundamental analysis (i.e., fundamental analysis should be aimed at predicting EPS but not explaining returns), Abarbanell and Bushee (1997) tested the explanatory power of the fundamental accounting variables with respect to changes in future EPS changes. They found strong associations between individual fundamental signals (i.e., the fundamental accounting variables provide information that act to “signal” forecasters of possible future events) and future EPS changes. Thus, specific financial data can be useful in explaining security returns and predicting EPS. Subsequent research has shown that investment strategies based on the yearly fundamental analysis information does earn significant abnormal returns (Abarbanell & Bushee, 1998; Piotroski, 2000).

In our study, we follow the view of Abarbanell and Bushee (1997) and focus on the predictive power of fundamental variables with respect to future EPS. Instead of fitting the fundamental variables in a model where future earnings are the dependent variable and analyzing the goodness-of-fit measure, we gauge the predictive power of the variables by examining the forecast error for a holdout sample. Research has long found that a model with a solid goodness-of-fit measure may not perform as well in forecasting tasks (Watts & Leftwich, 1977), and this has been described as the “descriptive-predictive paradox” or “regression fallacy” (Lorek & Willinger, 1996). Furthermore, different from Abarbanell and Bushee (1997) where a cross-sectional methodology is used, we use firm-specific forecast models to allow the weights on the fundamental accounting variables to vary from one firm to another. A firm-specific approach is further warranted given Abarbanell and Bushee’s (1997) findings that industry membership seems to condition some of the relations between fundamental accounting variable information and future EPS.

The use of the NN method in this forecasting setting is appropriate because it addresses a nonlinearity issue suspected by Abarbanell and Bushee (1997). Abarbanell & Bushee (1997) found that some identified fundamental financial signals are not significantly related to future EPS as hypothesized, and suggest a nonlinear relationship between the future EPS and some fundamental financial signals. To address this possibility, they tested to see if future EPS are related (1) to the quadratic forms of fundamental accounting variables, (2) to the interaction terms between fundamental financial signals and current EPS, or (3) to the selected pair-wise interactions of fundamental accounting variables. They detected no presence of nonlinearity. In our paper, we circumvent the difficult task of specifying a nonlinear relation between fundamental financial signals and future EPS by resorting to the arbitrary function mapping ability of the neural network model (Zhang, Patuwo, & Hu, 1998), where the underlying unknown mechanism through which the empirical data are generated can be identified by a repetitive learning process. As Church and Curram (1996, p. 261) point out, NN can be used as a regression tool to learn the relationship between a set of explanatory variables and the dependent variable, and in doing so, is used in a similar way to the ordinary least squares (OLS) method allowing highly nonlinear relationships to be fitted if they exist. In a sense, we employ a statistical procedure, as opposed to a predetermined theoretical specification, to account for the potential nonlinear relations between fundamental financial signals and future EPS.
Our choice of using the NN method to address the factor of nonlinearity is based on an extensive review of literature. Our research has shown that for univariate models, NN models outperform other nonlinear approaches in 6 studies; and for multivariate models, NN models achieve better forecasting results than other nonlinear methods in 16 studies. Among the studies that we are aware of, there is no evidence that NN models are outperformed by other nonlinear methods. We conclude that NN models generally perform better than other nonlinear models, and as such, we choose NN as opposed to other nonlinear methodologies to construct our nonlinear forecasting models.

**Hypotheses**

Using a database of corporation EPS, we undertake to statistically test and comparatively evaluate the forecasting accuracy of two sets of forecasting models, linear methods and nonlinear NN methods, and their use of fundamental accounting variables based on the (null) hypotheses that:

- **H1**: There is no forecasting accuracy difference in linear methods and nonlinear NN methods.

The H1 hypothesis requires a comparison between univariate linear with univariate NN, and a comparison between multivariate linear with multivariate NN models. Thus, for purposes of statistical testing we break the above hypothesis into two parts:

- **H1a**: There will be no forecasting accuracy difference in univariate linear and univariate NN nonlinear models when fundamental accounting variables are absent from the univariate models.

- **H1b**: There will be no forecasting accuracy difference in multivariate linear models and multivariate nonlinear NN models when fundamental accounting variables are present in the multivariate models.

We expect that, based on prior research, the consideration of nonlinearity in forecasting quarterly EPS will cause an improvement in forecasting accuracy, resulting in a rejection of the null H1.

Because fundamental accounting variables have predictive power with respect to future earnings and the relation between fundamental accounting variables and future EPS may be nonlinear, we further expect that the multivariate nonlinear NN models will result in the most accurate forecast as they simultaneously consider nonlinearity and fundamental accounting variables. With the other models in contrast, these two aspects can be considered separately a third part of H1 and stated as:

- **H1c**: There will be no forecasting accuracy difference between multivariate NN models and other forecasting models (i.e., univariate linear, univariate NN, and multivariate linear).

After we compare the accuracy of alternative forecasting methods, we try to determine what forecasting method the investors are most likely to follow by
conducting a market return test employed by Foster (1977) and Bathke and Lorek (1984). Deviation from the investors’ earnings expectation will trigger investors’ reevaluation of firm value and cause the occurrence of unsystematic (or abnormal) returns. Here, the abnormal returns are a proxy for investors’ revision of firm value subsequent to the newly released actual earnings. For example, if the firm’s actual earnings beat the earnings expectation, then investors will revise the firm’s share price upward and positive abnormal returns will result; on the other hand, if the firm’s earnings fall short of the expectation, then investors will revise the share price downward and negative abnormal returns will result. A significant positive relation between the forecast error and abnormal returns for a forecasting method will be indicative of the fact that investors have relied upon the method in setting earnings expectations. The more closely the investors follow a certain forecasting method, the higher the correlation between the forecast error and the market’s valuation revision (measured as abnormal returns) for that method.

Using corporation EPS, we undertake to statistically test and comparatively evaluate the degrees to which investors utilize different forecasting models based on the (null) hypothesis that:

\[ H_2: \text{There will be no difference in the forecasting methods in respect of capturing earnings expectations in capital markets.} \]

As previously stated research has suggested, the greatest association between the forecast error and abnormal returns will be present for the forecasting method that the market actually uses. Thus, \( H_2 \) can also be rephrased as: there is no difference in the associations between forecast errors and abnormal security returns for the different forecasting models.

The \( H_2 \) hypothesis requires a comparison of association measures across the four forecasting models: univariate linear, univariate NN, multivariate linear, and multivariate NN models. If investors choose a forecasting method based on its accuracy, the most accurate forecasting model (i.e., the multivariate NN model) will exhibit the strongest association between investors’ revision of firm value (i.e., abnormal security returns) and the forecast error. We therefore expect to reject the null \( H_2 \).

Detecting the method the investors use in the capital market helps address the question of whether the decision makers in the capital market use all the relevant information in decision making. If we find evidence that the strongest association between forecast error and abnormal security returns is present for a less accurate forecast method, then obviously investors have not followed the most accurate forecast model. Failure to rely upon the most accurate forecasting model will point to the limitations in investors’ decision making.

At a more practical level, the market’s failure to use all relevant information provides an exploitable opportunity to form profitable portfolios. As a simple illustration, suppose the market expects a firm’s EPS to be 17 cents while a more accurate method predicts the EPS to be 12 cents. Later, the actual earnings are announced to be 12 cents, triggering a decline in the firm’s share price because the actual EPS falls short of the market’s expectation of 17 cents. For investors who have the more accurate forecast, they can wait until after the earnings release to buy the firm’s shares (or sell the firm’s shares prior to the earnings release).
Table 1: A two-by-two research design assessing the comparative performances for linear and artificial neural network (NN) models in forecasting quarterly EPS.

<table>
<thead>
<tr>
<th>Fundamental accounting variables</th>
<th>Linear Models</th>
<th>Nonlinear NN Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>absent</td>
<td>Univariate Linear Forecast</td>
<td>Univariate Neural Network Forecast</td>
</tr>
<tr>
<td></td>
<td>{Category 1}</td>
<td>{Category 3}</td>
</tr>
<tr>
<td></td>
<td>{Models: M1.1, M1.2, M1.3, M1.4}</td>
<td>{Models: M3}</td>
</tr>
<tr>
<td>present1</td>
<td>Multivariate Linear Forecast</td>
<td>Multivariate Neural Network Forecast</td>
</tr>
<tr>
<td></td>
<td>{Category 2}</td>
<td>{Category 4}</td>
</tr>
<tr>
<td></td>
<td>{Model: M2.1, M2.2}</td>
<td>{Model: M4}</td>
</tr>
</tbody>
</table>

1We include the following fundamental accounting variables: INV (inventory); AR (accounts receivables); CAPX (capital expenditure); GM (gross margin); SA (selling and administrative expenses); ETR (effective tax rate); LF (labor force).

Therefore, if a group of investors have the advantage of possessing more accurate earnings forecasts, these investors may easily “beat the market” by utilizing the more accurate forecasts.

METHODOLOGY

To test the basic research hypotheses on forecasting quarterly EPS, we compare the performance of four categories of model specifications. Specifically, we construct four categories of models using the linear and nonlinear (NN) methods, and examine both situations with the fundamental accounting variables both absent and present. The four categories of models are defined in this section below, and are conceptually outlined in Table 1.

Linear Models

Category 1: Univariate-linear models

Following Callen et al. (1996), we examine three ARIMA specifications.

M1.1: From Brown & Rozef (1979), the (1,0,0) × (0,1,1)4 specification in (1):

\[ E(Y_t) = Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) - \Theta_1a_{t-4} + \delta \]  

M1.2: From Griffin (1977), Watts (1975), and Watts and Leftwich (1977), the (0,1,1) × (0,1,1)4 specification in (2):

\[ E(Y_t) = Y_{t-4} + (Y_{t-1} - Y_{t-5}) - \theta_1a_{t-1} - \Theta_1a_{t-4} - \theta_1\Theta_1a_{t-5} + \delta \]  

M1.3: From Foster (1977), the (1,0,0) × (0,1,0)4 specification in (3):

\[ E(Y_t) = Y_{t-4} + \phi_1(Y_{t-1} - Y_{t-5}) + \delta \]  

We define \( \delta \) as the constant term of the ARIMA models, \( a_t \) as the disturbance term at period \( t \), \( \phi \) as the autoregressive parameter, \( \theta \) as the moving-average parameter, and \( \Theta \) as the seasonal moving-average parameter. \( Y \) stands for quarterly EPS. In addition to time-series specifications, we also investigate a univariate OLS specification to
benchmark the incremental explanatory power of fundamental variables for future EPS in an OLS setting.

M1.4: From Lawrence and Rozeff (1979), the univariate OLS specification in (4):

\[ E(Y_t) = a + b_1 Y_{t-1} + b_2 Y_{t-4} + e_t \] (4)

Category 2: Multivariate-linear models

This category contains multivariate time-series EPS prediction models incorporating fundamental information. We will test two specifications.

M2.1: Similar to Abarbanell and Bushee’s (1997) study and the model used by Lorek and Willinger (1996) for forecasting cash flows, we will use the following multivariate time-series models with fundamental variables lagged by one quarter in (5):

\[ E(Y_t) = a + b_1 Y_{t-1} + b_2 Y_{t-4} + b_3 INV_{t-1} + b_4 AR_{t-1} + b_5 CAPX_{t-1} \\
+ b_6 GM_{t-4} + b_7 SA_{t-4} + b_8 ETR_{t-4} + b_9 LF_{t-4} + e_t \] (5)

M2.2: Following Abarbanell and Bushee (1997) and Lorek and Willinger (1996) and taking into account the potential seasonality of quarterly financial data, this multivariate time-series model now possesses fundamental variables lagged by four quarters in (6):

\[ E(Y_t) = a + b_1 Y_{t-1} + b_2 Y_{t-4} + b_3 INV_{t-4} + b_4 AR_{t-4} + b_5 CAPX_{t-4} \\
+ b_6 GM_{t-4} + b_7 SA_{t-4} + b_8 ETR_{t-4} + b_9 LF_{t-4} + e_t \] (6)

where

\[ Y = \text{Quarterly EPS}; \]
\[ INV = \text{Inventory}; \]
\[ AR = \text{Accounts receivables}; \]
\[ CAPX = \text{Capital expenditure per Schedule V (Schedule V contains disclosures of property, plant, and equipment under the Securities \\& Exchange Commission [SEC] Regulation S-X.); since Compustat reports only annual capital expenditure, we divide the annual amount by four to approximate to the quarterly capital expenditure;} \]
\[ GM = \text{Gross margin, defined as sales less cost of goods sold;} \]
\[ SA = \text{Selling and administrative expenses}; \]
\[ ETR = \text{Effective tax rate, defined as income taxes divided by pretax income;} \]
\[ LF = \text{Labor force, defined as sales divided by the number of employees; since Compustat reports the number of employees on a yearly basis, we take this number as the quarterly, assuming that the number of employed remains constant across the four quarters. (In some situations, firms report this number as an average number of employees, and in some situations, as the number of employees at} \]
Variables $INV, AR, CAPX, GM,$ and $SA$ are scaled by the weighted average number of common shares used in calculating the basic quarterly EPS. Further, we use the log form of the $LF$ variable to align $LF$ with other variables in scales. If not in the log form, the variable is usually in thousands or even millions, on a much larger scale than all the other variables. Also note that in our multivariate models, two fundamental accounting variable signals identified by Abarbanell and Bushee (1997) are not included: the EPS quality variable (measured as the inventory policy adopted by the firm) and auditor opinion variable. This is because for the model estimation period, a large number of firms (242 out of the 283 identified firms) consistently used one type of inventory policy (LIFO versus all other) or obtained one type of auditor opinion (unqualified versus all other).

We make two clarifications here. First, the choice of the fundamental variables is based on the previous literature, mainly Lev and Thiagarajan (1993) and Abarbanell and Bushee (1997). These two studies conducted a directed search of the written pronouncements of financial analysts to identify the set of variables claimed to be useful in security valuation and earnings prediction. Later studies such as Abarbanell and Bushee (1998) and Beneish et al. (2001) included the identified variables in their tests and confirmed the theory underlying the choice of these variables. In the Supplemental Tests section of this paper, we will add or drop certain signals and will also test various combinations of the variables to evaluate the robustness of our results.

Second, to overcome the constraints of a lack of quarterly data for capital expenditure and labor force, we make the assumptions that the capital expenditure is spread evenly across the quarters and that the number of employees remains relatively constant throughout the year. We argue that if any bias is introduced in making these assumptions for the quarterly measures, it is very likely that the bias will be identical across the multivariate linear and multivariate non-linear models where these variables are included. Furthermore, in our Supplemental Tests, we will drop the labor force variable and replace the capital expenditure per Schedule V with the capital expenditure reported by the statement of cash flows for which the quarterly information is available, with the major difference being that the former includes the capital expenditure for the acquired companies while the latter does not. The alternative measure is also used by Park and Pincus (2003). We will evaluate the sensitivity of our results to different variable measurements.

**Neural Network Models**

Based on the work of Callen et al. (1996), we use three-layer neural networks in both univariate and multivariate NN settings. According to Qi (1999), it has been widely accepted that a three-layer feedforward network with an identify transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous functions arbitrarily well, given sufficiently
many middle-layer units. The generic three-layer network model can be expressed in (7):

\[ Y_t = f[(X, \alpha, \beta) = \sum_{j=1}^{n} \alpha_j \log \text{sig} \left( \sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j} \right) \]  (7)

where \( Y_t \) is the network’s output, \( X \) is the input vector, \( x_i \) is \( i \)th input, \( n \) is the number of units in the middle layer, \( k \) is the number of inputs, \( \alpha \) represents a vector of the coefficients (weights) from the middle to output layer units, \( \beta \) indicates a matrix of the coefficients from the input to middle-layer units, \( \alpha_j \) is the weight of the output layer that connects the \( j \)th hidden layer unit to the output, \( \beta_{ij} = \{ \beta_{ij}, i = 1, 2, \ldots, k \} \) is the weight vector of the \( j \)th unit of the middle layer, \( \beta_{0j} \) is the bias weight of the \( j \)th unit of middle layer unit, and \( \log \text{sig} \) is the logistic transfer function \( \log \text{sig}(a) = 1/(1 + \exp(-a)) \).

**Category 3 (M3): Univariate-neural network model**

To forecast one-quarter ahead EPS, we use rolling samples of 30 quarters which are further broken down into 26 adjacent point 4-tuples of structure \( X \) as inputs and 26 corresponding outputs of the structure \( Y_t \). In a univariate or time series NN model, \( X_t = Y_t \). The univariate-neural network model is expressed in (8):

\[ Y_t = f[(Y_{t-4}, Y_{t-3}, Y_{t-2}, Y_{t-1}), \alpha, \beta] = \sum_{j=1}^{n} \alpha_j \log \text{sig} \left( \sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j} \right) \]  (8)

**Category 4 (M4): Multivariate-neural network model**

By the same token, we need to use 4-tuples of structure \( X \) as inputs and 26 corresponding outputs of the structure \( Y_t \). Moreover, in a multivariate setting, \( X_t \) becomes a vector itself: \( X_t = \{ Y_{t-1}, Y_{t-4}, INV, AR, CAPX, GM, SA, ETR, LF \} \). The multivariate-neural network model is expressed in (9):

\[ Y_t = f[\{Y_{t-4}, Y_{t-7}, INV, AR, CAPX, GM, SA, ETR, LF\}, \times \{Y_{t-3}, Y_{t-6}, INV, AR, CAPX, GM, SA, ETR, LF\}, \times \{Y_{t-2}, Y_{t-5}, INV, AR, CAPX, GM, SA, ETR, LF\}, \times \{Y_{t-1}, Y_{t-4}, INV, AR, CAPX, GM, SA, ETR, LF\}], \alpha, \beta] \]  
\[ = \sum_{j=1}^{n} \alpha_j \log \text{sig} \left( \sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j} \right) \]  (9)

**Forecasting Accuracy Procedure**

While a total of 40 observations are used in forecasting for each company, we use 30 observations to estimate the forecasting models and make ten one-step-ahead forecasts (from the fourth quarters of 1999 to the first quarter of 2002) based on the identified models. We use a rolling estimation method (i.e., for each firm, we
use the models identified by the data from the 1st to 30th quarters to forecast the EPS for the 31st quarter, and use the models identified by the data from the 2nd to 31st quarters to forecast the EPS for the 32nd quarter, and so on. We compute the following two error metrics to measure forecast accuracy in (10) and (11):

\[
\text{Mean Absolute Percentage Error (MAPE)} = \frac{1}{10} \sum_{t=31}^{40} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \tag{10}
\]

\[
\text{Mean Squared Error (MSE)} = \frac{1}{10} \sum_{t=31}^{40} \left( \frac{Y_t - \hat{Y}_t}{Y_t} \right)^2 \tag{11}
\]

where \( \hat{Y}_t \) is the forecasted value of quarterly EPS for period \( t \). Observations with a zero-quarterly-EPS are eliminated. Following Lorek and Willinger (1996) and Brown and Rozeff (1979), we set a forecast error as one when it exceeds 100%.

**Market Expectation and Returns Tests**

We assess our forecast models’ ability to approximate the security market’s expectation of future EPS based on Foster (1977) and Bathke and Lorek (1984). We argue that, if the market relies on a certain forecast model to set its earnings expectations, the association between forecast errors and abnormal security returns should be the strongest for that forecast model. We assess this association for each model by examining the relation between the directions of EPS forecast errors and directions of market price movements for each model.

We adjust firms’ returns by the market-wide factors using the following specification in (12):

\[
r_{i,t} = \alpha_{i,t} + \beta_{i,t} r_{m,t} + \varepsilon_{i,t}, \tag{12}
\]

where \( r_{i,t} \) is the firm-specific return, \( r_{m,t} \) is the NYSE/AMEX/Nasdaq value-weighted market index, and \( \varepsilon_{i,t} \) is the abnormal return on day \( t \). Denoting the quarterly EPS announcement date (identified from Compustat database) as \( t = 0 \), we calculate the abnormal return \( \varepsilon_{i,t} \) as \( r_{i,t} - \bar{\alpha}_{i,t} - \bar{\beta}_{i,t} r_{m,t} \), where \( \bar{\alpha} \) and \( \bar{\beta} \) are estimated parameters using daily returns from a 260-day period comprising \((t - 162, t - 33)\) and \((t + 33, t + 162)\). We then calculate the cumulative abnormal returns in (13):

\[
\text{CAR}_t = \sum_{t=-32}^{t+32(\text{ort}+1)} \varepsilon_{i,t} \tag{13}
\]

We examine the proposed association for two periods: \( t - 32 \) to \( t + 32 \) and \( t - 32 \) to \( t + 1 \).

As in previous research, for each forecast model, if the actual EPS are higher than the forecast, we classify the firm’s quarter into the “good news” portfolio; if the actual EPS are lower, then the firm’s quarter is classified into the “bad news” portfolio. We expect that the “good news” portfolio will be associated with positive abnormal returns while the “bad news” portfolio associated with negative abnormal returns. We use a two-by-two contingency table to assess the significance of this association and report the chi-square statistics together with the average CARs.
to the “good news” portfolio, to the “bad news” portfolio, and to an investment strategy of buying securities with good news and selling securities with bad news.

**Hypotheses-Testing Procedure**

Based on the results of the forecast accuracy statistics, two statistical comparisons (using the Friedman “F” test and chi-square tests) of sets of models (refer to Table 1) are used in the basic hypotheses testing design. As previously stated we test H1 by examining H1a, H1b, and H1c separately. For H1a (i.e., whether the use of the nonlinear NN method improves the forecast accuracy vis-à-vis the linear method when fundamental accounting variables are absent), we will compare models of Category 1 against 3. For H1b (i.e., whether the use of the nonlinear NN method improves the forecast accuracy vis-à-vis the linear method when fundamental accounting variables are present), we will compare models of Category 2 against 4. For H1c (i.e., whether the multivariate NN model outperforms all others), we will compare Category 4 against 1, 2, and 3.

To determine the significance of H1a, H1b, and H1c, we conduct Friedman tests (Conover, 1980) to gauge the differences in the EPS forecast models for individual quarters and for the forecasts pooled across quarters. In testing H1a, for each firm’s quarter, we assign 1 to the model with the lowest MAPE and 5 to the model with the highest MAPE, and calculate the average ranks for each model. We then use the Friedman test to conduct pair-wise comparisons for all the possible combinations of the five forecast models. Based on the pair-wise results, we assign a performance score of 1 to the model that outperforms all others, 5 to the model that performs the worst, and average scores to ties. For example, if from pair-wise comparisons, we observe that Model 1.1 ≻ Model 1.2 and Model 1.2 ≻ Model 1.3 (where we define “≽” as a model that produces statistically more accurate forecasts and “∼” when two models are not statistically different in forecast performance), then we assign the performance scores as the followings: 1 for M1.1, 2 for M1.2, and 3 for M1.3. If, however, Model 1.1 ≻ Model 1.2 but Model 1.1 ∼ Model 1.3 and Model 1.2 ∼ Model 1.3, then we take the performance across the three models as indistinguishable and will assign an average score of 2 to each of the three models (i.e., \((1 + 2 + 3)/3 = 2\), assuming that we only have three models to rank). For H1b, the same procedures as H1a apply except that we compare only three models. Thus, we assign a score of 1 to the most accurate forecast model and 3 to the least accurate one. For H1c, the most accurate model has a score of 1 and the least accurate one has a score of 8.

For H2, where we seek to compare the strength of associations between forecast error and abnormal security returns for the univariate models (linear and NN) and multivariate models (linear and NN), we will compare the chi-square statistics for each model obtained from a \(2 \times 2\) contingency table. A general null for the \(2 \times 2\) contingency table test is that “the event ‘an observation is in row \(i\)’ is independent of the event ‘that same observation is in column \(j\)’” (Conover, 1980, p. 159). In the context of our paper, \(i\) could be “good news” and “bad news,” and \(j\) could be “positive abnormal returns” and “negative abnormal returns.” For our forecast models, we expect that “good news” is associated with positive abnormal returns and “bad news” is associated with negative abnormal returns. We expect that the forecast model, which is the most likely surrogate for the market’s earnings
expectations, will exhibit higher association (or dependence) between news type and abnormal returns and, thus, will have the highest chi-square statistic. Thus, chi-square statistics from a $2 \times 2$ contingency table will be reported for each forecast model, consistent with prior forecasting study methodologies used by Foster (1977) and Bathke and Lorek (1984).

An alternative method for evaluating the strength of the associations is to examine the magnitude of the returns to the “good news” portfolio, “bad news” portfolio, and the composite returns to the investment strategy of buying securities experiencing “good news” and selling securities experiencing “bad news.” We expect the most likely surrogate for the market’s earnings expectations to have the highest magnitude of abnormal returns and composite returns.

RESULTS AND DISCUSSION

Data Profile

We collected quarterly EPS and fundamental information data from Compustat. Although Compustat provides quarterly EPS data for a relatively long period of time (approximately 80 quarters), the longest time-series for fundamental variables is 45 quarters. We further lose five data points due to differencing. As a result, we obtain a sample of 283 firms, each having 40-quarter observations from the second quarter of 1992 to the first quarter of 2002. Table 2 presents a breakdown of sample firms into industries based on two-digit SIC codes. Our sample comprises a total of 41 industries. Among them, the industry of electrical and electronic equipment accounts for the largest portion of the sample (9.54%). Such industries as chemical

<table>
<thead>
<tr>
<th>Two-Digit SIC Code</th>
<th>Industry</th>
<th>Number of Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Oil and gas extraction</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>Food and kindred products</td>
<td>13</td>
</tr>
<tr>
<td>26</td>
<td>Paper and allied products</td>
<td>17</td>
</tr>
<tr>
<td>27</td>
<td>Printing and published</td>
<td>8</td>
</tr>
<tr>
<td>28</td>
<td>Chemicals and allied products</td>
<td>24</td>
</tr>
<tr>
<td>30</td>
<td>Rubber and miscellaneous plastic products</td>
<td>9</td>
</tr>
<tr>
<td>33</td>
<td>Primary metal industries</td>
<td>8</td>
</tr>
<tr>
<td>34</td>
<td>Fabricated metal products</td>
<td>16</td>
</tr>
<tr>
<td>35</td>
<td>Industrial machinery and equipment</td>
<td>24</td>
</tr>
<tr>
<td>36</td>
<td>Electrical and electronic equipment</td>
<td>27</td>
</tr>
<tr>
<td>37</td>
<td>Transportation equipment</td>
<td>12</td>
</tr>
<tr>
<td>38</td>
<td>Instruments and related products</td>
<td>22</td>
</tr>
<tr>
<td>39</td>
<td>Miscellaneous manufacturing industries</td>
<td>6</td>
</tr>
<tr>
<td>50</td>
<td>Wholesale trade—durable goods</td>
<td>16</td>
</tr>
<tr>
<td>51</td>
<td>Wholesale trade—nondurable goods</td>
<td>7</td>
</tr>
<tr>
<td>73</td>
<td>Business services</td>
<td>10</td>
</tr>
<tr>
<td>Other Industries</td>
<td></td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>283</td>
</tr>
</tbody>
</table>
products, industrial machinery, and instruments also represent a relatively large portion of the sample (8.48%, 8.48%, and 7.77%, respectively).

**Results on H1 Hypothesis**

Table 3 reports the averages of the two measures for individual quarters as well as for the forecasts pooled across quarters. Pertaining to H1a, we find that the MAPEs of the univariate linear models in Category 1 (mostly between 0.50 and 0.60) are consistently higher than the MAPE measures of the univariate nonlinear (NN) model of Category 3 (between 0.48 and 0.54) for each quarter and for all the quarters combined, supporting the superior performance of the NN model in a univariate setting. A comparison of MAPEs between the multivariate linear and multivariate nonlinear (NN) model, as suggested by H1b, yields even stronger evidence in favor of the NN model. The multivariate model using the neural network approach (Category 4) has much lower MAPEs (between 0.35 and 0.38) than those of the multivariate linear approaches in Category 2 (mostly above 0.60). Pertaining to H1c, it is apparent that the multivariate NN model of Category 4 outperforms all other models with the lowest MAPEs. Across the eight models, the MSE measure and the percentage of large errors (defined as forecast errors above 100%) exhibit the same patterns.

The evidence in Table 3 also indicates fundamental variables are conditionally value-adding in forecasting future EPS based on the models used in forecasting. On the one hand, simply adding a linear combination of fundamental accounting variables to the linear forecasting models does not result in more accurate forecasts. The forecast accuracy, instead, is adversely affected by including fundamental accounting variables into the linear models, as evidenced by a comparison of Category 1 and Category 2. Judge, Hill, Griffiths, Lutkepohl, and Lee (1988) commented on the reason why the use of a particular variable may not provide a superior forecast over just using the past values of the variable: “Our information about the underlying sampling mechanism is generally incomplete, and thus economic and econometric models are at best rough approximations to reality. Therefore it should not be surprising that time-series models that use only the information from a set of observations on a single variable have in some instances provided forecast that are superior to predictions from a large-scale econometric model” (Judge et al., 1988, p. 675). On the other hand, including the fundamental signals in a nonlinear forecasting model improves forecasting accuracy notably, as evidenced by a comparison of error metrics of Category 3 and of Category 4. The seemingly conflicting results, obtained from using and not using the neural network approach, suggests that the manner in which fundamental variables are integrated into future EPS is nonlinear, and that the use of the neural network approach is appropriate here in that it captures the unknown form of nonlinearity. This result is what Papatla et al. (2002) found in their use of NN to explain unseen forces that could not be captured in a linear model. Therefore, we argue that the reduction in MAPE measures for the model of Category 4 should be attributed to both the inclusion of the fundamental variables and the use of the neural network approach.
Table 3: Comparing forecast accuracy of one-step-ahead quarterly EPS for four categories of forecast methods.

<table>
<thead>
<tr>
<th>Category 1</th>
<th>1st Quarter</th>
<th>2nd Quarter</th>
<th>3rd Quarter</th>
<th>4th Quarter</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE(^1)</td>
<td>MSE(^2)</td>
<td>MAPE</td>
<td>MSE</td>
<td>MAPE</td>
</tr>
<tr>
<td>M1.1</td>
<td>0.575</td>
<td>0.480</td>
<td>35.5%</td>
<td>0.541</td>
<td>0.449</td>
</tr>
<tr>
<td>M1.2</td>
<td>0.603</td>
<td>0.511</td>
<td>37.6%</td>
<td>0.546</td>
<td>0.451</td>
</tr>
<tr>
<td>M1.3</td>
<td>0.564</td>
<td>0.467</td>
<td>33.7%</td>
<td>0.529</td>
<td>0.444</td>
</tr>
<tr>
<td>M1.4</td>
<td>0.548</td>
<td>0.443</td>
<td>31.0%</td>
<td>0.544</td>
<td>0.438</td>
</tr>
<tr>
<td>Category 2</td>
<td>M2.1</td>
<td>0.631</td>
<td>0.543</td>
<td>41.6%</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td>M2.2</td>
<td>0.629</td>
<td>0.541</td>
<td>42.6%</td>
<td>0.552</td>
</tr>
<tr>
<td>Category 3</td>
<td>M3</td>
<td>0.534</td>
<td>0.432</td>
<td>29.4%</td>
<td>0.492</td>
</tr>
<tr>
<td>Category 4</td>
<td>M4</td>
<td>0.372</td>
<td>0.268</td>
<td>16.5%</td>
<td>0.368</td>
</tr>
<tr>
<td>Number of</td>
<td>N = 836</td>
<td>N = 562</td>
<td>N = 562</td>
<td>N = 840</td>
<td>Total = 2800(^4)</td>
</tr>
</tbody>
</table>

\(^1\) Mean Absolute Percentage Error (MAPE)
\(^2\) Mean Squared Error (MSE)
\(^3\) We set forecast error \(\left| \frac{\hat{Q}_t - Q_t}{Q_t} \right| \) as 100% when this expression exceeds 100% (large forecast error).
\(^4\) We excluded firm-quarters for which the EPS is zero, thus reducing the total firm-quarters from 2,830 to 2,800.
To provide a more formal testing of H1, we conduct the Friedman (Conover, 1980) test to gauge the statistical significance of difference in EPS forecast models for individual quarters and for the forecasts pooled across quarters. In Table 4, we report the models’ average ranks and overall performance scores obtained from the binary comparisons as specified by the hypothesis-testing procedures.

Panel A Table 4 presents the evidence for testing H1a. Note that we report the F-statistic here because Iman and Davenport (1980) have shown that the Friedman test statistic is better approximated using an F-distribution than the previously identified chi-square distribution. For the sample pooled across quarters, the Friedman test for multiple treatments has an F statistic of 14.036 (\( p = 0.000 \)), indicating that forecast accuracy, measured as ranks, varies significantly among forecast models of Category 1 and 3. The univariate NN model (Category 3) has the lowest average rank (2.832) and lowest performance score (1), thus representing the most accurate forecast model among the five. The four univariate linear models are indistinguishable from one another, and are assigned an overall performance score of 3.5 on a scale of 1 to 5. Thus, H1a in its null form is rejected for the sample pooled across quarters. A more detailed look at the quarter-by-quarter results shows that the results for the first and fourth quarters are consistent with those for the overall sample. However, the results for the second and third quarters suggest that the univariate NN model does not significantly dominate other univariate linear models. Therefore, we reject H1a for the sample pooled across quarters, but we cannot reject H1a for two (the second and third) of the four quarters.

Panel B Table 4 presents test results related to H1b. Among the three models of Category 2 and 4, for both individual quarters and all quarters, the multivariate neural network model (Category 4) consistently dominates. It has the lowest average rank (between 1.460 and 1.504) and is assigned a performance score of 1 across all quarters. The two multivariate linear models with fundamental variables lagged by either one or four quarters (Category 2) have higher average ranks, implying worse performance than that of the multivariate NN model. Therefore, H1b in its null form is rejected for both observations pooled across quarters and observations for specific quarters.

H1c is tested by comparing the forecast accuracy across all eight models. Panel C of Table 4 presents substantial evidence that the multivariate NN model is the most accurate forecasting model. It has the lowest average rank (between 2.743 and 3.124) and is assigned a score of 1 across all the quarters. Thus, H1c in its null form is also rejected.

Based on the results reported in Tables 3 and 4, we conclude that there is a significant difference in the forecasting accuracy between linear and nonlinear NN models. The difference is larger in magnitude between the multivariate linear and multivariate NN models when fundamental accounting variables are present. Simply stated, we reject H1 and conclude that using NN models does improve forecasting accuracy when forecasting quarterly EPS, and that the improved performance is more evident when fundamental accounting variables are added to the forecasting models. Therefore, in the current context, a nonlinear approach (i.e., the neural network approach) incorporating fundamental accounting variable information apparently represents a better forecast model.
Table 4: Ranking forecast accuracy of one-step-ahead quarterly EPS for eight forecast models.

<table>
<thead>
<tr>
<th>1&lt;sup&gt;st&lt;/sup&gt; Quarter</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Quarter</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Quarter</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; Quarter</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Quarter</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Quarter</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Quarter</td>
<td>4&lt;sup&gt;th&lt;/sup&gt; Quarter</td>
<td>Overall</td>
</tr>
<tr>
<td>Average Ranks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Average Ranks</td>
<td>Average Ranks</td>
<td>Average Ranks</td>
<td>Average Ranks</td>
</tr>
<tr>
<td>M1.1</td>
<td>3.056</td>
<td>3</td>
<td>2.988</td>
<td>2.5</td>
</tr>
<tr>
<td>M1.2</td>
<td>3.170</td>
<td>5</td>
<td>3.010</td>
<td>2.5</td>
</tr>
<tr>
<td>M1.3</td>
<td>2.961</td>
<td>3</td>
<td>2.931</td>
<td>2.5</td>
</tr>
<tr>
<td>M1.4</td>
<td>2.974</td>
<td>3</td>
<td>3.185</td>
<td>5</td>
</tr>
<tr>
<td>M3</td>
<td>2.839</td>
<td>1</td>
<td>2.886</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Panel A: Comparisons of Univariate Linear and NN Models (H1a)

F = 6.529 (p = 0.000)<sup>3</sup> | F = 3.720 (p = 0.005) | F = 2.233 (p = 0.063) | F = 8.497 (p = 0.000) | F = 14.036 (p = 0.000)

Panel B: Comparisons of Multivariate Linear and NN Models (H1b)

F = 298.297 (p = 0.000) | F = 176.745 (p = 0.000) | F = 154.348 (p = 0.000) | F = 293.373 (p = 0.000) | F = 904.213 (p = 0.000)
Table 4: (continued) Ranking forecast accuracy of one-step-ahead quarterly EPS for eight forecast models.

<table>
<thead>
<tr>
<th></th>
<th>1st Quarter</th>
<th>2nd Quarter</th>
<th>3rd Quarter</th>
<th>4th Quarter</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
<td>Performance Scores Determined by the Friedman Test</td>
</tr>
<tr>
<td></td>
<td>Average Ranks</td>
<td>Average Ranks</td>
<td>Average Ranks</td>
<td>Average Ranks</td>
<td>Average Ranks</td>
</tr>
</tbody>
</table>

**Panel C: Comparisons of the Multivariate NN Model and All Others (H1c)**

<table>
<thead>
<tr>
<th>Model</th>
<th>1st Quarter</th>
<th>2nd Quarter</th>
<th>3rd Quarter</th>
<th>4th Quarter</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1.1</td>
<td>4.658</td>
<td>4</td>
<td>4.530</td>
<td>3.5</td>
<td>4.736</td>
</tr>
<tr>
<td>M1.2</td>
<td>4.835</td>
<td>6</td>
<td>4.553</td>
<td>3.5</td>
<td>4.710</td>
</tr>
<tr>
<td>M1.3</td>
<td>4.519</td>
<td>4</td>
<td>4.435</td>
<td>3.5</td>
<td>4.771</td>
</tr>
<tr>
<td>M1.4</td>
<td>4.493</td>
<td>4</td>
<td>4.823</td>
<td>6.5</td>
<td>4.742</td>
</tr>
<tr>
<td>M2.1</td>
<td>5.120</td>
<td>7.5</td>
<td>5.345</td>
<td>8</td>
<td>4.936</td>
</tr>
<tr>
<td>M2.2</td>
<td>5.175</td>
<td>7.5</td>
<td>4.798</td>
<td>6.5</td>
<td>5.040</td>
</tr>
<tr>
<td>M3</td>
<td>4.322</td>
<td>2</td>
<td>4.391</td>
<td>3.5</td>
<td>4.321</td>
</tr>
<tr>
<td>M4</td>
<td>2.878</td>
<td>1</td>
<td>3.124</td>
<td>1</td>
<td>2.743</td>
</tr>
</tbody>
</table>

F = 100.311 (p = 0.000) F = 49.651 (p = 0.000) F = 47.990 (p = 0.000) F = 104.735 (p = 0.000) F = 290.493 (p = 0.000)

1We assign 1 to the model with the lowest MAPE and 5 to the model with the highest MAPE in Panel A and average the ranks across all firm-quarters. The same procedure is implemented for Panels B and C but the model with the highest MAPE in Panel B is assigned to 3 and the highest MAPE in Panel C is assigned to 8.

2In Panel A, we use the Friedman test for two treatments to conduct pair-wise comparisons for all the possible combinations of the five models. Based on the pair-wise results, we assign a performance score of 1 to the model which outperforms all others, 5 to the model that performs the worst, and average scores to ties. We define two models as being significantly different in forecast accuracy when the two-tailed t-test yields a probability of equal to or lower than 0.1.

3F statistics are obtained from the Friedman test for multiple treatments (in our case, five treatments for Panel A, three treatments for Panel B, and eight treatments for Panel C).
Results on H2 Hypothesis

The chi-square, “good news,” “bad news,” and composite statistics for the H2 research question are presented in Table 5. We find that the most accurate forecast model, the multivariate-neural network model of Category 4, does not exhibit the strongest association between the sign of forecast errors and abnormal returns. For the period of \((t - 32, t + 1)\), the \(\chi^2\) for the multivariate-neural network model is 10.955. Significance levels for the \(\chi^2\)-distribution with one-degree of freedom is 2.706 for \(p = 0.1\), 3.841 for \(p = 0.05\), and 6.635 for \(p = 0.01\). Although significant, it is lower than those for the univariate models (Categories 1 and 3). Among the four categories of forecast models, the \(\chi^2\) statistics are the highest for the univariate linear models (Category 1). Over the \((t - 32, t + 1)\) period, the returns to the “good news” portfolio are all positive while those to the “bad news” portfolio are all negative. The composite returns are again the highest (1.8% to 2.4%) for the investment strategy based on the EPS forecasts yielded by the univariate-linear forecast models. The results for the period of \((t - 32, t + 32)\) suggest the same interpretations. Only, all chi-square statistics are much lower, and the composite returns are also slightly lower over this period compared to those for the \((t - 32, t + 1)\) period.

Thus, the market return tests indicate that the market expectations of one-step-ahead quarterly EPS are more in line with the univariate linear forecast models, although the neural network approach incorporating fundamental accounting signals obviously generates more accurate forecasts. While there are statistically significant associations for all the models, the degree of association, measured by the significant \(\chi^2\) statistics, is better for the less accurate univariate linear models than the NN model with fundamental accounting variables. We therefore reject H2, that there is no difference in the degree to which the market relies upon different forecasting models in forming its expectations. The evidence points to just the opposite, that NN models with fundamental accounting variables have weaker associations between forecast errors and abnormal security returns than the less accurate univariate linear models. These results, though, are consistent with the findings of Foster (1977) and Bathke and Lorek (1984) who document that the market adjusts for the seasonality of the quarterly EPS data.

Our return tests show that the most accurate forecasting model did not capture the market’s expectation well. There could be two explanations for this result. First, forecasting accuracy is not a major concern to market participants (i.e., investors); second, investors have failed to recognize or use the most accurate forecasting method. The first explanation is not likely, given the importance the investors place on earnings forecast accuracy (Elgers et al., 1995). If the second explanation is adopted, then limitations in the information processing capacity of decision makers (i.e., investors) may be implied. Specifically, if the market had used all information (for example, information represented by fundamental signals) in an appropriate way (for example, a nonlinear approach) in setting its earnings expectations, then its earnings forecasts would have been the closest to those predicted by the multivariate NN model (which is the most accurate forecasting method), and consequently, the association between forecast errors and abnormal security returns (revision in the market’s valuation of the firm) would have been the greatest
Table 5: Testing the association between the unexpected earnings and abnormal market returns for forecast models.

<table>
<thead>
<tr>
<th>Category</th>
<th>For the Period of ((t - 32, t + 1)^1)</th>
<th>For the Period of ((t - 32, t + 32))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chi-Square</td>
<td>Good News(^2)</td>
</tr>
<tr>
<td>Category 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1.1</td>
<td>28.947 (p = 0.000)</td>
<td>0.027</td>
</tr>
<tr>
<td>M1.2</td>
<td>17.251 (p = 0.000)</td>
<td>0.022</td>
</tr>
<tr>
<td>M1.3</td>
<td>31.984 (p = 0.000)</td>
<td>0.030</td>
</tr>
<tr>
<td>M1.4</td>
<td>22.921 (p = 0.000)</td>
<td>0.021</td>
</tr>
<tr>
<td>Category 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2.1</td>
<td>7.157 (p = 0.007)</td>
<td>0.017</td>
</tr>
<tr>
<td>M2.2</td>
<td>11.938 (p = 0.001)</td>
<td>0.019</td>
</tr>
<tr>
<td>Category 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>12.884 (p = 0.000)</td>
<td>0.018</td>
</tr>
<tr>
<td>Category 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>10.955 (p = 0.001)</td>
<td>0.017</td>
</tr>
</tbody>
</table>

---

\(^1\)We define the earnings announcement date (identified from Compustat) as \(t = 0\).

\(^2\)This column reports the mean abnormal returns to the “Good News” portfolio. Good news is defined as the actual quarterly EPS higher than the forecasted value.

\(^3\)Bad news is defined as the actual quarterly EPS lower than the forecasted value.

\(^4\)Composite returns are obtained by buying the securities with good news and selling the securities with bad news.
for the multivariate NN model. Obviously, the evidence from the return tests shows that this is not the case. Therefore, the market has failed to recognize or utilize the most accurate forecasting method. We suggest that either investors have ignored the information relevant to forming earnings expectations or the investors have not employed complicated information processing techniques in forming their expectations although such techniques prove to be beneficial in predicting future EPS. For example, even if the market has neglected the fundamental information, a non-linear approach (NN) in forecasting quarterly EPS will still be able to improve the forecasting accuracy. However, our H2 results suggest compared to the univariate neural network approach, the traditional univariate time-series models that account for the seasonality of quarterly EPS are more likely to have been used by the investors in forming EPS expectations, as shown by a stronger association between the sign of forecast errors and the sign of abnormal returns for the univariate time-series models.

Supplemental Tests

Quarterly, yearly, and industry-specific forecasts

Our forecast period includes the second and third quarters twice and the first and fourth quarters three times. To address the possibility that the frequency of each quarter appearing in the sample might bias our results, we conduct a sensitivity test by truncating the forecast period to the third quarter of 2001 so that each quarter is covered by the forecast period twice. The forecasting results remained the same.

To validate that our results regarding forecast accuracy are not driven by a specific quarter or year, we present the ranks of each forecast model on a quarter-by-quarter and year-by-year basis in Panel A Table 6 and Panel B Table 6, respectively. It is apparent that our results are not influenced by a specific time period. We also examine the ranks of forecasting models for each industry and find consistent patterns across industries (results available upon request). Therefore, we believe that the dominance of the multivariate NN model is applicable to firms from any industries for all periods covered by our sample.

We further examined whether the dominance of the multivariate NN model is linked with firm size or earnings variability. We calculated firm size as the natural log of the number of shares outstanding at the end of the quarter multiplied by stock price, and calculated the earnings variability as the standard deviation divided by the average of the EPS over the estimation period. We ran a simple logit model where the dependent variable is coded as 1 if the multivariate NN outperforms others and zero otherwise. We found no relation between the probability of multivariate NN dominance and firm size or earnings variability.

Forecasts with variations in the fundamental variable set

The choice of fundamental accounting variables is made based on the evidence presented by prior literature, including Lev and Thiagarajan (1993), Abarbanell and Bushee (1997, 1998), Beneish et al. (2001), Khurana and Raman (2003), Swanson, Rees, and Juarez-Valdes (2003), and Bao and Bao (2004). These variables were first identified by Lev and Thiagarajan (1993) and have been documented by later
Table 6: Rankings of model forecast accuracy for individual quarters and years.

Panel A: Quarterly Rankings

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M1.1</td>
<td>4.5</td>
<td>3.5</td>
<td>4.5</td>
<td>5.0</td>
<td>5.0</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>4.5</td>
</tr>
<tr>
<td>M1.2</td>
<td>4.5</td>
<td>7.0</td>
<td>4.5</td>
<td>5.0</td>
<td>5.0</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>4.5</td>
</tr>
<tr>
<td>M1.3</td>
<td>4.5</td>
<td>3.5</td>
<td>4.5</td>
<td>5.0</td>
<td>5.0</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>4.5</td>
</tr>
<tr>
<td>M1.4</td>
<td>4.5</td>
<td>3.5</td>
<td>4.5</td>
<td>5.0</td>
<td>5.0</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>4.5</td>
</tr>
<tr>
<td>M2.1</td>
<td>7.0</td>
<td>7.0</td>
<td>8.0</td>
<td>5.0</td>
<td>5.0</td>
<td>7.5</td>
<td>5.5</td>
<td>7.5</td>
<td>5.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>M2.2</td>
<td>8.0</td>
<td>7.0</td>
<td>4.5</td>
<td>5.0</td>
<td>5.0</td>
<td>7.5</td>
<td>5.5</td>
<td>7.5</td>
<td>5.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>M3</td>
<td>2.0</td>
<td>3.5</td>
<td>4.5</td>
<td>5.0</td>
<td>5.0</td>
<td>4.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>M4</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Number of Forecasts</td>
<td>N = 281</td>
<td>N = 279</td>
<td>N = 281</td>
<td>N = 280</td>
<td>N = 281</td>
<td>N = 278</td>
<td>N = 281</td>
<td>N = 282</td>
<td>N = 276</td>
<td>N = 279</td>
<td>N = 2,800</td>
</tr>
</tbody>
</table>

Panel B: Yearly Rankings

<table>
<thead>
<tr>
<th></th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1.1</td>
<td>5.0</td>
<td>4.0</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>M1.2</td>
<td>5.0</td>
<td>4.0</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>M1.3</td>
<td>5.0</td>
<td>4.0</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>M1.4</td>
<td>5.0</td>
<td>4.0</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>M2.1</td>
<td>5.0</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>M2.2</td>
<td>8.0</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>M3</td>
<td>2.0</td>
<td>4.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>M4</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>F</td>
<td>35.288</td>
<td>88.653</td>
<td>139.117</td>
<td>39.589</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Number of Forecasts</td>
<td>N = 281</td>
<td>N = 1123</td>
<td>N = 1117</td>
<td>N = 279</td>
</tr>
</tbody>
</table>
studies useful in explaining future security returns and earnings. We randomly picked another variable, accounts payable, and experimented with the addition of the variable to evaluate whether, with a change in the fundamental signal set, our results will hold. Our results remained unaffected.

Because labor force and capital expenditure variables are only approximations of quarterly measures, we dropped the labor force variable and used an alternative definition of capital expenditure, whose quarterly data could be obtained from the statement of cash flows. The difference between our initial Schedule V measure and the alternative cash flow measure is that the former includes the capital expenditure for the acquired companies while the latter does not. Again, our results were unchanged.

Schuh and Triest (2000) document that multiplant firms tend to own big plants, and that the employment size of plants tends to raise with the number of plants a firm operates. Job flows are higher at single-plant firms and firms operating only a few plants than at multiplant firms operating more plants. One implication of their findings is that firms with large employment may be less likely to experience drastic changes in aggregate employment than firms with small employment. During the period from 1992 to 2001, our sample of 283 firms has an average of 16,213 employees per year, while the rest of the COMPUSTAT firms have only 5,837 employees per year. The sample firms in our study are larger in employment and, therefore, the rate of job creation and destruction will be lower, lending support to our assumption of a constant employment throughout the year.

Park and Pincus (2003) used a quarterly capital expenditure measure that is the same as our alternative measure. We compared this alternative measure to our initial capital expenditure measure, and found that for 283 firms from 1992 to 2001, difference exists between the two measures for 238 firm/years out of a total of 2,830 firm/years. Additionally, we found that, if we had used the alternative measure, the mean capital expenditure would have been 33 cents per share for the first quarter, 37 for the second, 36 for the third, and 41 for the fourth. Thus, by allocating capital expenditure equally to each quarter, we would have overstated the capital expenditure by 3.75 cents for the first, understated 0.25 cents for the second, overstated 0.75 for the third, and understated 4.25 for the last quarter if, again, we had used the alternative measure as in Park and Pincus (2003). Some researchers suggest that new capital projects do not have an immediate impact on earnings but the related depreciation expenses do (Abarbanell & Bushee, 1997). Suppose the newly purchased assets, on average, have a life of 10 years (40 quarters), depreciation expense would have been overstated by 0.094 cents for the first quarter, understated by 0.006 cents for the second, overstated by 0.019 cents for the third, and understated by 0.106 cents for the quarter. Given the low incidence of differences between our initial capital expenditure measure (from Schedule V) and the alternative measure (from cash flow statements), and given the relatively smaller magnitude of overstating or understating in depreciation expenses compared to the magnitude of related expenditure if we had used the alternative measure, we consider our study’s dividing annual capital expenditures equally into four quarters a reasonable approximation of quarter capital expenditures.
**Forecasting more recent periods**

Our forecast period ends with the first quarter of 2002. As we have more recent accounting data available, we are able to extend our forecast period to the first quarter of 2003. After we added the most recent four quarters, all evidence is consistent with previous results whether we tested the four quarters separately or tested a combined sample pooling the more recent period with the original ten quarters. Therefore, the superiority of the multivariate NN model still holds in forecasting more recent periods’ earnings.

Due to the unavailability of more recent returns data (from CRSP Center for Research in Security Prices), we are unable to extend our returns tests to include more quarters. However, we did break down the sample and examined on a quarter-by-quarter basis what is the forecast model best capturing the market’s expectation. Among the ten quarters, the multivariate NN model showed a slightly higher association measure only for the last quarter of our forecast period (the first quarter of 2002). For all the other nine quarters, the univariate models were a better surrogate for market expectations. Although this evidence may suggest the possibility that, starting from the first quarter of 2002 after the market experienced many ups and downs, the market participants might have become more attentive to the fundamental signals of firm earnings and have made more accurate forecasts, we believe that it requires an examination of more recent quarters of 2002 and of 2003 before we can comfortably draw that conclusion.

**Forecasting with stepwise methods**

We tested whether the use of stepwise regressions could significantly increase the forecast accuracy for multivariate linear models of Category 2. Our results show that this statistical procedure of selecting variables did result in improvements in forecast accuracy for Category 2 models compared to when the whole set of variables are used. However, the improved forecast accuracy of multivariate linear models still falls short of that of the univariate linear and NN (both univariate and multivariate) models. Therefore, the inferiority of the multivariate linear models compared to other types of forecast models does not seem to lie in the types of variables included. The findings are consistent with the majority of the studies on comparison between NN based models and linear stepwise regression methods. Six out of seven of such studies we could find show that NN based models are superior to linear stepwise regression methods (Franses & Draisma, 1997; Gorr, 1994; Hardgrave, Wilson, & Walstrom, 1994; Indro, Jiang, Patuwo, & Zhang, 1999; Nath, Rajagopalan, & Ryker, 1997; Tsai, 2000). The only exception is the study conducted by Williams (2002) that shows a regression model produces superior predictions compared to the best performing neural network model.

We also agree with the notion that there are some advantages of stepwise approach for NN models. For instance, the most significant advantage of the stepwise approach is that the weight connections of the network can be updated easily, when a new input is given later after the network has been trained. The stepwise approach also is able to update weights instantly when a new neuron is added to the existing network if the desired error criterion cannot be met. Based on a dynamic stepwise updating algorithm for neural networks proposed by Chen and
Wan (1999), we rerun all the NN models. However, the results show that there is no significant difference between stepwise NN models and the NN models based on the backpropagation algorithm used in the study.

**Industry-specific forecasting models**

Similar to Abarbanell and Bushee (1997), we divided our sample into three general industry sectors: manufacturing (with SIC codes beginning with the numbers 2 and 3), service (with SIC codes beginning with numbers 7 and 8), and wholesale and retail (with SIC codes beginning with 5). Classifying the sample firms into these three sectors retains the majority of the sample (259 firms out of 283 firms), acknowledges differences among individual sectors yet avoids constructing overly many industry-specific models. For the manufacturing sector, we included gross margin, inventory, and labor force variables. For wholesale and retail sectors, we included gross margin and inventory. For the service sector, we included only gross margin. Gross margin is an important predictor for future profitability for all firms, thus is included as a forecast variable for all industries. Inventory is important to manufacturing and wholesale/retail industries, but has little relevance to service industry. Finally, labor force is important to the manufacturing sector as our forecasting period is characterized by both an expansion and a contraction in the demand of labor.

Our results indicate that including different sets of variables for different industries improved the forecasting accuracy of multivariate linear models, and moved Category 2 models closer to univariate ARIMA models in terms of forecasting accuracy. However, with industry-specific models, the multivariate NN method still outperformed all other methods. Thus, our major finding stands even when we consider industry-specific factors in forecasting.

**Generalizability issues**

The previous supplemental tests attest to the robustness of our main argument that the multivariate NN models strongly outperform all other forecasting models. About 98.3% of our sample firms (278 out of 283, untabulated) have at least one quarter where multivariate NN dominates. Among the 2,800 forecasts for all the firms and quarters, multivariate NN dominates for 956 forecasts, accounting for 34.14% of all the forecasts. The next best performer, univariate NN, dominates for only 335 firm-quarters. The sharp discrepancy between the best and the second-best models leads us to believe that, in forecasting quarterly EPS, the multivariate NN method deserves a fair amount of attention regardless the industries the firm belongs to or the time periods we are focused on.

Although the fundamental signals are repeatedly referred to in analysts’ reports and financial statement analysis texts, prior literature also presents some exceptions to the defined relation between the identified fundamental signals and future earnings. For example, accounts receivables and capital expenditures are found to be related to future earnings in an unexpected direction (Abarbanell & Bushee, 1997). Additionally, the behavioral bias of financial analysts in describing their forecasting experience can also contribute to the bias in the search of fundamentals (Slovic, Fleissner, & Bauman, 1972; Abarbanell & Bushee, 1997).
In view of these limitations, we suggest that a comparison of forecast accuracy between the multivariate NN model and financial analysts’ judgmental forecasts is much warranted to further evaluate the performance of the multivariate NN model. Nonetheless, we believe that the NN model incorporating fundamental accounting signals represents a promising decision aid in the domain of forecasting quarterly EPS.

CONCLUSIONS

The purpose of this paper was to examine the accuracy of using NN in forecasting EPS. We document that the neural network approach improves forecast accuracy, whether for the univariate or multivariate models. However, the improved forecasting accuracy is more pronounced when we include a collection of fundamental accounting variables. The evidence further indicates that the fundamental signals possess incremental value with respect to future quarterly EPS forecasts, but the incremental value of the fundamental information can only be achieved in a nonlinear manner.

It should be noted that our evidence on the superiority of univariate NN forecast models over linear ARIMA models (albeit weak) contradicts that documented by Callen et al. (1996). In their study, Callen et al. (1996) identify their models using a longer time-series and validate the models using a longer holdout period. Constrained by the data availability of fundamental information, we are unable to obtain long time-series spanning as many quarters as theirs. This may suggest the additional feature that the neural network approach performs better when a limited number of observations are available to researchers, and that the performance of the neural network is context-sensitive.

We also examine the relative ability of our forecast models in approximating the market expectations, and find that more accurate forecasts do not capture the market’s expectations better. This implies that for the period that we are forecasting, investors have not adopted a more accurate forecasting method (i.e., the multivariate NN method), probably either because of their neglecting the fundamental accounting information in valuing securities or because of their failure to incorporate the fundamental signals into future earnings in an appropriate manner. Either explanation will be the indirect evidence that the investors have not used the available information to the fullest extent. Therefore, the limitations in the investors’ decision-making process are suggested.

It will be interesting to see if in the long run, the market revises its valuations of firm value based on quarterly fundamental variables, and as a result, security valuation becomes more aligned with the forecasts of the multivariate-neural network model. It will also be interesting to see whether in the long run, portfolios formed on the basis of the quarterly EPS forecasts from a neural network model incorporating quarterly fundamental information outperform portfolios based on other forecast models. Finally, comparing forecasting accuracy of financial analyst forecasts with that of the mechanical methods we examined in this paper is important to further developing the neural network approach into a decision aid to assist financial analysts’ forecasting of quarterly EPS. We leave these issues to be addressed by future research. [Received: May 2003. Accepted: February 2004.]
REFERENCES


**Wei Zhang** is an assistant professor of accounting in the School of Business at Clarkson University. She holds a PhD from the College of Business Administration at the University of Nebraska-Lincoln. She has published in journals such as the *Journal of Accounting, Accountability & Performance*. Her current research interests are in the areas of insider trading, incentives and compensation, analyst forecasts, and earnings management.

**Qing Cao** is an assistant professor of management information systems in the H.W. Bloch School of Business & Public Administration at the University of Missouri-Kansas City. He holds a PhD from the College of Business Administration at the University of Nebraska-Lincoln. He has coauthored a book, *E-commerce Operations Management*. He has published in journals such as *International Journal of Production Research, Journal of Database Management, International Journal of Production Economics, Quality Management Journal*, and *International Journal of Physical Distribution and Logistics*. His current research interests are software quality metrics, electronic commerce, supply chain information management, and artificial intelligence applications.
