

ORIGINAL RESEARCH

Open Access

Forecasting S&P 500 index using artificial neural networks and design of experiments

Seyed Taghi Akhavan Niaki* and Saeid Hoseinzade

Abstract

The main objective of this research is to forecast the daily direction of Standard & Poor's 500 (S&P 500) index using an artificial neural network (ANN). In order to select the most influential features (factors) of the proposed ANN that affect the daily direction of S&P 500 (the response), design of experiments are conducted to determine the statistically significant factors among 27 potential financial and economical variables along with a feature defined as the number of nodes of the ANN. The results of employing the proposed methodology show that the ANN that uses the most influential features is able to forecast the daily direction of S&P 500 significantly better than the traditional logit model. Furthermore, experimental results of employing the proposed ANN on the trades in a test period indicate that ANN could significantly improve the trading profit as compared with the buy-and-hold strategy.

Keywords: S&P 500 index, Financial time series, Artificial neural networks, Design of experiments, Analysis of variance, Logit model

Background

Due to the great volume of trading in stock markets, significant profit can be made by improving trading performance using adequate forecasting of financial variables such as stock prices, stock market indices, and prices of financial derivatives; that is why several research works in different fields of study have been performed on this subject so far. Atsalakis and Valavanis (2009) and Vanstone and Finnie (2009) comprehensively surveyed these works and their involved methodologies. Among other financial variables, stock market indices have received significant attention, and many researchers such as Atsalakis and Valavanis (2009), Vanstone and Finnie (2009), and Leung et al. (2000) proposed different methodologies to forecast them.

The efficient market hypothesis states that all of the available information in a market is reflected in the price of a stock at any moment, and therefore, forecasting the future price of a stock using the historical data and consequently gaining a return above the market average is impossible (Peters 1991). However, many empirical research works questioned this hypothesis and claimed

that markets often do not follow the efficient market hypothesis completely. Jensen (1978) stated that because of the psychological factors along with varying groups of human decision makers, markets do not react immediately to the newly released information and that this phenomenon makes the markets inefficient. Additionally, Ferson (1989), Fama and French (1989), and Fama and Schwert (1977) showed that financial variables could be forecasted using time series on some other financial and economical variables.

Artificial neural networks (ANNs) are one of the relatively newly developed methods that have been widely used for forecasting purposes in different fields. Zhang et al. (1998) made a comprehensive review of using ANNs as a forecasting tool.

One of the fields of study in which the ANN application received a significant attention is finance. Trippi and Turban (1996) reviewed the application of ANNs in different areas in finance and investment. One of the areas of finance, in which using ANNs gained attention, is forecasting financial time series. Financial time series have some characteristics that make them hard to forecast especially when a traditional statistical method is employed (see for example Eakins and Stansell 2003; Hussain et al. 2007; Lam 2004; Lin et al. 2006; Motiwalla

* Correspondence: Niaki@Sharif.edu
Department of Industrial Engineering, Sharif University of Technology, P.O.
Box 11155-9414 Azaadi Ave, Tehran 145889694, Iran

and Wahab 2000; Thawornwong and Enke 2004; Versace et al. 2004; Yao et al. 1992). These characteristics are as follows:

1. Non-stationarity. It means that because of the different business and economic cycles, the statistical properties of financial data change over time.
2. Nonlinearity. It means that the relationship between the financial and economical independent variables and the desired dependent variable is not linear.
3. Noisiness. It means that there are day-to-day variations in financial time series.

There are two main approaches of univariate and multivariate analyses in forecasting financial time series where ANNs can be used in both (Cao and Tay 2001). In univariate analyses, the features are limited to the variable being forecasted, while in multivariate case, any variable that is thought to be related to the output is considered a feature (an input of the model). The following points justify the use of ANNs to forecast financial time series (Eakins and Stansell 2003; Hornik et al. 1989; Hussain et al. 2007; Lam 2004):

1. ANNs are nonlinear. It means that they can capture nonlinear relations between feature (input or independent) and response (output or dependent) variables.
2. ANNs are data driven. It means that they do not need any explicit assumption on the model between inputs and outputs.
3. ANNs can generalize. It means that after training, they can produce good results even when they face new input patterns.
4. Unlike statistical techniques, ANNs do not need assumptions on the distribution of input data.

Although ANNs have the above-mentioned advantages, the robustness of their outcomes has been questioned (Saad et al. 1998). Besides, the ANNs' limitations are as follows (Hussain et al. 2007; Lam 2004):

1. Determining the optimal combination of the network parameters such as learning rate, momentum, number of hidden layers, and number of hidden nodes in each layer is difficult.
2. Selecting the relevant features of an ANN is not an easy job.
3. Great volume of data is required to train the network to achieve an accurate result.

Designing an optimal ANN with only relevant features is essential in obtaining good forecasting results (Zhu et al. 2008). In fact, using irrelevant features along with

not using relevant ones can reduce the performance of an ANN (Thawornwong and Enke 2004). For example, Cao et al. (2005) concluded that extending the model from CAPM to that of the 3-factor model of Fama and French (1989) negatively affects the forecast accuracy in the Shanghai stock exchange. Hence, the key factor to success in forecasting the financial time series is designing an ANN that has the least complexity with only relevant and most influential features (Atsalakis and Valavanis 2009).

There are several techniques in the literature to select relevant features of ANNs. These techniques can be classified into two main categories. The first category uses variable relevance analysis to examine whether a variable has useful information to forecast the output variable. If the answer to this question is yes, then the variable will be used as an input for the ANN. Thawornwong and Enke (2004) used this method to select features of an ANN. The disadvantage of the methods in this category is that even if a variable has some useful information, ANNs may not be able to extract that information and use it to capture the relationship between that variable and the output.

The techniques of the second category compare the performance of different ANNs that contain different features. Zhu et al. (2008), Cao et al. (2005), and O'Connor and Madden (2006) used one of these techniques. The advantage of the techniques of this category is that they not only select relevant features, but they also propose a method to extract information on the features to forecast the output. That is why the techniques of this category are more common in the literature. However, there are several drawbacks in the use of the techniques of this category in previous studies. The main problem is that they lack a scientific design of experiments, which leads to robust results that can be statistically trusted. In other words, most of the available studies did not undergo statistical analysis to prove that their designed ANN would significantly improve the results (Lam 2004).

The main objective of this research is to forecast the daily direction of Standard & Poor's (S&P 500) index using an ANN that uses the most influential financial and economical features. We are also going to use tests of hypothesis to prove statistically that the proposed ANN significantly improves the forecasting ability of the S&P 500 index.

The remainder of this paper is organized as follows. 'The propose ANN' section describes the developed ANN and the designed experiments that will be used to select the most influential features. In the 'Conducting the experiments' section, the experiments are conducted and the results are analyzed. In the 'Examining the financial performances of the proposed ANN' section, the hypothesis that using the forecasts produced by the

proposed ANN significantly increases the profit is tested. The 'Forecasting S&P 500 daily direction using regression' section contains the comparison results of employing both the proposed ANN and the traditional logit model approach. Finally, the research is concluded in the 'Conclusions and further research' section.

The proposed ANN

ANNs are flexible computing methods that have the ability to capture the patterns among variables. ANNs have some characteristics (mentioned in the 'Background' section) that make them reasonable to be used for a wide range of problems.

In order to design an appropriate ANN for a particular problem, one should decide on the network topology, number of network layers, number of nodes in each layer, activation function of the nodes, and finally, the learning algorithm.

Based on the topology, ANNs are mainly divided into two groups of feed-forward and recurrent networks. For univariate forecasting analysis, the use of recurrent topology is more common than the feed-forward networks. Hussain et al. (2007), Lin et al. (2006), and Saad et al. (1998) are the ones who employed recurrent topology. However, in the field of financial forecasting, especially in multivariate forecasting analysis, the feed-forward topology has gained much more attention. As a result, the latter topology will be used in this research as well.

The number of network layers depends on the complexity of the problem being modeled. In addition to the input and the output layers that are essential for an ANN design, many feed-forward networks have one or more hidden layers. Several researchers have proposed different methods to determine the optimal number of hidden layers and hidden nodes. However, these methods are very complex and hard to apply (Zhang et al. 1998). Moreover, none of the existing methods can guarantee that the obtained ANNs are optimal. Therefore, the common practice of identifying the proper network design, which is comparing the performances of ANNs with different designs and selecting the network that results in the best performance, is taken in this research (Hosseini et al. 2006). In addition, while different researchers in the literature recommended not using more than two hidden layers, in this research the optimal number of hidden nodes in only one hidden layer is determined using design of experiments (DOE). In other words, the number of hidden nodes is considered a potential influential factor in DOE that affect the forecasting ability of the proposed ANN. The levels of this factor will be discussed later in this section.

The input layer of an ANN consists of the input variables (features) that seem to be influential to the output variable. These influential features are determined using DOE as well. Again, the features of the proposed ANN are the

potential factors of DOE. These features along with their levels will be discussed later in this section.

The output layer of an ANN consists of nodes associated with the dependent variables. Since the objective of this research is to forecast the daily direction of S&P 500 index properly and since the direction is either positive or negative, the output layer of the proposed ANN consists of only one node.

As the activation function, the tangent hyperbolic sigmoid (Tansig) function, which is the most common one in the relevant literature, is used for the nodes of all layers. Furthermore, the error back-propagation algorithm is employed to train the designed ANN.

In order to design, train, and simulate the proposed ANN, the neural network toolbox of the MATLAB 7 package software (Mathworks 2004) is used in this research. For a detailed description of the neural network toolbox, the reader is referred to the work of Demuth and Beale (1998) who demonstrated the application of ANNs in this toolbox comprehensively.

Data compilation

The dataset of this research has been compiled containing the daily closing value of S&P 500 as the output and 27 financial and economical variables as the potential features for 3,650 days partially taken on the period starting 01 March 1994 to 30 June 2008. The first 80% of the compiled data (for 2,920 days) is used for the training purpose, the next 10% is used for verification, and finally, the last 10% is used to test the performance of the designed ANN. The verification and the test periods consist of the data for 365 days each.

The main consideration for selecting the potential features is whether they have significant influence on the output variable (the direction of S&P 500) in the next day. While some of the features considered in this study were used in previous studies, some of them are new and are used in this study for the first time. Analyzing the results of the designed experiments will indicate which of these features have a meaningful impact on the response. The list, the description, and the sources of the potential features are given in Table 1.

Selecting the design and choosing the factor levels

In order to determine the relevant and the most influential variables of the forecasting process, a factorial design is employed in this research. The factors include the 27 features given in the Table 1, each with two levels (0 and 1). Level 1 of a factor indicates that the corresponding factor is used as a feature, and level 0 represents the nonexistence of the factor. In addition to the 27 features, the number of hidden nodes is also considered as a potential factor in designing the experiments. Note that experiments with fewer input

Table 1 The 27 potential influential features

| | | |
|----|---------|--|
| 1 | SPYt-1 | The return of the S&P 500 index in day t-1. Source: finance.yahoo.com |
| 2 | SPYt-2 | The return of the S&P 500 index in day t-2. Source: finance.yahoo.com |
| 3 | SPYt-3 | The return of the S&P 500 index in day t-3. Source: finance.yahoo.com |
| 4 | Oil | Relative change in the price of the crude oil (Cushing, OK WTI Spot Price FOB (dollars per barrel)). Source: Energy Information Administration, available at http://tonto.eia.doe.gov/dnav/pet/hist/rwtcd.htm |
| 5 | Gold | Relative change in the gold price. Source: www.usagold.com |
| 6 | CTB3M | Change in the market yield on US Treasury securities at 3-month constant maturity, quoted on investment basis. Source: H.15 Release - Federal Reserve Board of Governors |
| 7 | CTB6M | Change in the market yield on US Treasury securities at 6-month constant maturity, quoted on investment basis. Source: H.15 Release - Federal Reserve Board of Governors |
| 8 | CTB1Y | Change in the market yield on US Treasury securities at 1-year constant maturity, quoted on investment basis. Source: H.15 Release - Federal Reserve Board of Governors |
| 9 | CTB5Y | Change in the market yield on US Treasury securities at 5-year constant maturity, quoted on investment basis. Source: H.15 Release - Federal Reserve Board of Governors |
| 10 | CTB10Y | Change in the market yield on US Treasury securities at 10-year constant maturity, quoted on investment basis. Source: H.15 Release - Federal Reserve Board of Governors |
| 11 | AAA | Change in the Moody's yield on seasoned corporate bonds - all industries, Aaa. Source: H.15 Release - Federal Reserve Board of Governors |
| 12 | BBB | Change in the Moody's yield on seasoned corporate bonds - all industries, Baa. Source: H.15 Release - Federal Reserve Board of Governors |
| 13 | USD-Y | Relative change in the exchange rate between US dollar and Japanese yen. Source: OANDA.com |
| 14 | USD-GBP | Relative change in the exchange rate between US dollar and British pound. Source: OANDA.com |
| 15 | USD-CAD | Relative change in the exchange rate between US dollar and Canadian dollar. Source: OANDA.com |
| 16 | HIS | Hang Seng index return in day t-1. Source: finance.yahoo.com |
| 17 | FCHI | CAC 40 index return in day t-1. Source: finance.yahoo.com |
| 18 | FTSE | FTSE 100 index return in day t-1. Source: finance.yahoo.com |
| 19 | GDAXI | DAX index return in day t-1. Source: finance.yahoo.com |
| 20 | V | Relative change in the trading volume of S&P 500 index. Source: finance.yahoo.com |
| 21 | XOM | Exxon Mobil stock return in day t-1. Source: finance.yahoo.com |

Table 1 The 27 potential influential features (Continued)

| | | |
|----|------|---|
| 22 | GE | General Electric stock return in day t-1. Source: finance.yahoo.com |
| 23 | MSFT | Micro Soft stock return in day t-1. Source: finance.yahoo.com |
| 24 | PG | Procter and Gamble stock return in day t-1. Source: finance.yahoo.com |
| 25 | JNJ | Johnson and Johnson stock return in day t-1. Source: finance.yahoo.com |
| 26 | DJI | Dow Jones Industrial Average index return in day t-1. Source: finance.yahoo.com |
| 27 | IXIC | NASDAQ composite index return in day t-1. Source: finance.yahoo.com |

factors need fewer hidden nodes and vice versa. In other words, the number of hidden nodes of the proposed ANN for experiments on a large number of features should be high enough to have the ability of capturing the patterns. Therefore, the number of hidden nodes is not constant in all experiments and is considered a factor in the experiments. Considering the number of other features, four levels of the number of hidden nodes including 5, 15, 30, and 60 are considered in this research. As a result, the total number of combinations is 2^{29} (four for the number of hidden nodes and 2^{27} for the other features). We also note that only one replicate of this design needs 2^{29} experiments, which is impractical. To make the total number of experiments practical, two approaches are used.

In the first approach, in each pair of the features that are highly correlated, only one is kept and the other is excluded. This is based on the fact that if two features are highly correlated, then both are inferred to contain similar information. Hence, using one of them as the input is almost sufficient for an ANN to extract the desired pattern. By 'highly correlated,' we mean that the absolute value of the correlation is at least 0.5. The application of this approach leads to exclude 7 of the 27 initial financial and economical variables. The omitted variables of this approach are DJI, IXIC, CTB6M, CTB1Y, CTB5Y, CTB10Y, and BAA.

The second approach of decreasing the number of features is based on the concept of grouping. In this approach, the remaining features of the first approach with almost similar nature are placed in one group, making them group factors. In this way, the existence and nonexistence of the group factors matters. Zhu et al. (2008) and O'Connor and Madden (2006) used a similar technique for comparing the performance of different ANNs. Employing the second approach makes the remaining 20 features as the six group factors, as shown in Table 2.

Conducting the experiments

Based on the derivations given in the 'Selecting the design and choosing the factor levels' subsection, there are now six group factors, each with two levels along with

Table 2 Grouping features

| Group number | Group description | Group title | Included input variables |
|--------------|--|-------------|--------------------------------|
| 1 | S&P 500 index return in three previous days | G1 | SPYt-1 SPYt-2 SPYt-3 |
| 2 | Financial and economical indicators | G2 | Oil Gold CTB3M AAA |
| 3 | The return of the five biggest companies in S&P 500 | G3 | XOM GE MSFT PG JNJ |
| 4 | Exchange rate between USD and three other currencies | G4 | USD-Y USD-GBP USD-CAD |
| 5 | The return of the four world major indices | G5 | HIS FCHI FTSE GDAXI |
| 6 | S&P 500 trading volume | G6 | V |

the number of hidden nodes with four levels. This decreases the total number of the factor level combination to a practical value of 2^8 . However, four of these combinations are not possible. Since one cannot have an ANN without input, the combinations in which the level of all features is zero are not feasible. The number of these infeasible combinations is four. Consequently, the number of possible combinations would be 252. Each combination corresponds to a specific ANN; each observation for a combination uses the specified ANN to produce the response variable, and a total of 10 replicates are considered for each combination, leading to the total number of 2,520 experiments. Experiments are conducted in a way that for each observation of each possible combination, the corresponding ANN is initialized with a random value. The initialized ANN is then trained with the training dataset until the early stopping^a rule terminates the training process. Next, the trained network is simulated with the test dataset, and the forecasts that are generated by the ANN in the test period are gathered. Since the test period consists of 365 days, each ANN corresponding to an experiment forecasts 365 daily directions of S&P 500 in the test period. The response variable is the number of correct forecasts that the ANN produces in the test period.

Determining the influential factors

After conducting the experiments and obtaining the results, the analysis of variance (ANOVA) approach is used to determine the factors that have statistically significant effects on the response variable. The general linear model procedure of the SAS software package (SAS Institute Inc. 2009) is employed for this purpose. We assume that only the main effects and the second-order mutual (interaction) effects exist, and mutual effects of higher orders can be neglected.

The underlying assumptions of the ANOVA approach are that the error terms are independently and normally distributed with mean zero and a constant variance. By a residual analysis (not shown here), we concluded that the assumptions hold, and hence, the results of ANOVA are valid. The results are summarized in Table 3.

The results in Table 3 show that four main effects, including $G1$, $G4$, $G6$, and hidden nodes, and one mutual effect, $G2 \times G3$, are statistically significant at 5% significance level. This means that the existence or nonexistence of the other factors including $G1$ and $G5$ does not meaningfully influence the response variable.

In order to determine the levels of the selected factors on which the response has the best value, Duncan's multiple range test is used. Table 4 presents the optimal level of the influential factors.

Table 3 The ANOVA results

| Source | Type III sum of squares | df | Mean square | F |
|-------------------|-------------------------|-------|-------------|---------|
| G1 | 1,070.88 | 1 | 1,070.88 | 30.26* |
| G2 | 121.07 | 1 | 121.07 | 3.42 |
| G3 | 23.34 | 1 | 23.34 | 0.66 |
| G4 | 5,057.11 | 1 | 5,057.11 | 142.91* |
| G5 | 109.34 | 1 | 109.34 | 3.09 |
| G6 | 258.63 | 1 | 258.63 | 7.31* |
| Hidden nodes | 1,844.41 | 3 | 614.8 | 17.37* |
| G1 × G2 | 117.66 | 1 | 117.66 | 3.33 |
| G1 × G3 | 22.83 | 1 | 22.83 | 0.65 |
| G1 × G4 | 29.77 | 1 | 29.77 | 0.84 |
| G1 × G5 | 30.62 | 1 | 30.62 | 0.87 |
| G1 × G6 | 0.51 | 1 | 0.51 | 0.01 |
| G1 × hidden nodes | 151.75 | 3 | 50.58 | 1.43 |
| G2 × G3 | 154.65 | 1 | 154.65 | 4.37* |
| G2 × G4 | 6.98 | 1 | 6.98 | 0.2 |
| G2 × G5 | 7.98 | 1 | 7.98 | 0.23 |
| G2 × G6 | 9.84 | 1 | 9.84 | 0.28 |
| G2 × hidden nodes | 72.87 | 3 | 24.29 | 0.69 |
| G3 × G4 | 0.59 | 1 | 0.59 | 0.02 |
| G3 × G5 | 0.63 | 1 | 0.63 | 0.02 |
| G3 × G6 | 130.42 | 1 | 130.42 | 3.69 |
| G3 × hidden nodes | 48.41 | 3 | 16.14 | 0.46 |
| G4 × G5 | 92.6 | 1 | 92.6 | 2.62 |
| G4 × G6 | 23.59 | 1 | 23.59 | 0.67 |
| G4 × hidden nodes | 257.08 | 3 | 85.69 | 2.42 |
| G5 × G6 | 16.4 | 1 | 16.4 | 0.46 |
| G5 × hidden nodes | 239.54 | 3 | 79.85 | 2.26 |
| G6 × hidden nodes | 121.54 | 3 | 40.51 | 1.14 |
| Error | 87,650.55 | 2,477 | 35.39 | |
| Total | 97,862.62 | 2,519 | | |

*p value < 0.05.

Table 4 Influential factors and their optimal level

| Significant factors | Optimal level | Description |
|---------------------|------------------|--|
| G1 | 0 | This means that the variables of group 1 should not be used as a feature of the optimal ANN. |
| G4 | 1 | This means that the variables of group 4 should be used as a feature of the optimal ANN. |
| G6 | 0 | This means that the variables of group 6 should not be used as a feature of the optimal ANN. |
| Hidden nodes | 60 | This means that the optimal ANN should have 60 hidden nodes. |
| G2 × G3 | G3 = 1 G2 = 0 | This means that the variables of group 3 should be used, while the variables of group 2 should not be used as features of the optimal ANN. |

Examining the financial performances of the proposed ANN

There are many criteria to evaluate the performances of a forecasting method. While the number of correct forecasts in the test period has been used as a criterion so far, the main goal of developing any financial forecasting method is making profit in real markets. Deciding on to what extent the optimal ANN can result in making profit comparing with other trading strategies is the purpose of this section. The common trading strategies that have been used in most of the relevant literature are as follows:

1. Buy-and-hold strategy (passive strategy).
2. Investing on an index that reflects the whole market.

However, since the objective of this research is to properly forecast the direction of S&P 500 index in a day, the buy-and-hold strategy and the investing strategy on S&P 500 are the same.

The following assumptions are made in trading in the S&P 500 market:

1. At the beginning of the test period, there is \$100,000 available as initial investment capital.
2. Every trade is made at the end of the day with the closing price of that day.
3. Transaction cost is flat rate and is \$8 per trade (O'Connor and Madden 2006).
4. Transaction costs are not deducted directly from the investment capital and are paid from a separate source. At the end of the test period for calculating the final capital, this cost will be considered.

Since the proposed ANN produces different results when initialized with different random values, one should initialize the designed ANN with different random values a number of times and compare the mean value of the final capitals produced by the ANN with the final capitals produced by other trading strategies. This helps to reach a statistically supportable conclusion. It should be noted that some of the previous studies such as those of Leung et al. (2000), Hussain et al. (2007), Yao et al. (1992), and O'Connor and Madden (2006) lacked the consideration of the effect of initialization.

For simulating trades in the real market using the forecasts produced by the proposed ANN, the following five steps are performed 100 times:

1. Initialize the network and train it by the training dataset.
2. Simulate the trained ANN with the test dataset and obtain the signals that the ANN produces in the test period.

3. When the ANN forecasts an increase in S&P 500 in the next day (the output of the ANN is positive), all of it will be used for buying SPDRs^b if the capital is in the form of cash, and no trade will be performed if the capital is in the form of stocks.
4. When the ANN forecasts a decrease in S&P 500 in the next day (the output of the ANN is negative), all of it will be liquidated if the capital is in the form of stocks, and no trade will be performed if the capital is in the form of cash.
5. At the end of the test period, all of the capital will be liquidated, and the total transaction cost is deducted. This leads us to the final capital.

In the buy-and-hold strategy, we spend all of the initial capital investment to buy SPDRs at the beginning of the test period and liquidate all of the purchased stocks at the end of the test period. This strategy produces an amount of \$89,593.3 as the final capital in the specified test period. For testing whether using the ANN produces more final capital than the buy-and-hold strategy does, one needs to test the following hypothesis:

$$H_0 : \mu \leq 89,593.3,$$

$$H_1 : \mu > 89,593.3,$$

in which μ is the average of 100 final capitals produced by trading in the market using the proposed ANN.

Since the variance of the final capitals is unknown and needs to be estimated by the sample variance and since the sample size is large enough (more than 30), to test the hypotheses, a t test was carried out with a p value of 0.02503. Hence, it is concluded that at the 5% significance level, the null hypothesis is rejected in the favor of the alternative hypothesis. The result of accepting alternative hypothesis is that using the proposed ANN significantly increases the final capital in comparison with the buy-and-hold strategy.

Forecasting S&P 500 daily direction using regression

According to the relevant literature, the relation between financial and economical variables is nonlinear. One of the reasons to employ a nonlinear method (ANN) to forecast the S&P 500 daily direction in this research is the nonlinearity of the relation. However, in order to test whether the nonlinearity assumption is true and to evaluate the performance of the proposed ANN methodology, in this section, we compare the performance of the proposed methodology with the ones of a linear model. Furthermore, since the response variable of interest, the daily direction of S&P 500, which is a binary variable, a logistic regression model (the logit model) is employed. For this model, the first 3,284 data (including the training and the verification data) are used to model

the relationship between the independent and the dependent variables. All of the 27 initial financial and economical variables are used as the independent variables of the logit model. In order to eliminate the redundant input variables, the forward selection method of the SAS software package is used. The estimated regression function becomes:

$$z = -0.1394 + 6.5569(SPY_{t-2}),$$

$$P = f(z) = \frac{1}{1 + e^{-z}},$$

in which SPY_{t-2} is the return of S&P 500 in day $t-2$ and P is the probability that the S&P 500 index decreases in the next day.

In order to validate the estimated regression function, a residual analysis based on the Box-Pierce method is conducted to see whether the error terms are white noise, i.e., they are independent standard normal random variables. The p value of the test statistic becomes 0.1145; indicating the null hypothesis cannot be rejected at 5% confidence level. Thus, the regression model is valid for forecasting. Applying the obtained regression equation to the test data, the logit model produces 189 correct forecasts in the test period.

For comparing the performance of the logit model with the ones of the proposed ANN, a statistical test of hypothesis is conducted as follows.

$$H_0 : \mu_{NN} \leq 189,$$

$$H_1 : \mu_{NN} > 189,$$

in which μ_{NN} is the mean of the number of correct forecasts that the optimal ANN produces with 50 random initializations. Once more, a t test may be employed for the test of hypothesis for which the p value is obtained as 0.001. Therefore, at 5% significance level, the null hypothesis is rejected in favor of the alternative hypothesis. This means that the proposed ANN as a nonlinear model can significantly outperform the logit model. Thus, it can be inferred that the relation between the independent variables and the dependent variable is nonlinear and that using a nonlinear model such as ANN is absolutely reasonable.

Conclusions and further research

In this research, the financial and economical features that have a significant influence on the daily direction of S&P 500 index were determined using DOE. Furthermore, an artificial neural network as a forecasting method was employed for which the influential features were used to forecast the daily direction of the S&P 500 index. The main findings of this research are as follows:

1. A factor that significantly influences the daily direction of S&P 500 is the exchange rates between the US dollar and three main currencies including the British pound, Canadian dollar, and Japanese yen.
2. Some of the variables that were found influential in previous studies on monthly forecasts do not have a significant influence on the daily forecast.
3. Using some of the variables including return of S&P 500 in previous days and relative change in the trading volume of S&P 500 not only is unhelpful, but also confuses the ANN and decreases the performance of ANNs for forecasting the daily direction of S&P 500.
4. Unlike some of the previous studies that did not consider the effect of network initialization, this paper does consider this effect.
5. The results obtained in this research are statistically supportable.
6. The factorial design of experiments helped to examine not only the main effects of the factors, but also the mutual effects between them. This led us to find out that the combination of the network features is also important for an ANN to produce good results.
7. By statistical analysis, it was proven that the designed ANN outperforms the logit model regarding the number of correct forecasts and outperforms the buy-and-hold strategy in terms of the obtained profit.

It should be mentioned that the conclusions are based on the use of only the feed-forward networks, and the scope extends to see if better results are achievable through the use of recurrent networks and other feed-forward varieties such as radial basis, adaptive neuro-fuzzy, and some other architectures. Future research may also include making the proposed ANN more reliable. As the results of the designed experiments showed, the mean value of the profit produced using the proposed ANN was more than the profit produced by the buy-and-hold strategy. However, the outputs of the proposed ANN are still not quite reliable because of two reasons. First, ANNs with different initializations produce different signals for a certain feature. Second, the proposed ANN with a certain initialization produces incorrect signals in some cases. Therefore, the first step in making the optimal ANN more reliable is initializing the optimal ANN in a way to produce as more correct signals as possible. Also, the second step identifies market situations in which using the proposed ANN with the best initialization is suitable and more likely to produce correct signals. Hence, it is recommended to use the signals produced by the proposed ANN in only certain market conditions. For example, signals produced by the proposed ANN may be useful in only a bull or

bear market. As a result, by examining the relation between market conditions and signals produced by the proposed ANN, one can discover the market conditions in which the signals produced are more reliable.

Endnotes

^aEarly stopping rule is a method to avoid over-fitting, but it improves the generalization ability of ANNs. This method halts the training process when the performance with validation data stops improving.

^bA short form of Standard & Poor's depository receipt, an exchange-traded fund managed by the State Street Global Advisors that tracks the S&P 500 index. Each share of spider contains one-tenth of the S&P index and trades at roughly one-tenth of the dollar-value level of the S&P 500.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

STAN and SH forecasted the daily direction of S&P 500 index using ANN. In order to select the most influential features of the proposed ANN that affect the daily direction of S&P 500, design of experiments were conducted to determine the statistically significant factors among 27 potential financial and economical variables along with a feature defined as the number of nodes of ANN. The results of employing the proposed methodology showed that ANN which uses the most influential features is able to forecast the daily direction of S&P 500 significantly better than the traditional logit model. Furthermore, experimental results of employing the proposed ANN on the trades in a test period indicated that ANN could significantly improve the trading profit as compared with the buy-and-hold strategy. All authors read and approved the final manuscript.

Authors' information

STAN is a professor of Industrial Engineering at Sharif University of Technology. His research interests are in the areas of Simulation Modeling and Analysis, Applied Statistics, and Multivariate Quality Control. Before joining Sharif University of Technology, he worked as a systems engineer and quality control manager for Iranian Electric Meters Company. He received his Bachelor of Science in Industrial Engineering from Sharif University of Technology in 1979, his Master's and PhD degrees both in Industrial Engineering from West Virginia University in 1989 and 1992, respectively. He is a member of $\sigma\mu\pi$. SH received his Bachelor of Science and Master's of Science degrees in Industrial Engineering from Sharif University of Technology in 2007 and 2009, respectively. His research interests are in the areas of Time Series Analysis and Finance.

Acknowledgments

The authors are thankful for the constructive comments of the anonymous reviewers as well as the editorial changes suggested by the language editor that certainly improved the presentation of the paper.

Received: 6 May 2012 Accepted: 24 November 2012

Published: 28 February 2013

References

- Atsalakis GS, Valavanis KP (2009) Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Syst Appl* 36:5941–5950
- Cao L, Tay F (2001) Financial forecasting using vector machines. *Neural Comput Appl* 10:184–192
- Cao Q, Leggio KB, Schniederjans MJ (2005) A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. *Comput Oper Res* 32:2499–2512
- Demuth H, Beale M (1998) *Neural network toolbox: for use with MATLAB*, 5th edn. The Math Works, Inc., Natick, MA

- Eakins SG, Stansell SR (2003) Can value-based stock selection criteria yield superior risk-adjusted returns: an application of neural networks. *Int Rev Financ Anal* 12:83–97
- Fama EF, French KR (1989) Business conditions and expected returns on stocks and bonds. *J Financ Econ* 25:23–49
- Fama EF, Schwert WG (1977) Asset returns and inflation. *J Financ Econ* 5:115–146
- Ferson W (1989) Changes in expected security returns, risk, and the level of interest rates. *J Finance* 44:1191–1217
- Hornik K, Stinchcombe M, White H (1989) Multilayer feed forward networks are universal approximators. *Neural Netw* 2:359–366
- Hosseini H, Luo D, Reynolds KJ (2006) The comparison of different feed forward neural network architectures for ECG signal diagnosis. *Medical Eng Phys* 28:372–378
- Hussain AJ, Knowles A, Lisboa PJG, El-Deredey W (2007) Financial time series prediction using polynomial pipelined neural networks. *Expert Syst Appl* 35:1186–1199
- Jensen M (1978) Some anomalous evidence regarding market efficiency. *J Financ Econ* 6:95–101
- Lam M (2004) Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. *Decis Support Syst* 37:567–581
- Leung MT, Daouk H, Chen A (2000) Forecasting stock indices: a comparison of classification and level estimation models. *Int J Forecast* 16:173–190
- Lin C, Huang J, Gen M, Tzeng G (2006) Recurrent neural network for dynamic portfolio selection. *Appl Math Comput* 175:1139–1146
- Mathworks (2004) *MATLAB*, version 7.0.0.19920 (R14). Mathworks, Inc, Natick, MA
- Motiwalla L, Wahab M (2000) Predictable variation and profitable trading of US equities: a trading simulation using neural networks. *Comput Oper Res* 27:1111–1129
- O'Connor N, Madden MG (2006) A neural network approach to predicting stock exchange movements using external factors. *Knowl Base Syst* 19:371–378
- Peters EE (1991) *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*. Wiley, New York
- Saad EW, Prokhorov DV, Wunsch DC (1998) Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Trans Neural Netw* 9:1456–1470
- SAS Institute Inc (2009) *SAS 9.1 TS Level 1M0*. SAS Institute Inc, Cary, NC
- Thawornwong S, Enke D (2004) The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing* 56:205–232
- Trippi RR, Turban E (1996) *Neural networks in finance and investing: using artificial intelligence to improve real-world performance*. Irwin Professional Pub, Burr Ridge
- Vanstone B, Finnie G (2009) An empirical methodology for developing stock market trading systems using artificial neural networks. *Expert Syst Appl* 36:6668–6680
- Versace M, Bhatt R, Hinds O, Shiffer M (2004) Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks. *Expert Syst Appl* 27:417–425
- Yao J, Tan CL, Poh H (1992) Neural networks for technical analysis: a case study on KLCI. *Int J Theor Appl Finance* 2:221–241
- Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. *Int J Forecast* 14:35–62
- Zhu X, Wang H, Xu L, Li L (2008) Predicting stock index increments by neural networks: the role of trading volume under different horizons. *Expert Syst Appl* 34:3043–3054

doi:10.1186/2251-712X-9-1

Cite this article as: Niaki and Hoseinzade: Forecasting S&P 500 index using artificial neural networks and design of experiments. *Journal of Industrial Engineering International* 2013 **9**:1.