The adaptive selection of financial and economic variables for use with artificial neural networks

Suraphan Thawornwong, David Enke*

Intelligent Systems Center, 1870 Miner Circle 204, Eng Management, University of Missouri - Rolla, Rolla, MO 65409-0370, USA

Received 25 October 2001; received in revised form 20 March 2003; accepted 30 May 2003

Abstract

It has been widely accepted that predicting stock returns is not a simple task since many market factors are involved and their structural relationships are not perfectly linear. Recently, a promising data mining technique in machine learning has been proposed to uncover the predictive relationships of numerous financial and economic variables. Inspired by the fact that the determinant between these variables and their interrelationships over stock returns changes over time, we explore this issue further by using data mining to uncover the recent relevant variables with the greatest predictive ability. The objective is to examine whether using the recent relevant variables leads to additional improvements in stock return forecasting. Given evidence of non-linearity in the financial market, the resulting variables are then provided to neural networks, including probabilistic and feed-forward neural networks, for predicting the directions of future excess stock return. The results show that redeveloped neural network models that use the recent relevant variables generate higher profits with lower risks than the buy-and-hold strategy, conventional linear regression, and the random walk model, as well as the neural network models that use constant relevant variables.

Keywords: Neural networks; Variable relevance analysis; Financial and economic variables; Stock market prediction

1. Introduction

Over the past two decades many important trends have changed the environment of the financial markets. Traditional capital market theory has been adapted and methods...
of financial analysis have been improved [35]. Investors are becoming more dependent on advanced computer and communication technologies to benefit from a wider range of investment choices [13]. Artificial neural networks are one of the technologies that have caused the most recent excitement in this financial environment. They provide an interesting technique that theoretically can approximate any non-linear continuous function on a compact domain to any designed degree of accuracy [9]. The novelty of neural networks lies in their ability to model non-linear processes without a priori assumptions about the nature of the generating process [20]. This is useful in security investment and other financial areas where much is assumed, and little is known about the nature of the processes determining asset prices [5].

Recently, promising results were obtained by incorporating a data mining technique used in machine learning to uncover the predictive powers of numerous financial and economic variables [45]. This approach seems particularly attractive in selecting the variables when the usefulness of the data is unknown, especially when non-linearity exists in the financial market. In the study, the forecasting models were developed to estimate the value (level) of future excess stock return on the S&P 500 stock portfolio. A consecutive experiment reported in [47] continued with a similar objective but different approach. That is, the neural networks were modeled to predict (classify) the direction (sign) of future excess stock return. The results indicate that the trading strategies guided by the neural network classification models generated higher risk-adjusted profits than the buy-and-hold strategy and those guided by the level-estimation models.

Driven by the fact that historical data may not fully represent patterns of current stock behavior, this paper uses the above-mentioned data mining technique to uncover the predictive powers of various financial and economic variables under different time periods. The objective is to conduct an investigation of the predictive effect of recent (adaptive) relevant variables on the accuracy of stock return forecasting. The specific objective is to evaluate the impact of historical data on profitability based on trading guided by several forecast models. The sign of excess stock return generated by the developed models, as compared to a risk-free return in a one-month T-bill, is used as a trading decision since it provides a measure of how well the models perform relative to the minimum returns gained from depositing the money in a risk-free account [25].

The remainder of this paper is organized as follows: a brief history of stock return forecasting is provided in the next section. The third section discusses the methodology for variable selection and the relevant variables that were obtained from the data mining analysis. The fourth section briefly reviews the two types of neural networks used for classification. Neural network modeling to predict the sign of future excess stock returns on the S&P 500 stock portfolio is also described in this section. The fifth section gives the resulting data selection and model development of the classical linear regression that is used as the benchmark for performance comparisons. The sixth section explains the methodology used to evaluate the effect of recent relevant variables on stock return forecasting. The performance measures used in this study are also discussed in this section. Empirical results are reported in the seventh section. The discussion and conclusion are summarized in the eighth and ninth sections, respectively. Lastly, the data source and description are given in the appendix.
2. Predicting stock returns

Stock return or stock market prediction is an important financial subject that has attracted researchers’ attention for many years. It involves an assumption that past publicly available information has some predictive relationship to future stock returns [14,15]. The samples of such information include economic variables such as interest rates and exchange rates, industry specific information such as growth rates of industrial production and consumer price, and company specific information such as income statements and dividend yields. An attempt to predict stock returns, however, is opposed to the general perception of market efficiency. As noted by [22], the efficient market hypothesis states that all available information affecting the current stock values is accounted for by the market before the general public can make trades based on it. Therefore, it is impossible to forecast future prices since they already reflect everything that is currently known about the stocks. It is also believed that an efficient market will instantaneously adjust prices of stocks based on news which arrives at the market in a random fashion [2]. This line of reasoning supports the rationale for the so-called “random walk” model which implies that the best prediction of the next period’s stock price is simply its current value [27]. Nonetheless, this is still a debating issue because there is considerable evidence that markets are not fully efficient, and it is possible to predict the future stock prices or indices with results that are better than random [26].

During the past few years there is considerable evidence to prove that markets are not fully efficient. In fact, many researchers provide evidence that stock market returns are predictable by means of publicly available information such as time-series data on financial and economic variables, especially those with an important business cycle component [3,4,6,16–19,23,42]. These studies identify that such variables as interest rates, monetary growth rates, changes in industrial production, and inflation rates are statistically important for predicting a portion of the stock returns. However, most of the studies attempting to capture the relationship between the available information and the stock returns rely on linear assumptions. Practically, there is no evidence that the relationship between the stock returns and the financial and economic variables is perfectly linear. This is due to the fact that there exists significant residual variance of the actual stock return from the prediction of the regression equation. Therefore, it is possible that non-linear models may be able to better explain this residual variance and produce more reliable predictions of the stock price movements [29,36].

In recent years, the discovery of non-linear movements in the financial markets has been greatly emphasized by various researchers and financial analysts (see [1]). Even though there are a number of non-linear statistical techniques that have been used to produce better predictions of future stock returns or prices, most techniques are model-driven approaches which require that the non-linear model be specified before the estimation of parameters can be determined. In contrast, neural networks are data-driven approaches which do not require a pre-specification during the modeling process because they independently learn the relationship inherent in the variables. Thus, neural networks are capable of performing non-linear modeling without a priori knowledge about the relationship between input and output variables. As a result, there has been a growing interest in applying neural networks to capture future stock
behaviors; see [25,2,7,8,48,11,30,38,37] for the previous work on the S&P 500 stock predictions.

Although there currently exists a vast number of articles addressing the predictabilities of stock market return, most studies rely on various input variables and assumptions. Often no justification is given as to why a particular series of inputs were chosen. While uncertainty in selecting the predictive variables to forecast stock returns still exists, as can be observed from a variety of input variables used in a recent literature survey [46], no studies however have incorporated all available variables previously mentioned in the literature to uncover input data that may be effective in predicting stock returns. Obviously, many of variables used may be irrelevant or redundant to the prediction of stock returns. In fact, leaving out relevant variables or keeping irrelevant variables may be detrimental, causing confusion to the neural networks. Unfortunately, there is no consistent method that has been used to pick out the useful variables in stock return forecasting. This may be due to the fact that the behavior of this data is not well known. Obviously, a systematic approach to determining what inputs are important is necessary.

3. Methodology for variable selection

There have been many studies in various areas of data mining (e.g., machine learning, fuzzy logic, statistics, and rough set theory) on variable relevance analysis for data understanding [21]. The general idea behind variable relevance analysis is to compute some measures that can be used to quantify the relevance of variables hidden in a large data set with respect to a given class or concept description. Such measures include information gain, the Gini index, uncertainty, and correlation coefficients.

An inductive learning decision tree algorithm that integrates an information gain analysis technique with a dimension-based data analysis method was selected for this study as it can be effectively used for variable subset selection [21]. The resulting method removes the less information producing variables and collects the variables that produce more information. Therefore, it may be the most appropriate data mining technique to perform variable subset selection when the usefulness of the data is unknown. While using the information gain analysis technique, the predicted directions of excess stock return were used as class distributions for the experiment. The resulting variables with high information are chosen as the relevant input variables provided to the neural network models. The following paragraphs give a brief introduction to the information gain calculation. Readers who are interested in full details of the information gain algorithm should refer to [39].

Let $S$ be a set consisting of $s$ data samples. Suppose the class label variable has $m$ distinct values defining $m$ distinct classes, $C_i$ (for $i = 1, 2, \ldots, m$). Let $s_i$ be the number of samples of $S$ in class $C_i$. The expected information needed to classify a given sample is given by

$$I(s_1, s_2, s_3, \ldots, s_m) = - \sum_{i=1}^{m} p_i \log_2 (p_i),$$

(1)
where \( p_i \) is the probability that an arbitrary sample belongs to class \( C_i \) and is estimated by \( s_i/s \). Note that a log function to the base 2 is used since the information is encoded in bits. Let variable \( A \) have \( v \) distinct values denoted in order from small to large values as \( \{a_1, a_2, a_3, \ldots, a_v\} \). Any split value lying between \( a_i \) and \( a_{i+1} \) will have the same effect of dividing the samples into those whose value of the variable \( A \) lies in \( \{a_1, a_2, a_3, \ldots, a_i\} \) and those whose value is in \( \{a_{i+1}, a_{i+2}, a_{i+3}, \ldots, a_v\} \). However, the midpoint of each interval is usually chosen as the representative split, defined as \((a_i + a_{i+1})/2\). Thus, there \( v-1 \) possible splits on \( A \), all of which are examined. Note that examining all \( v-1 \) splits is necessary to determine the highest information gain of \( A \).

Variable \( A \) can therefore be used to partition \( S \) into 2 subsets, \( \{S_1, S_2\} \), where \( S_j \) contains those samples in \( S \) that have values \( \{a_1, a_2, a_3, \ldots, a_i\} \) or \( \{a_{i+1}, a_{i+2}, a_{i+3}, \ldots, a_v\} \) of \( A \). Let \( S_j \) contain \( s_{ij} \) samples of class \( C_i \). The expected information based on this partitioning by \( A \), also known as the “entropy” of \( A \), is given by

\[
E(A) = \sum_{j=1}^{v} \frac{s_{ij} + s_{2j} + \cdots + s_{mj}}{s} I(s_{1j}, s_{2j}, \ldots, s_{mj}).
\]

The term \((s_{1j} + s_{2j} + \cdots + s_{mj})/s\) acts as the weight of the \( j \)-th subset and is the number of samples in the subset (i.e., having value \( a_j \) of \( A \)) divided by the total number of samples in \( S \). Note that for a given subset \( S_j \),

\[
I(s_{1j}, s_{2j}, \ldots, s_{mj}) = -\sum_{i=1}^{m} p_{ij} \log_2(p_{ij}),
\]

where \( p_{ij} = s_{ij}/s_j \) and is the probability that a sample in \( S_j \) belongs to class \( C_i \). The information gain obtained by this partitioning of the split on \( A \) is defined by

\[
\text{Gain}(A) = I(s_1, s_2, s_3, \ldots, s_m) - E(A).
\]

In this approach to relevance analysis, the highest information gain for each of the variables defining the samples in \( S \) can be obtained. The variable with the highest information gain is considered the most discriminating variable of the given set. By computing the information gain for each variable, a ranking of the variables can be obtained. Finally, the relevant threshold is determined to select only the strong relevant variables to be used in the forecasting models.

In this study, \( PP_{t-1}, CP_{t-1}, IP_{t-1}, M_{1_t-1}, T_{3t}, T_{6_t}, T_{12_t}, T_{60_t}, T_{120_t}, CD_{1_t}, CD_{3_t}, CD_{6_t}, AAA_t, BAA_t, DIV_t, T_{1_t}, SP_t, DY_t, TE_{1_t}, TE_{2_t}, TE_{3_t}, TE_{4_t}, TE_{5_t}, TE_{6_t}, DE_{1_t}, DE_{2_t}, DE_{3_t}, DE_{4_t}, DE_{5_t}, DE_{6_t}, \) and \( DE_{7_t} \) are the 31 financial and economic variables which were available for sign prediction of excess stock return (\( ER_{t+1} \)) on the S&P 500 stock portfolio (the source and definition of all the variables are given in the appendix). These variables were collected monthly from March 1976 to December 1999, for a total of 286 months. They contained a mixture of the predictive variables reported in [25,11,30,38]. However, two variables often used in the literature, long-term treasury rates and commercial papers, were not applicable due to the fact that the 30-year treasury rate provided by the Federal Reserve Board of Governors started
The first and second period sets for each sliding period

<table>
<thead>
<tr>
<th>Sliding periods</th>
<th>First period set (months)</th>
<th>Second period set (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training and validation</td>
<td>Total</td>
</tr>
</tbody>
</table>

from February 1977, while the series of commercial papers had been discontinued because of a change in methodology as of September 1997. Therefore, several financial instruments, such as CD and T-bill rates with additional maturities, were included to supplement unavailable data in the study.

Generally, large-scale deterministic components, such as trends and seasonal variations, should be eliminated from the inputs since the networks will attempt to learn the trend and use it in the prediction [31]. Therefore, the above-mentioned variables, excluding $DIV, T1, SP, DY$, and $ER$, were seasonally adjusted allowing the networks to concentrate on the important details necessary for an accurate prediction. In addition, due to the lag associated with the publication of macroeconomic indicators as mentioned in [38], certain data, in particular $PP, IP, CP$, and $M1$, were included in the base set with a 2-month time lag. The rest of the variables were included in the base set with a 1-month time lag. Constructing the data in this manner ensures that the forecasting models using these variables will be similar to real-world practice. Specifically, only observable (not future) data were employed as inputs to the forecasting models. As a result, these time lags were used throughout the experiment to maintain realistic situations when data are gathered. In this study, the differences $[Pt - Pt_{-1}]$ of variables were provided to the networks so that different input variables can be compared in terms of change with the monthly excess stock return, since the level changes of the variable when forecasting financial time series may be more meaningful to the models than those using the original values [32].

For comparison of all models, the whole data samples were divided into four sliding periods. This was done for analyzing the predictive effect of adaptive relevant variables in different time periods and for establishing the robustness of the out-of-sample forecasting performance. Each of these sliding periods was further split into two period sets. The first period set of each sliding period was used for training and validating the forecasting models, while the second period set was reserved for out-of-sample evaluation and comparison of performance among the forecasting models. Table 1 represents the four sliding periods used to examine the predictive powers of the financial and economic variables evaluated in the study. The selection of input variables is a modeling decision that can greatly affect the model performance. For the neural network modeling, the information gain data mining analysis was employed to find subsets of relevant variables from the 31 variables available in the study. After performing the analysis for all of the first period sets, the variables that indicate strong relevant
information for each set were obtained, as presented in Table 2. These selected variables were therefore used consistently as the input variables for training and validating the neural networks throughout the modeling stage.

As illustrated in Table 2, the strong relevant variables determined by the information gain data mining analysis varied from sliding period to sliding period. In fact, the variance of these variables was essentially caused by the 22 sliding months. This is no surprise since it has been long believed in the finance literature that the determinant between the variables and their interrelationships over stock returns changes over time [40]. Therefore, this finding suggests the importance of choosing appropriate historical data for stock return forecasting since some data may not fully represent current patterns of stock behavior. When these relevant variables were examined more closely, it was observed that six variables, namely $CP; M$, $1; T$, $3; T$, $6; T$, $60; T$, $CD$, $1; CD$, $3; CD$, $6; CD$, $SP; TE$, $2; TE$, $3; TE$, $4; TE$, $DE$, $2; DE$, $3; DE$, $5; DE$, $7; DE$, appeared in three out of the four sets of the sliding periods. This suggests that these variables were reasonably important for explanation of the predicted sign of excess stock return on the S&P 500 stock portfolio. More importantly, there are three variables, including $T$, $6; SP$, and $TE$, $2; TE$, which were consistently included in all sets of the sliding periods. This indicates that these three variables were particularly significant for predicting the signs of excess stock return in the study.

On the other hand, there are only three variables, namely $PP; T$, $12$, and $T$, $1$, which were not incorporated in any first period sets of all sliding periods evaluated in this study. This observation implies that most variables selected by the recent studies were important for explaining residual variance of the stock returns. This may also explain why the proposed forecasting models developed in those studies relied on various input variables. This is especially true when the study periods were different, diversely affecting their criteria for variable selection. In fact, it may be for this reason that the selected variables were effective during their respective modeling efforts, even when the models or the performance criteria were different.

4. Neural network models

The theory of neural network computation provides interesting techniques that mimic the human brain and nervous system. A neural network is characterized by the pattern of connections among the various network layers, the numbers of neurons in each
layer, the learning algorithm, and the neuron activation functions. In general, a neural network is a set of connected input and output units where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to correctly predict or classify the output target of a given set of input samples. Given the numerous types of neural network architectures that have been developed in the literature, two important types of neural networks, namely the probabilistic and feed-forward neural networks, were implemented in this study to compare their predictive ability against a classical linear regression model. The following two subsections give a brief introduction to these neural network models. The network modeling is presented in the third subsection.

4.1. Feed-forward neural network

Feed-forward neural networks have been widely used for financial forecasting due to their ability to correctly classify and predict the dependent variable [49]. For each training sample the input variables are fed simultaneously into a layer of processing units making up the input layer. The weighted outputs of these units are, in turn, fed simultaneously to a second layer of processing units known as a hidden layer. The hidden layer’s weighted outputs can be input to another hidden layer, and so on. The weight outputs of the last hidden layer are input to units making up the output layer which issues the network’s prediction for a given set of samples.

Backpropagation is by far the most popular neural network training algorithm that has been used to perform learning on feed-forward neural networks. It is a method for assigning responsibility for mismatches to each of the processing units in the network, which is achieved by propagating the gradient of the activation function back through the network to each hidden layer, down to the first hidden layer. The weights are then modified so as to minimize the mean squared error between the network’s prediction and the actual target. This is a supervised learning procedure that attempts to minimize the error between the desired and predicted outputs. The output value for a unit \( j \) is given by the following function:

\[
O_j = G \left( \sum_{i=1}^{n} w_{ij} x_i - \theta_j \right),
\]

where \( x_i \) is the output value of the \( i \)th unit in the preceding layer, \( w_{ij} \) is the weight on the connection from the \( i \)th unit, \( \theta_j \) is the threshold, \( n \) is the number of units in the preceding layer, and \( G() \) is the activation function. Fig. 1 shows the configuration of a single-hidden layer neural network which will be used later in this study. Since feed-forward neural networks are well known, the various network structures and backpropagation algorithms are not described here. Readers who are interested in greater detail can refer to [41] for an explanation of the backpropagation algorithm used to train feed-forward neural networks.

During neural network modeling, it has been suggested that the proper number of hidden layer nodes requires validation techniques to avoid under-fitting (too few neurons) and over-fitting (too many neurons) [28]. Generally, too many neurons in the hidden layers, and, hence, too many connections, produces a neural network that memorizes
the data and lacks the ability to generalize. One approach that can be used to avoid over-fitting is $n$-fold cross-validation [34]. A five-fold cross-validation was used in this experiment and can be described as follows: The data sample is randomly partitioned into five equal-sized folds and the network is trained five times. In each of the training passes, one fold is omitted from the training data and the resulting model is validated on the cases in that omitted fold, which is also known as a validation set. The first period set (200 months) of each sliding period is used for the five-fold cross-validation experiment, leaving the second period set for truly untouched out-of-sample data. The average root-mean squared error over the five unseen validation sets is normally a good predictor of the error rate of a model built from all the data.

Another approach that can be used to achieve better generalization in trained neural networks is called early stopping [10]. This technique can be effectively used with the cross-validation experiment. The validation set is used to decide when to stop training. When the network begins to over-fit the data, the error on the validation cases will typically begin to rise. In this study the training was stopped when the validation error increased for five iterations, causing the return of the weights and biases to the minimum of the validation error. The average error results of the validation cases (40 months in each fold) from the $n$-fold cross-validation experiment can finally be used as criteria for determining the optimal network structure, namely the number of hidden layers, number of neurons, learning algorithms, learning rates, and activation functions.

4.2. Probabilistic neural network

Like feed-forward neural networks, the probabilistic neural network (PNN) can be used for classification problems. The PNN is a parallel implementation of a non-parametric method called Parzen windows and is a four-layer network that can perform pattern classification [43]. It is based essentially on the estimation of probability density functions for various classes learned from training samples. The PNN learns from the sample data instantaneously and uses these probability density functions to compute the non-linear decision boundaries between classes in a way that is related to the Bayes
optimal [12]. The PNN formula can be briefly explained as follows:

\[
f_1(x) = \frac{1}{(2\pi)^{p/2}\sigma^p n} \sum_{i=1}^{n} z_i, \tag{6}\]

where \(f_1(x)\) is the probability density function estimator for class 1, \(p\) is the dimensionality of training vector, \(z_i = \exp[-D_i/(2\sigma^2)]\) is the output of hidden neuron, \(D_i = (x-u_i)^T(x-u_i)\) is the distance between the input vector \(x\) and the training vector \(u\) from category 1, and \(\sigma\) is a smoothing parameter.

Theoretically, the PNN can classify out-of-sample data with the maximum probability of success when enough training data is given [50]. Fig. 2 presents the PNN architecture. When an input is presented, hidden layer 1 computes distances from the input vector to the training vectors and produces a resulting vector whose elements indicate how close the input is to the vectors of the training set. Hidden layer 2 then sums these elements for each class of inputs to produce a vector of probabilities as its net output. Finally, the activation function of the PNN output layer picks the maximum of these probabilities and places it into specific output classes.

4.3. Neural network modeling

For the feed-forward neural networks, a sigmoid hyperbolic tangent function was selected as the activation function to generate an even distribution over the input values. A single hidden layer was also chosen for the neural network model since it has been successfully used for financial classification and prediction [44]. Accordingly, the feed-forward neural networks were built with three layers, including the input layer, hidden layer, and output layer. Each of the relevant input variables was assigned a separate input neuron within the input layer. For this reason, the five-fold cross-validation experiment of the first period sets (of the sliding periods) 1, 2, 3, and 4 contained 15, 10, 13, and 16 neurons (see Table 2) at the input layer, respectively. For instance, each of the 15 relevant variables for the first period set 1 was assigned separately to a neuron of the input layer containing 15 input neurons. Since there are two classes
of the signs of excess stock return, two output neurons were employed for the output layer to represent the different classes of predicted excess stock return (see Fig. 1, with $K = 2$). In this study, the vectors $[+1-1]$ and $[-1+1]$ represent the predicted positive and negative signs of excess stock return, respectively.

In addition, the resilient backpropagation-learning algorithm was employed to train the feed-forward neural networks since this optimization method is generally much faster than the standard steepest descent algorithm and requires only a modest increase in memory requirements. The output neuron with the highest value was taken to represent the predicted sign of excess stock return based on a given set of input variables. Note that the values of the relevant input variables for each of the first period sets were preprocessed by normalizing them within a range of $-1$ and $+1$ to minimize the effect of magnitude among the inputs and increase the effectiveness of the learning algorithm. The connection weights and biases were initially randomized and then set during the backpropagation training process. The appropriate learning rate for maximizing performance was also determined during neural network training.

After many experimentations, the feed-forward neural networks employing 27, 14, 15, 22 neurons in the hidden layer were found to be the best network architectures with the lowest average root-mean squared error (RMSE) over the five fold cross-validation experiment of the first period sets 1, 2, 3, and 4, respectively. Note that the RMSE used in this experiment is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} ((y_{1i} - t_{1i})^2 + (y_{2i} - t_{2i})^2)}{2m}}, \quad (7)$$

where $y_1$ and $y_2$ are the predicted classes of excess stock return of the two output neurons, $t_1$ and $t_2$ are the actual classes of excess stock return, and $m$ is the number of validation cases (40 in this study). By conducting the five-fold cross-validation experiment, the forecasting results will not be based on only one network output since five neural network models of each first period set were developed from the five different data sets. In other words, the predicted sign of excess stock return based on a given set of input variables can be derived from each one of the five network models. Therefore, in this study the majority of the signs of the five network outputs were used to determine the decisive predicted sign of excess stock return. For example, when the five network models generate three positive- and two negative-predicted signs of excess stock return based on a given set of input variables, the decisive predicted sign of the excess stock return is resolved to be positive.

In addition, a portfolio neural network model consisting of the network architecture producing the lowest RMSE in each omitted fold cross-validation experiment of the first period set was also explored. Therefore, each neural network model generating the “lowest” RMSE from each omitted fold experiment was chosen as one of the five neural networks deliberately combined as the portfolio network model. The resulting portfolio network model using the lowest RMSE in each omitted fold experiment is provided in Table 3. It is observed from each of the first period sets that the suitable number of neurons used in the hidden layer of the five combined portfolio networks that
were trained based on the various omitted folds were in fact different. This observation suggests the importance of network modeling for each separate omitted fold experiment since they may potentially lead to better trained (generalized) neural networks that produce a lower RMSE on the validation cases. Again, the decisive predicted sign of excess stock return of the portfolio network model was derived from the majority of the five combined portfolio network outputs.

Unlike the feed-forward neural networks, the design of the PNN is fast and straightforward. In fact, neither a training nor early stopping technique are required during its design. This produces no need to randomly partition the data into equal-sized folds for a cross-validation experiment. Therefore, the first period sets (200 months) for each sliding period were used in network training for predicting the sign of excess stock returns of the second period sets. In this study, a smoothing parameter equal to 1.00 was selected to consider all nearby design vectors. For each of the first period set experiments, the PNN design employed the same input variables and pre-processing techniques as those of the feed-forward neural network models.

5. Linear regression forecast

The backward stepwise regression for dimensionality reduction was employed to assume a linear additive relationship for the classical linear regression forecasts. This technique started with the full set of variables in the model. The worst of the original variables was determined and removed from the full set. At each subsequent iteration or step, the worst of the remaining variables was removed from the last updated set. The significant $t$-statistics were used as criteria for retention of the significant input variables in the linear regression model. The remaining variables were then used in predicting excess stock returns. Table 4 provides the significant variables for each of the linear regression models after the backward stepwise regression elimination technique was performed on the first period sets.

According to Table 4, the backward stepwise technique kept 10, 12, 13, and 12 variables as the significant relevant variables in the regression models ($\alpha = 0.05$) for the first period sets 1, 2, 3, and 4, respectively. The resulting regression models have
Table 4
The relevant variables for each of the first period sets (backward stepwise analysis)

<table>
<thead>
<tr>
<th>First period set</th>
<th>Significant relevant variables</th>
<th>Number of variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PP, M1, T3, T12, T60, CD1, CD6, BAA, SP, DE7</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>PP, IP, M1, T6, T60, CD1, CD3, CD6, T1, DY, TE4, TE5</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>PP, CP, IP, T6, T12, CD1, CD6, BAA, DIV, T1, DY, TE2</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>IP, M1, T3, T12, T60, T120, CD6, DIV, SP, TE5, TE6, DE4</td>
<td>12</td>
</tr>
</tbody>
</table>

the following functions:

Period 1: \( ER_{t+1} = -0.444 + (0.959 \times PP_{t-1}) + (0.100 \times M1_{t-1}) \)
+ \((2.525 \times T3_t) + (5.981 \times T12_t) + (-4.584 \times T60_t)\)
+ \((-1.050 \times CD1_t) + (-5.472 \times CD6_t) + (-0.437 \times BAA_t)\)
+ \((-0.027 \times SP_t) + (8.295 \times DE7_t)\),

Period 2: \( ER_{t+1} = -0.362 + (0.850 \times PP_{t-1}) + (0.145 \times IP_{t-1}) \)
+ \((0.099 \times M1_{t-1}) + (-8.126 \times T60_t) + (-1.754 \times T60_t)\)
+ \((1.243 \times CD1_t) + (-4.833 \times CD3_t) + (5.730 \times CD6_t)\)
+ \((3.727 \times T1_t) + (2.871 \times DY_t) + (-2.781 \times TE4_t)\)
+ \((3.209 \times TE5_t)\),

Period 3: \( ER_{t+1} = 0.414 + (0.967 \times PP_{t-1}) + (-0.679 \times CP_{t-1}) + (0.174 \times IP_{t-1}) \)
+ \((-8.536 \times T6_t) + (2.070 \times T12_t) + (1.258 \times CD1_t)\)
+ \((-5.282 \times CD3_t) + (6.304 \times CD6_t)\)
+ \((-0.452 \times BAA_t) + (0.956 \times DIV_t)\)
+ \((0.479 \times T1_t) + (2.790 \times DY_t) + (-3.904 \times TE2_t)\),

Period 4: \( ER_{t+1} = 0.588 + (0.369 \times IP_{t-1}) + (-0.052 \times M1_{t-1}) \)
+ \((-4.535 \times T3_t) + (8.161 \times T12_t)\)
+ \((2.694 \times T60_t) + (-4.216 \times T120_t)\)
+ \((3.434 \times CD6_t) + (2.398 \times DIV_t)\)
+ \((-0.027 \times SP_t) + (19.367 \times TE5_t) + (-20.219 \times TE6_t)\)
+ \((-2.194 \times DE4_t)\).
where all the regression coefficients were significant and the $F$-statistics for the first period sets 1, 2, 3, and 4 were $2.027$ ($p$-value $0.033$, $F_{crit} = 1.91$), $1.946$ ($p$-value $0.031$, $F_{crit} = 1.83$), $1.904$ ($p$-value $0.032$, $F_{crit} = 1.80$), and $1.914$ ($p$-value $0.035$, $F_{crit} = 1.83$), respectively. These indicate that the retained relevant variables contained information about future excess stock returns. By grouping the regression coefficients for all of the first period sets, the regression models show that the changes of $PP, IP, T12, DIV, T1, DY, TE5$, and $DE7$ had a positive effect on predictions of excess stock return, whereas the effect of $CP, T6, T120, CD3, BAA, SP, TE2, TE4, TE6$, and $DE4$ on excess stock returns was negative. Interestingly, it was found that $M1, T3, T60, CD1$, and $CD6$ had both positive and negative effects on excess stock return predictions depending on the period of time when the variable was analyzed.

In addition, the significant variables obtained from the backward stepwise technique varied from sliding period to sliding period, which is also similar to those determined by the data mining analysis. This result once again confirms that the structural relationship of financial and economical variables over stock returns changes over time and that this interrelationship should be determined frequently to insure representation of the latest behavior of current stock return. It is also noted that the backward stepwise technique retained one variable ($CD6$) in the first period sets of all sliding periods and kept six variables ($PP, M1, IP, T12, T60$, and $CD1$) in three out of the four sets. This indicates that these variables are important for linear explanation of the prediction of excess stock return on the S&P 500 stock portfolio. In contrast, there are eight variables, namely $AAA, TE1, TE3, DE1, DE2, DE3, DE5$, and $DE6$, which were consistently removed from all first period sets of the sliding periods. This result suggests that using these removed variables to model the linear regression forecast may not arrive at more accurate predictions of S&P 500 stock portfolio return.

6. Adaptive and constant modeling

To examine the predictive effect of adaptive relevant variables on forecasting accuracy of stock return, an experiment involving forecasting models that utilize adaptive relevant variables was conducted. Four forecasting models, namely the original classification feed-forward neural network using the lowest average $RMSE$ (original NN), the portfolio classification feed-forward neural network using the lowest $RMSE$ in each omitted fold (portfolio NN), the PNN, and the linear regression model (regression) were compared in this study. Since the models were repeatedly developed four times using the four first period sets (see Table 1), there are three modeling comparisons (MCs) that can be used to evaluate whether frequent remodeling efforts of adaptive relevant variables leads to additional improvements in stock return forecasting. The experimental design used to evaluate the predictive effects of the three MCs between the constant and adaptive modeling is presented in Table 5.

According to Table 5, the forecasting models developed by the first period sets 1, 2, and 3 were further used to predict the signs of excess stock return until December 1999. For instance, the portfolio NN that was developed using the first period set 1
Table 5
The experimental design for constant and adaptive modeling forecasts

<table>
<thead>
<tr>
<th>Modeling comparisons (MC)</th>
<th>Forecasting models (developed using the first period set)</th>
<th>Out-of-sample forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant modeling (CM)</td>
<td>Adaptive modeling (AM)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1–4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2–4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3–4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1–4 11/92–12/99 86</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2–4 09/94–12/99 64</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3–4 07/96–12/99 42</td>
</tr>
</tbody>
</table>

of the sliding period 1 is also used to predict the signs of excess stock return of the second period sets 2, 3 and 4, in addition to those of the second period set 1. In other words, this portfolio NN is employed to forecast the signs of excess stock return of the last 64 out-of-sample months (09/1994–12/1999) aside from those of the original 22 out-of-sample months (11/1992–08/1994) in the sliding period 1. The out-of-sample forecasts using the forecasting model which was developed by one set of relevant variables is called constant modeling (CM) for this study. The results generated using the CM will then be used to compare against results generated by the forecasting models that were repeatedly developed to forecast the specific second period sets (as given in Table 1). For example, the four portfolio NNs (developed by the first period sets 1, 2, 3, and 4) are used to forecast the signs of excess stock return of the specific months (see Table 1), as compared against the portfolio NN of the CM (developed by the first period set 1). The out-of-sample forecasts using the forecasting model which was redeveloped by more than one set of relevant variables is referred to as adaptive modeling (AM) in this study.

6.1. Performance measures

The predictive forecasting performances of the developed models are evaluated using the untouched out-of-sample data (second period sets). This is due to the fact that the superior in-sample performance does not always guarantee the validity of the forecasting accuracy. One possible approach for evaluating the forecasting performance is to investigate whether traditional error measures such as those based on the RMSE or correlation between the actual out-of-sample returns and their predicted values are small or highly correlate, respectively. However, there is some evidence in the finance literature suggesting that traditional measures of forecasting performance may not be strongly related to profits from trading [33]. An alternative approach is to look at the proportion of time that the signs of excess stock returns (SIGN) are correctly predicted. In fact, the forecast performance based on the sign measure matches more closely to the profitability performance than do traditional criteria [24]. Recently, a profitability performance guided by certain forecasts and trading strategies has been used extensively to measure how well the forecasting models perform [25,2,7,8,48,11,30,38,37]. Therefore, the SIGN and returns based on trading were selected as the performance measures to
report the predictability and profitability results of the developed forecasting models in this study.

7. Empirical results

The average performance of the second period sets for both the CM and AM was used for comparing the robustness of the forecasting models. In fact, it is impossible to report all the details of the results because there were a total of 72 second period set forecasts for each of the four forecasting models (the original NN, portfolio NN, PNN, and regression) and for the 18 second period sets of the three MCs (eight, six, and four second period set forecasts of MC1, MC2, and MC3, respectively). After performing all out-of-sample forecasts, the average SIGN over the second period sets between the two modeling efforts were calculated and are provided in Table 6.

Table 6
Predictability and profitability results between constant and adaptive modeling

<table>
<thead>
<tr>
<th></th>
<th>SIGN</th>
<th>Monthly return</th>
<th>Std. of return</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>0.6273</td>
<td>1.51</td>
<td>2.99</td>
</tr>
<tr>
<td>Portfolio NN</td>
<td>0.6977</td>
<td>1.72</td>
<td>3.16</td>
</tr>
<tr>
<td>PNN</td>
<td>0.6034</td>
<td>1.26</td>
<td>2.67</td>
</tr>
<tr>
<td>Regression</td>
<td>0.4761</td>
<td>0.89</td>
<td>2.49</td>
</tr>
<tr>
<td>AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original NN</td>
<td>0.7080</td>
<td>1.65</td>
<td>3.27</td>
</tr>
<tr>
<td>Portfolio NN</td>
<td>0.7205</td>
<td>1.78</td>
<td>3.26</td>
</tr>
<tr>
<td>PNN</td>
<td>0.7409</td>
<td>1.66</td>
<td>3.52</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5228</td>
<td>1.04</td>
<td>3.05</td>
</tr>
<tr>
<td>MC2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>0.6682</td>
<td>1.77</td>
<td>3.60</td>
</tr>
<tr>
<td>Portfolio NN</td>
<td>0.6364</td>
<td>1.82</td>
<td>3.82</td>
</tr>
<tr>
<td>PNN</td>
<td>0.6985</td>
<td>1.77</td>
<td>3.97</td>
</tr>
<tr>
<td>Regression</td>
<td>0.4712</td>
<td>1.02</td>
<td>3.29</td>
</tr>
<tr>
<td>AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original NN</td>
<td>0.7167</td>
<td>1.95</td>
<td>3.62</td>
</tr>
<tr>
<td>Portfolio NN</td>
<td>0.7333</td>
<td>2.07</td>
<td>3.67</td>
</tr>
<tr>
<td>PNN</td>
<td>0.7151</td>
<td>1.86</td>
<td>4.01</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5152</td>
<td>1.17</td>
<td>3.38</td>
</tr>
<tr>
<td>MC3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>0.6636</td>
<td>2.01</td>
<td>4.22</td>
</tr>
<tr>
<td>Portfolio NN</td>
<td>0.7591</td>
<td>2.30</td>
<td>4.25</td>
</tr>
<tr>
<td>PNN</td>
<td>0.6864</td>
<td>2.00</td>
<td>4.73</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5477</td>
<td>1.43</td>
<td>4.28</td>
</tr>
<tr>
<td>AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original NN</td>
<td>0.6887</td>
<td>2.09</td>
<td>4.31</td>
</tr>
<tr>
<td>Portfolio NN</td>
<td>0.7591</td>
<td>2.33</td>
<td>4.36</td>
</tr>
<tr>
<td>PNN</td>
<td>0.6864</td>
<td>2.00</td>
<td>4.73</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5228</td>
<td>1.23</td>
<td>4.05</td>
</tr>
</tbody>
</table>
According to Table 6, the SIGN results indicate that the forecasting models of the AM outperform those of the CM as can be seen from the fact that the regression of the AM produces higher SIGN than that of the CM for MC1 and MC2. Similarly, the SIGNs generated by all neural network models of the AM are always greater than those of the CM for MC1 and MC2. The SIGN results for MC3 indicate that the original NN of the AM performs better than its counterpart, while the portfolio NN and PNN of the AM generate similar results as those of the CM. In particular, the result shows that the portfolio NN of the AM has the highest SIGN for MC2 and MC3, whereas the highest SIGN for MC1 can be achieved by the PNN forecast of the AM. More importantly, this result confirms that forecasting accuracy can be further improved by uncovering the adaptive relevant variables for retraining the forecasting models.

7.1. Trading simulation

A trading simulation was adopted in an effort to further examine if the forecasting models of the AM could practically have been used to generate higher profits than those of the CM. The operational details of the simulated trading are explained as follows: The simulated trading assumes that in the beginning of each monthly period the investor makes an asset allocation decision of whether to shift assets into T-bills or an S&P 500 stock portfolio. It should be noted that the S&P 500 is a value-weighted index in which the index level can be used to perfectly track capital gains on the underlying portfolio if investors were to buy each share in the index in proportion to its outstanding market value. This strategy is applicable for portfolio and fund managers who have the ability to shape a 1-month ahead asset allocation in the equity markets, yet it does not seem possible for small investors. However, investors can purchase shares in mutual index funds that hold shares in proportion to their representation in the S&P 500. Further, it is assumed that the money invested in either T-bills or a stock portfolio becomes illiquid and remains detained in that security until the end of the month.

In the beginning of each month the investor has to decide whether to purchase the S&P 500 portfolio or T-bills, depending on whether the predictions generated by the forecasting models call for a positive or negative excess stock returns in the next month, respectively. The above-mentioned strategies imply full investment in either a stock or T-bill for the whole month. Leveraging or short selling when investing is not allowed in this study, since several factors such as uptick exchange rules, dividends paid during short selling, and margin calls must be considered to reflect a more realistic trading practice. Dividends and transaction costs are also ignored for this study. To summarize, the following describes the simple trading strategies:

If $C_{t+1} = +1$, then
Fully invest in stocks or maintain, and receive the actual stock return for the period $t + 1(R_{t+1})$;
Else (if $C_{t+1} = -1$), then
Fully invest in Treasury bills or maintain, and receive the actual Treasury bill return for the period $t + 1(T1H_t)$.
where $C$ is the sign of excess stock return given by the forecasting models. After performing the trading simulation for evaluating the impact of the AM on stock trading profitability, the resulting mean or monthly return on investment and standard deviation of the monthly return generated from each forecasting model over the second period sets are calculated and presented in the last two columns of Table 6.

The results in Table 6 show that the monthly returns based on trading guided by the original and portfolio NNs of the AM are consistently higher than those of the CM for all three MCs. Similarly, the PNN of the AM greatly outperform that of the CM for MC1 and MC2, although no difference of the monthly returns between the CM and AM was found for MC3. In the same fashion, the monthly returns generated by the regression of the AM are clearly higher than those of the CM for MC1 and MC2. However, it was found that the regression of the AM for MC3 fails to generate higher monthly returns than that of the CM. This result implies that the AM approach may not be effectively utilized by the regression. As a result, based on the predictability and profitability measures, it may be concluded that the interrelationship of the financial and economic variables over stock returns changes over time for a number of reasons. First, the sets of relevant variables selected over the different sliding periods by the variable relevance analysis techniques, including the data mining analysis and the backward stepwise regression elimination, are considerably different as previously mentioned. Second, the results from the study show that the adaptive remodeling efforts help improve the forecasting accuracy in terms of the SIGN. Undoubtedly, this superior forecasting ability would then help generate higher trading profitability than forecasts that are based on constant variables.

7.2. Adaptive modeling performance

Each forecasting model of the AM was examined more closely since the previous analysis concentrated primarily on the overall performance comparisons between the CM and AM forecasts. In fact, neither the forecasting performance nor the trading profitability of the particular forecasting model was statistically evaluated. The results reported in Table 7 presents the detail performance measures of each second period set generated by the forecasting models that were redeveloped based on the adaptive relevant variables. Note that the “overall results” of all second period sets provided in Table 7 are actually the results obtained from the AM for MC1 (see Table 6).

Regarding Table 7, the SIGN results of the redeveloped linear regression model show that it is the worst performer as it disappointingly generates the lowest SIGN in comparison to that of the redeveloped neural network models in all of the second period sets. In fact, the SIGN result generated by the redeveloped neural network models is by far more accurate and consistently predictive than that of the redeveloped linear regression forecast. This is because the correct signs produced by all of the redeveloped neural network models are always greater than or equal to 0.6000 for all second period sets. Particularly, the highest SIGN in two out of the four-second period sets can be obtained by using either the redeveloped portfolio NN or the redeveloped PNN. Particularly, the overall result shows that the redeveloped PNN is the best model that generates the highest average SIGN, while the redeveloped original and portfolio NNs
Table 7
Predictability and profitability results (adaptive modeling) for each second period set

<table>
<thead>
<tr>
<th>Second period set 1: Original NN</th>
<th>SIGN</th>
<th>Monthly return</th>
<th>Std. of return</th>
<th>Sharpe ratio</th>
<th>Equal-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/92–08/94 Portfolio NN</td>
<td>0.6818*</td>
<td>0.80</td>
<td>1.70</td>
<td>0.32</td>
<td>0.64</td>
</tr>
<tr>
<td>PNN</td>
<td>0.8182*</td>
<td>1.10</td>
<td>1.29</td>
<td>0.66</td>
<td>1.05</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5455</td>
<td>0.69</td>
<td>1.76</td>
<td>0.25</td>
<td>0.56</td>
</tr>
<tr>
<td>Buy-and-hold</td>
<td>—</td>
<td>0.61</td>
<td>2.32</td>
<td>0.16</td>
<td>0.44</td>
</tr>
<tr>
<td>RW</td>
<td>—</td>
<td>0.34</td>
<td>1.76</td>
<td>0.05</td>
<td>0.31</td>
</tr>
<tr>
<td>T-bill</td>
<td>—</td>
<td>0.25</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second period set 2: Original NN</th>
<th>SIGN</th>
<th>Monthly return</th>
<th>Std. of return</th>
<th>Sharpe ratio</th>
<th>Equal-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/94–06/96 Portfolio NN</td>
<td>0.6818*</td>
<td>1.57</td>
<td>1.74</td>
<td>0.66</td>
<td>1.42</td>
</tr>
<tr>
<td>PNN</td>
<td>0.7727*</td>
<td>1.60</td>
<td>2.07</td>
<td>0.57</td>
<td>1.28</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5000</td>
<td>1.04</td>
<td>1.51</td>
<td>0.41</td>
<td>1.04</td>
</tr>
<tr>
<td>Buy-and-hold</td>
<td>—</td>
<td>1.60</td>
<td>2.07</td>
<td>0.57</td>
<td>1.28</td>
</tr>
<tr>
<td>RW</td>
<td>—</td>
<td>1.15</td>
<td>1.93</td>
<td>0.38</td>
<td>0.99</td>
</tr>
<tr>
<td>T-bill</td>
<td>—</td>
<td>0.42</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second period set 3: Original NN</th>
<th>SIGN</th>
<th>Monthly return</th>
<th>Std. of return</th>
<th>Sharpe ratio</th>
<th>Equal-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>07/96–04/98 Portfolio NN</td>
<td>0.8182*</td>
<td>2.84</td>
<td>3.60</td>
<td>0.67</td>
<td>2.84</td>
</tr>
<tr>
<td>PNN</td>
<td>0.7727*</td>
<td>2.41</td>
<td>4.19</td>
<td>0.47</td>
<td>2.13</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5455</td>
<td>1.60</td>
<td>3.72</td>
<td>0.32</td>
<td>1.56</td>
</tr>
<tr>
<td>Buy-and-hold</td>
<td>—</td>
<td>2.41</td>
<td>4.19</td>
<td>0.47</td>
<td>2.13</td>
</tr>
<tr>
<td>RW</td>
<td>—</td>
<td>1.65</td>
<td>3.73</td>
<td>0.33</td>
<td>1.61</td>
</tr>
<tr>
<td>T-bill</td>
<td>—</td>
<td>0.42</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second period set 4: Original NN</th>
<th>SIGN</th>
<th>Monthly return</th>
<th>Std. of return</th>
<th>Sharpe ratio</th>
<th>Equal-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/98–12/99 Portfolio NN</td>
<td>0.7000*</td>
<td>1.78</td>
<td>5.10</td>
<td>0.28</td>
<td>1.26</td>
</tr>
<tr>
<td>PNN</td>
<td>0.6000</td>
<td>1.54</td>
<td>5.35</td>
<td>0.22</td>
<td>1.08</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5000</td>
<td>0.82</td>
<td>4.45</td>
<td>0.10</td>
<td>0.70</td>
</tr>
<tr>
<td>Buy-and-hold</td>
<td>—</td>
<td>1.54</td>
<td>5.35</td>
<td>0.22</td>
<td>1.08</td>
</tr>
<tr>
<td>RW</td>
<td>—</td>
<td>1.31</td>
<td>3.23</td>
<td>0.29</td>
<td>1.31</td>
</tr>
<tr>
<td>T-bill</td>
<td>—</td>
<td>0.37</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall results: Original NN</th>
<th>SIGN</th>
<th>Monthly return</th>
<th>Std. of return</th>
<th>Sharpe ratio</th>
<th>Equal-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/92–12/99 Portfolio NN</td>
<td>0.7205</td>
<td>1.78</td>
<td>3.26</td>
<td>0.43</td>
<td>1.56</td>
</tr>
<tr>
<td>PNN</td>
<td>0.7409</td>
<td>1.66</td>
<td>3.52</td>
<td>0.37</td>
<td>1.39</td>
</tr>
<tr>
<td>Regression</td>
<td>0.5228</td>
<td>1.04</td>
<td>3.05</td>
<td>0.22</td>
<td>0.98</td>
</tr>
<tr>
<td>Buy-and-hold</td>
<td>—</td>
<td>1.54</td>
<td>3.68</td>
<td>0.32</td>
<td>1.25</td>
</tr>
<tr>
<td>RW</td>
<td>—</td>
<td>1.11</td>
<td>2.77</td>
<td>0.27</td>
<td>1.11</td>
</tr>
<tr>
<td>T-bill</td>
<td>—</td>
<td>0.37</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

give slightly lower SIGN results than that of the redeveloped PNN. For statistical evaluation, the null hypothesis of no predictive effectiveness was calculated by conducting a one-sided test of $H_0: p = 0.50$ against $H_a: p > 0.50$. The SIGN of each second period set marked with an asterisk (*) in Table 7 indicates that they are significantly different from the benchmark of 0.5 at a 95% level of confidence. This verifies that the correct sign generated by most of the redeveloped neural network models is better than
random. This result also implies that the classical linear regression cannot be used to accurately account for directional forecasts of stock return.

Simulated trading was also performed to evaluate the impact of the AM on trading profitability guided by the forecasting models redeveloped in this study. The trading results of the redeveloped forecasting models, including the monthly return on investment, standard deviation, and Sharpe ratio, are reported in Table 7. The monthly returns obtained from trading using the random walk model (RW) and from always investing in a stock portfolio (buy-and-hold) or a short term T-bill (see $T1H$ in the appendix) were also provided as the benchmarks for profitability comparisons in the study. Note that the sign of excess stock return of the current month was simply taken as the forecast of the next month for the RW. The trading results show that all redeveloped forecasting models in the second period set 1 generate higher monthly returns than that of the buy-and-hold account. Specifically, it can be observed that the redeveloped original and portfolio NNs have obtained higher monthly returns than that of the buy-and-hold account in three out of the four second period sets. Noticeably, the redeveloped PNN has gained the same resulting monthly return as that of the buy-and-hold account in the second period sets 2, 3, and 4. Accordingly, the redeveloped PNN was examined more closely to understand why no monthly return difference exists in these second period sets. It was found that the redeveloped PNN always signals the positive signs of excess stock return in the last three second period sets. It may be due to the fact that these sliding periods contained more positive outputs of excess stock return. In fact, the S&P 500 Index during the study period represented an episode of significant rise in stock prices (bull market). Therefore, the smoothing parameter selected in this study may be partly responsible for the similar monthly returns produced from the PNN predictions.

In addition, the trading result shows that the monthly return based on trading guided by the redeveloped regression forecast is 0.56%, 0.81%, and 0.72% less than that of the buy-and-hold account, and is 0.11%, 0.05%, and 0.49% less than that of the RW in the second period sets 2, 3, and 4, respectively. This result suggests that the classical linear regression cannot be used as an efficient forecasting tool to generate profitable trading. Finally, the overall trading results show that the redeveloped portfolio NN generates the highest monthly return (1.78%), and is 0.24% greater than that of the buy-and-hold account (1.54%). In other words, the different of 0.24% is approximately equal to an annualized 2.88% of extra return as compared to that gained from investing in the buy-and-hold account. Similarly, the higher monthly return over the buy-and-hold account can also be obtained from trading driven by the redeveloped PNN and the redeveloped original NN. In fact, it is found that the monthly return (1.66%) guided by the redeveloped PNN is slightly better than the monthly return (1.65%) obtained by the redeveloped original NN. On the contrary, the monthly return (1.04%) generated by the redeveloped regression is not only the lowest but is also 0.50% less than that of the buy-and-hold account. A correlation between the SIGN and the monthly return of the four forecasting models for all second period sets was also calculated ($\rho_{\text{SIGN}&\text{Return}} = 0.5672$) to examine whether the SIGN is strongly related to trading profits. In fact, this resulting correlation means that the return obtained from trading is positively correlated to the SIGN with a fairly strong relationship.
Interestingly, it can be observed that the return obtained from the buy-and-hold account is approximately equal to an annualized 18.46% over the out-of-sample trading periods. This may be the reason why several redeveloped forecasting models could not significantly achieve performance better than that of the buy-and-hold account during this period. Fig. 3 illustrates samples of key cumulative investment returns from 11/1992 to 12/1999 guided by the redeveloped portfolio NN and the redeveloped regression as compared against the buy-and-hold and T-bill accounts. As can be observed from the figure, the cumulative investment return of the redeveloped regression model is slightly higher than that of the buy-and-hold account in the early period of the forecasting months (from 11/1992 to 09/1995). Perceptibly, the cumulative investment return of the redeveloped regression model keeps increasing at a declining rate as compared to that of the buy-and-hold account in the later period of the forecasting months (after 09/1995). Perhaps this result suggests that the redeveloped regression still has no ability to generate consistent profits even though it has been linearly remodeled four times for the out-of-sample forecasting periods (four second period sets). Nevertheless, the cumulative investment return guided by these redeveloped forecasting models is far better than that of the risk-free T-bill account.

A paired two-sample t-test for mean return differences of the whole out-of-sample trading periods (86 months) was also performed. The resulting t-statistics of the redeveloped original NN, portfolio NN, and PNN models in comparison to the redeveloped regression are 2.970, 3.260, and 2.738, respectively (t-critical = 1.988 and α = 0.05). The results indicate that all of the redeveloped neural network models significantly outperform the redeveloped linear regression forecast at a 95% level of confidence. The two-sample test was additionally performed between the redeveloped portfolio NN and the buy-and-hold account. It was found that the mean return improvement of the redeveloped portfolio NN model over the buy-and-hold account is statistically significant at a 85% level of confidence (t-statistic = 1.478, t-critical = 1.453, and α = 0.15). The Sharpe ratio analysis was also included in the study. The Sharpe ratio is simply the mean excess return of the trading divided by its standard deviation. The higher the
Sharpe ratio, the higher the return and the lower the volatility. It is observed that all of the neural network models redeveloped in the study generate a Sharpe ratio performance that is either equal to or higher than the buy-and-hold account. In contrast, the redeveloped regression can only outperform the buy-and-hold account in the second period set 1 for this Sharpe ratio performance. More importantly, the redeveloped portfolio NN, which yields a monthly return of 1.78% over the trading periods, is once again the best performer in terms of the Sharpe ratio performance (0.43) among the redeveloped forecasting models evaluated in the study.

For comparability, equivalent-variance portfolios (equal-variance), using a combination of the return based on trading from each redeveloped forecasting tool and the risk-free asset, were created to evaluate the investment returns under the same volatility. For instance, the redeveloped regression model has the lowest standard deviation for the second period set 2; therefore, the standard deviations of the other models in the second period set 2 were reduced to that of the redeveloped regression by combining them with the existing T-bill account. The results show that all of the redeveloped neural network models perform better than or equal to the buy-and-hold account under this risk-adjusted return calculation for all of the second period sets. Again, the redeveloped regression only does well in the second period set 1 under this comparison, which is similar to the results of the Shape ratio. Remarkably, the monthly return based on trading guided by the redeveloped portfolio NN model is 0.31% greater than that of the buy-and-hold strategy under the same investment uncertainty. In other words, the difference of 0.31% is approximately equal to an annualized 3.72% of investment return under the identical risk exposure.

8. Discussion

In this study an attempt to uncover the structural relationship of various financial and economic variables using data mining analysis and neural networks has been made on the S&P 500 stock portfolio. This approach seems particularly attractive in selecting the variables when the usefulness of the data is unknown, especially when non-linearity exists in the financial market. In particular, our results indicate that the interrelationship of the financial and economic variables over stock returns changes over time and that predicting stock directions based on past publicly available information is possible. In fact, this finding suggests the importance of frequent neural network remodeling to uncover predictive powers of recent relevant variables because of the potentially higher trading profitability that may be obtained from the most up-to-date forecasting model.

One interesting observation about the collected variables used in the study is that the certificates of deposit (CD1, CD3, and CD6) normally neglected by many studies have actually been selected by both the data mining and backward stepwise regression analysis techniques as strong relevant variables in nearly all of the sliding periods. This result implies that there may be additional variables containing predictive powers to explain current patterns of stock behavior that have not been practically tested in the previous finance and economic literature. Particularly, we examined the effectiveness
of the redeveloped neural network models that use the adaptive relevant variables for predicting the directions of excess stock return. The results show that these neural network models, especially the redeveloped portfolio NNs that use the adaptive relevant variables, generate higher profits with lower risks than the buy-and-hold strategy, the conventional linear regression, and the single neural network models that use constant relevant variables.

In particular, the results also show that higher correct signs may not always imply higher profitability. For instance, the redeveloped PNN having the correct sign that is higher than the redeveloped original NN in the second period set 3 (Table 7) does not outperform the redeveloped original NN in terms of the profits obtained from trading. This suggests that the forecast that has a higher percentage of correct sign may not necessarily yield higher profit. In fact, it may be due to the fact that the redeveloped original NN gives better prediction of signs when the actual monthly stock return is highly volatile; thus receiving higher trading profits. This observation suggests the importance of making an accurate asset allocation (between stock and T-bill) when the positive or negative actual stock return of the next month is significant. Therefore, potentially higher investment return may be obtained from training the networks to correctly predict the signs of trading only when significant profit opportunities exist. This may also explain why the redeveloped PNN in the sliding period 3, which is unable to predict the negative signs of excess stock return, could not achieve higher profitability than that of the redeveloped original NN.

This empirical result shows that the trading results based on the neural network forecasts can arrive at higher profitability improvement than the buy-and-hold strategy. However, this does not mean that buy-and-hold strategy can be totally ignored since holding the buy-and-hold account can also be very profitable. In fact, the profitability obtained from the neural network forecasts will likely be less if transaction costs are taken into consideration. In addition, feed-forward neural network training is usually not very stable since the training process may depend on the choice of a random start. Training is also computationally expensive in terms of training times used to determine the appropriate network structure. The degree of success, therefore, may fluctuate from one training pass to another. Although the portfolio neural networks yield impressive profits, it should raise concern that higher profits are derived at the expense of exposing the investors to higher computational risk.

Nonetheless, the empirical findings in this study show that our proposed development of the portfolio network models using the \( n \)-fold cross-validation and early stopping techniques does not sacrifice any of the first period data set used for training and validating the networks. This is especially useful when data size is limited. In particular, we find that the method for improving the generalization ability of feed-forward neural networks, a combination of \( n \)-fold cross-validation and early stopping techniques, obviously help improve the out-of-sample forecasts. In addition to early stopping, an advantage may be gained from the five-time network modeling which allows the networks to extract more useful information from the data. Thus, the prediction based on the majority of excess return sign could reasonably be used to reduce the prediction error. Evidently, the results found in this study suggest that the portfolio neural networks that direct the trading based on the majority of
the five network outputs could be developed and used as a more efficient forecasting tool.

9. Conclusion

Stock return and stock market predictions have been studied by both academics and practitioners for many years. Many studies conclude that stock returns can be predicted by some financial and economic variables. To this end, our finding suggests that stock market predictions are feasible but will remain difficult because they are also influenced by other factors, such as political, international, and even natural events. Obviously, this study covers only the analysis of fundamental available information, while the technical analysis technique remains intact. Far from perfect, technical analysis nonetheless has been shown to provide another aspect for stock price and stock return forecasting. Technical analysis to some extent has also been known to offer a relative mixture of human, political, and economical events. If both technical and fundamental approaches are thoroughly examined and included during the variable relevance analysis modeling, it would no doubt be a major improvement in predicting stock returns.

Finally, this study assumes trading strategies of investing in either the stock index portfolio or risk-free account in the absence of trading costs. During the simulated trading exercise we notice that the profitability results may change if different trading strategies are adopted by investors. In fact, it is possible that investors would benefit from further investigation on profits received from different trading strategies. Also, future research should consider the trading simulation under the scenarios of stock dividends, transaction costs, and individual-tax brackets to replicate realistic investment practices.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments and the Intelligent System Center at the University of Missouri-Rolla for financial support.

Appendix

\[ SP \] Nominal Standard & Poor’s 500 index at the close of the last trading day of each month. Source: Commodity Systems, Inc. (CSI)

\[ DIV \] Nominal dividends per share for the S&P 500 portfolio paid during the month. Source: Annual dividend record/Standard and Poor’s Corporation
Annualized average of bid and ask yields on one-month T-bill rate on the last trading day of the month. It refers to the shortest maturity T-bills not less than one month in maturity. Source: CRSP tapes, the Fama risk free rate file.

Monthly holding period return on one-month T-bill rate on the last trading day of the month, calculated as $T1/12$.

Nominal stock returns on the S&P 500 portfolio, calculated as $R_t = (SP_t - SP_{t-1})/SP_{t-1}$.

Excess stock returns on the S&P 500 portfolio, calculated as $ER_t = R_t - T1H_{t-1}$.

Dividend yield on the S&P 500 portfolio, calculated as $DY_t = DIV_t/SP_t$.

3-month T-bill rate, secondary market, averages of business days, discount basis. Source: H.15 Release—Federal Reserve Board of Governors

6-month T-bill rate, secondary market, averages of business days, discount basis. Source: H.15 Release—Federal Reserve Board of Governors

1-year T-bill rate, secondary market, averages of business days, discount basis. Source: H.15 Release—Federal Reserve Board of Governors

5-year T-bill constant maturity rate, secondary market, averages of business days. Source: H.15 Release—Federal Reserve Board of Governors

10-year T-bill constant maturity rate, secondary market, averages of business days. Source: H.15 Release—Federal Reserve Board of Governors

1-month certificate of deposit rate, averages of business days. Source: H.15 Release—Federal Reserve Board of Governors

3-month certificate of deposit rate, averages of business days. Source: H.15 Release—Federal Reserve Board of Governors

6-month certificate of deposit rate, averages of business days. Source: H.15 Release—Federal Reserve Board of Governors

Moody’s seasoned Aaa corporate bond yield, averages of business days. Source: The Federal Reserve Bank of St. Louis

Moody’s seasoned Baa corporate bond yield, averages of business days. Source: The Federal Reserve Bank of St. Louis


M1 Money Stock. Source: H.6 Release—Federal Reserve Board of Governors

Term spread between $T120$ and $T1$, calculated as $TE1 = T120 - T1$.

Term spread between $T120$ and $T3$, calculated as $TE2 = T120 - T3$.

Term spread between $T120$ and $T6$, calculated as $TE3 = T120 - T6$.

Term spread between $T120$ and $T12$, calculated as $TE4 = T120 - T12$.

Term spread between $T3$ and $T1$, calculated as $TE5 = T3 - T1$.

Term spread between $T6$ and $T1$, calculated as $TE6 = T6 - T1$.

Term spread between $T120$ and $T1$, calculated as $TE1 = T120 - T1$.
DE1  Default spread between BAA and AAA, calculated as $DE_1 = BAA - AAA$
DE2  Default spread between BAA and T120, calculated as $DE_2 = BAA - T120$
DE3  Default spread between BAA and T12, calculated as $DE_3 = BAA - T12$
DE4  Default spread between BAA and T6, calculated as $DE_4 = BAA - T6$
DE5  Default spread between BAA and T3, calculated as $DE_5 = BAA - T3$
DE6  Default spread between BAA and T1, calculated as $DE_6 = BAA - T1$
DE7  Default spread between CD6 and T6, calculated as $DE_7 = CD6 - T6$

References

Suraphan Thawornwong received his B.E. in Mechanical Engineering from Prince of Songkla University in 1995, and his M.B.A. and M.S. in Industrial Engineering from the University of Missouri-Columbia in 1999 and 2000, respectively. He received his Ph.D. degree in Engineering Management at the University of Missouri-Rolla in 2003. His research interests include applications of artificial intelligence, particularly neural networks, genetic algorithms, and expert systems, for financial forecasting and business decision-making.

David Enke received his B.S. in Electrical Engineering in 1990, and his M.S. and Ph.D. in Engineering Management in 1994 and 1997, respectively, each from the University of Missouri-Rolla (UMR). He was an assistant professor of System Science and Industrial Engineering at the State University of New York-Binghamton before returning to UMR as an assistant professor within the Engineering Management Department. His research interests involve the development of smart and intelligent systems, specifically using neural networks, knowledge-based systems, and data mining techniques in the areas of financial forecasting, financial engineering, investment, capital planning and budgeting, electrical load and price forecasting, and artificial vision.