DESIGN OF EXPERIMENTS ON NEURAL NETWORK'S PARAMETERS OPTIMIZATION FOR TIME SERIES FORECASTING IN STOCK MARKETS

Mu-Yen Chen[*], Min-Hsuan Fan[*], Young-Long Chen[†], Hui-Mei Wei[*]

Abstract: Artificial neural network (ANN) model has been used for years to conduct research in stock price prediction for three reasons. First, it has a higher prediction accuracy rate in empirical research. Second, it is not subject to the assumption of having samples from a normal distribution. Third, it can deal with non-linear problems. Nevertheless, the accuracy of prediction relies on the parameter settings of neural network as well as the complexities of problems and the neural network architecture; the results of the analysis could be even more significant with the selection of optimal parameters and network architecture. Currently, as a way of setting parameters, most researchers employed the trial and error method. However, this method is very time-consuming and labor-intensive and may not result in the optimal parameters. Therefore, this research took advantage of a back propagation neural network (BPNN) for the purpose of parameter optimization through constructing a model of stock price prediction, applying design of experiment (DOE) to systematize experiment scheduling, and methods of main effects analysis and interaction analysis. The research used two datasets of financial ratios from 50 blue chip companies in Taiwanese stock market and 40 listed American banks in New York stock exchange as experimental samples. Research results showed that the correlation forecasting, root mean squared error (RMSE), and computing time, which can effectively increase the accuracy of stock price prediction, are better than traditional statistical methods and conventional neural network model.

Key words: *Stock price prediction, back propagation neural network, design of experiment, financial ratios*

[*]Mu-Yen Chen, Min-Hsuan Fan, Hui-Mei Wei
Department of Information Management, National Taichung University of Science and Technology, Taichung 404, Taiwan, E-mail: `mychen@nutc.edu.tw, mfan@nutc.edu.tw, sampan@nutc.edu.tw`
[†]Young-Long Chen
Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taichung 404, Taiwan, E-mail: `ylchen66@nutc.edu.tw`

1. Introduction

Among diversified financial products, stock is an investment tool which is public and easy to obtain information. Stock price movements mean profits or losses on investments, and the ultimate purpose of investment and financial management is to gain maximum profits through minimum costs. Regardless of the means of evaluation to predict price, the stock price movement has always been the focus of concern not only for individual or institutional investors, but also for academic research.

Generally speaking, financial statements include an income statement, a balance sheet, a statement of stockholders equity and a statement of cash flows. The result of utilizing financial statements to analyze past operations or to evaluate current financial performance for the purpose of determining company stability is predictable [1]. Moreover, the investment behavior of investors is easily affected by accounting information released by each company [2]. Many scholars have focused their research on the relationship between company profitability and stock returns, and the research results proved that accounting information can be a reference base for investors to understand more about the business operations of each company [3-7]. Consequently, financial statements can provide information about financial position, business results, profitability and growth capacity of a company for investors as references of decision-making and investment planning.

The financial ratios are used to understand a company's financial information, such as a company's solvency, operating performance, operating capacity, growth potential and corporate risk. A variety of investors and stock market analysts utilize financial ratios to analyze the value of potential investment in a company [8-9]. However, there is often a non-linear relationship among related financial variables; unlike the analysis tools of artificial intelligence, traditional regression analysis cannot be used to more effectively simulate the non-linear relationship between financial variables. Nevertheless, for sales prediction [10], consumer preference [11], company financial diagnosis [12-13] and stock trading prediction [14-15], artificial neural networks (ANN) has been proven effective by numerous researchers. In addition, Chen (2012) discovered that in the application of predicting financial ratios, ANN has been proved to be more effective than traditional logistic regression and discriminant analysis methods [16]. Because of its effectiveness being applied to financial ratios, ANN has been used to determine the financial soundness of a company by many researchers [17], such as the application of bankruptcy prediction. Moreover, concerning stock price prediction, Dunis and Jalilov (2002) used the neural network regression (NNR) to construct the financial forecasting models and financial trading models for international stock markets [18]. Enke and Thawornwong (2005) indicated that there is a non-linear relationship between financial variables and the financial markets, and thought that ANN can effectively explore the relationship [19]. Lipinski (2005) discovered stock buying, selling and holding strategies through self-organizing maps and the K-means algorithm [20]. Zhu (2008) used ANN to analyze the effect of stock trading volume on the predictive ability to short-term, medium-term and long-term stock investment [21]. Al-Qaheri et al. (2008) applied three methods, rough set, genetic algorithms and ANN, to predict the stock price [22].

Empirical research has proven that ANN has a higher prediction accuracy rate, not being subject to the assumption of having samples from a normal distribution, and having no non-linear problems that variables have to deal with. Also, strengths of the ANN model include having less hypothesis restrictions and more capability of dealing with non-linear problems. However, the strength of the prediction function of ANN relies on the parameter settings, network architecture design, and the complexity of problem. If appropriate parameters and network architecture are selected, then the results are even more significant. Currently, most researchers manually adjusted the parameters to determine the optimal parameters of the neural network based on prior self-experience, recommendation settings in the literatures, or trial and error method adoption [23]. However, it may not be objective if the experimental parameters are determined merely by the experience of researchers. By doing so, only one factor variable can be changed at one time, the interaction between factors may not be able to be discussed, obtaining optimal parameters is very time and labor intensive, and the parameters found by the trial and error method may not be the optimal combination. Thus, instead of the traditional trial and error method which manually adjusts the parameters, this study employed the Design of Experiment (DOE) to establish a systematic research method for parameter design optimization, and developed a systematic prediction model to improve the accuracy of stock price prediction. The main purposes of this research are summarized as follows:

(1) in order to effectively increase the performance of back propagation neural network, this research took advantage of an experimental design to build a model for stock price prediction, and used financial indicators as research variables.

(2) in order to improve the weakness of the traditional trial and error method, specifically that manually adjusting the parameters is very time consuming, this research aims at assisting researchers in systematically conducting experiments through an experimental scheduling for saving time and costs, and conducting empirical research for the purpose of upgrading the prediction capacity of the model.

2. Literature Reviews

2.1 Back-Propagation Neural Network (BPNN)

A back-propagation neural network (BPNN) was first proposed by Werbos in 1974 [24], and its structure included:

(1) Input Layer: used to present the input variables of the network where the number of neurons are set based on the research questions.

(2) Hidden Layer: used to show the interaction between neurons. The numbers of hidden layers can be set according to the complexity of research problems, and the numbers of layers being set may impact the convergence effect of the neural network. In general, the optimal numbers of layers are determined by the trial and error method.

(3) Output Layer: used to display the output variables of the network where the number of neurons are set based on the research questions.

Basic principles of BPNN use the concept of the gradient steepest descent method to transmit data from the input layer to the hidden layer. After calculating and transforming data, the results will be transmitted to the output layer. When the gap between the output value and the actual value is calculated, the data will be returned to the hidden layer to correct weights. Through the process of constantly adjusting the magnitude of weight changes, error function between the output value of the network and the actual value will be minimized to achieve the network training until the error has been converged to the preset condition. The adjusted magnitude of the network and the error function to the sensitivity on weights are in direct proportion, i.e., the error function and the size of weighted partial differential value are in direct proportion. The relationship is shown as Eq. (1).

$$\Delta W = -\eta(\partial E/\partial W) \qquad (1)$$

Where ΔW is weight correction, η is learning rate used to control the magnitude of each weight correction, and E is error function. The definition of E is shown as Eq. (2).

$$E = \frac{1}{2} \sum (T_j - Y_j)^2 \qquad (2)$$

T_j: j-th output unit of target output value (actual value) at the output layer
Y_j: j-th output unit of network output value at the output layer

ANN is widely applied to the scope of financial prediction. Yao, Lim and Poh (1995) employed BPNN combined with data of previous day's closing price, moving average (MA), momentum, relative strength index (RSI) and stochastics (KD) to predict the next day's closing price of the stock market in an emerging stock market—Kuala Lumpur Stock Exchange—and compare the prediction accuracy with the autoregressive integrated moving average (ARIMA) time series model [25]. Results discovered that compared with ARIMA, the three-layer (6-4-3-1) ANN has a better prediction results and rate of return. Ramazan (1996) took the moving average method as a indicator while utilizing ANN to determine when to buy or sell stocks, and used two linear models—the auto-regression (AR) and the generalized autoregressive conditional heteroskedastic (GARCH) model—and a non-linear model, BPNN, for prediction [26]. Results showed that the predictive ability of the nonlinear BPNN is better. Fu and Xu (1997) integrated the genetic algorithm (GA) and the ANN model to predict the Shanghai stock index for the future trend. The results revealed that the predicted effect of short-term stock price prediction is great if the GA is applied with the ANN model [27].

2.2 Design of Experiment (DOE)

Design of Experiment (DOE) is proposed by Fisher in 1920, and applied to the genetic research on wheat varieties. The development of design of experiment can be divided into four periods. The first period is the agricultural period in 1920; Fisher proposed the analysis of variance (ANOVA) and applied the concept of design of

experiment to increase the production of agricultural products. In the application of the design of experiment, Fisher has stressed three important basic principles: randomization, replication and blocking. In 1980, Dr. Taguchi [28,29] proposed robust parameter design, at this time, design of experiment has officially entered the 3rd period. Due to the Taguchi method which is simple and easy to learn, and convenient to apply, engineers without statistical backgrounds love it. Meanwhile, western statistical scholars such as Box and Meyer [30], Nair [31] and Leon [32] have criticized the Taguchi method severely, which has also brought into the broad discussion and application of western scholars on the design of experiment, and further devoted into the study of the design of experiment. Design of experiment is therefore entered the fourth period. Many R&D has introduced engineer knowledge into the design of experiment, and further increase the applicability and efficiency.

Today, design of experiment has been broadly applied in agriculture, medical, commercial and manufacturing industries, and also has significant status in the scientific and industrial field [33]. Design of experiment finds the key variables through experiment, and improves the product quality, and further lowers the experiment frequency and shortens experiment time and improves the economic benefits of experiments. Currently, the known experiment arrangement has at least the following methods:

(1) Trial-and-error method

The trial and error method is completely based on personal experience and instinct to select experiment parameters. No data analysis needs to be processed, it is not a systematic method and depends on personal experience too much, and the trial-and-error experience is also difficult to pass it on to others.

(2) One-factor-at-a-Time (OFAT)

Only one factor is changed in each experiment, and other factors maintain in the original level until a satisfying effect appears. This experiment cannot prior predict the experiment frequency and budget, and the experimental treatment also cannot assure whether it is impacted by other factors [34].

(3) Full-factorial-experiments

The treatment of full factorial level will all appear in the experiment, and the reaction value of the treatment will be compared to find the best one. This method can find the best experimental treatment; however, the cost is too much, especially when the considering factors or levels increase, and it often leads to an oversized experiment scale and further resulted in incompletion.

As the above-mentioned, three methods all have its unavoidable shortcomings, therefore, Box and Hunter have applied the relationship defining method in 1961 to implement a specific part of the experiment, and it is also able to obtain the information on whether main effect and low stage interaction effect will have impact on product quality, this design is called "fractional factorial design", and they are currently widely applied in various product and process improvement field [33]. Fractional factorial experiments can be applied in three major objects: screening factors, optimization, and stability test. It applies some part of the full factorial

experiment to process experiment and statistical analysis, its purpose is to apply fewer experiments to obtain similar analysis result with the full-factorial design. Taking 2 levels and 4 factors as an example, if the full factorial design is applied, then 16 treatments will need to be processed. If fractional factorial design and half faction experiment is applied, then it is called 2^{4-1} fractional factorial experiment, and only 8 experiment treatment will be needed. Therefore, 2^{k-p} factor design has great assistance on screening important factors and saving cost. This research will explore how to effectively find and improve the ANN parameters when time and cost are under control through the design of experiment and expect to develop more potential application values for ANN.

2.3 Related Studies about DOE and BPNN for Time Series Forecasting

The ARIMA model presented by Box and Jenkins (1976),also known as the Box–Jenkins approach, has been widely applied and designed into the linear time series data corresponding to the ARIMA(p, d, q) structure [35].On the other hand, back-propagation neural networks with hidden layers are capable of modeling non-linear time series data. In the real world, most time series datasets are non-linear rather than linear, and neural network models outperform ARIMA models in many of these cases [36]. Researchers have recently begun using neural networks in finance engineering. Kaastra and Boyd (1996) discussed the application of neural network models into financial time series and modeled the forecasting process [37]. Zhang and Hu (1998) used a back-propagation neural network to forecast an exchange rate between the British pound and US dollar [38]. More recently, Yao et al. (2000) claimed the performance of a back-propagation neural network model is superior to the traditional Black–Scholes model for option price forecasting [39]. A substantial literature has been developed on the use of neural networks in financial time series forecasting. Chen et al. (2013) used an adaptive network-based fuzzy inference system (ANFIS) model to predict stock prices and directions for companies listed on the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX). The results indicated that the neural network model outperforms the multivariate model [40]. However, of the various ANN models in time series forecasting, the potential of the BPNN model has not been fully exploited due to several short-comings such as long training time, local minima issues, the lack of a fair and robust evaluation, and the lack of a proper network structure for effective modeling and forecasting [36].

In recent years, Balestrassi et al. (2009) proposed the full DOE design to tune suitable parameters of a multilayer perceptron neural network [41]. They tried to integrate the DOE into the ANN framework to forecast electricity demand of six industrial users in Brazil by using historical hourly time series. The empirical results showed that integrating DOE with the ANN model could obtained better performance than nonlinear autoregressive models. Their findings also suggested that the neural network models were quite competent for forecasting short-term time series datasets. Hsieh et al. (2011) integrated the DOE with BPNN to implement a robust DOE-based predictor [42]. As input components, they adopted technical indicators from seven publicly-listed companies in Taiwan. This hybrid system as

then used to forecast stock prices by a one-step sign rate. Empirical results showed that the proposed model could provide improved forecasting performance of stock prices.

Several studies have investigated the use of ANNs through experimental simulations. The literature survey on DOE and BPNN for stock price forecasting reveals three major drawbacks: (1) a lack of fair feature selection for input variables, (2) a lack of nonlinear and medium/long-term time series forecasting and (3) a lack of comparison with other artificial intelligence approaches. Therefore, the primary research motivation of this study is to integrate DOE and BPN to forecast nonlinear and seasonal time series, specifically real-world financial datasets for the Taiwan stock market and New York stock exchange. The stepwise regression method is used to select suitable financial indicators as the input variables. The seasonal accounting reports are then used as the experimental datasets to investigate performance for medium-term time series forecasting. Finally, the results produced using the proposed method is compared with those using other traditional statistical methods and artificial intelligence approaches.

3.   Research Methodology

3.1   Research Materials

This research applied the financial statement of Taiwan 50 Index in the Taiwan Stock Exchange Corporation (TSEC) stock market and 40 listed American banks in New York stock exchange (NYSE) as the analysis dataset. The data sampled is the seasonal report of the first season, the third seasonal, the semi-annual report, and the annual report. The Taiwan 50 Index companies are the leading listed companies of TSEC stock market, including cement, electronics, textiles, plastics, and financial holding companies. There are only 50 stocks chosen as constituent stocks, however, the total market value of constituent stocks already accounted 70% of the total market value of TSEC stock market, and the correlation coefficient of Taiwan 50 Index and TAIEX has also reached more than 98%, which shows that there is a relatively high correlation between them.

In addition, this research divided the dataset into two sub-datasets based on the subprime mortgage crisis in the U.S. The first sub-dataset is the financial ratio of Taiwan 50 Index companies from the first season of 2007 to first season of 2008, which has a total of 184 data. The second sub-dataset is the financial ratio of Taiwan 50 Index companies from the first season of 2010 to the fourth season of 2011, which has a total of 199 data. The purpose of applying two sub-datasets of Taiwan 50 Index is to verify whether the proposed model has the prediction ability that will not impact by the world economy status and local business cycle. This study also adopted 40 famous American banks listed on the NYSE, including 20 suffering financial distress and 20 operating normally. The experimental period was primarily from January 2006 to December 2009, amounting to 4 years coinciding with the recent financial crisis. The American banks datasets and experimental results are discussed in detail in Section 4.3.

3.2 Research Variables

The stepwise regression method fits the independent variables to the simple regression model and process F test, and then inspects the significance level of each variable. The standard value of the significance level is set as 0.05. Variables which fail to meet this threshold are discarded, while the remainder is selected for multiple linear regression according to their significance level [43]. The F test will be processed again on the variables included in the model after a variable is selected. If the significance level of the variable is lower than the standard value, the variable will be eliminated from the model. The procedure of the selection replication and variable elimination will be processed until all variables are inspected. This research refers to literature [44] [45] and applies stepwise regression method to process the screening of study variables.

No.	Code	Variables (Ratio)	Formula
1	X3	fixed assets to long-term debt	$\frac{[\text{fixed assets}]}{[\text{long} - \text{term debt}]}$
2	X8	Shareholders' equity to Long term debt	$\frac{[\text{shareholders equity}]}{[\text{Long term debt}]}$
3	X9	Capital to total assets	$\frac{([\text{liquid assets}] - [\text{liquid debt}])}{[\text{Total Assets}]}$
4	X12	Fixed assets to net worth	$\frac{[\text{Fixed assets}]}{[\text{shareholders equity}]}$
5	X20	Account Payables turnover	$\frac{[\text{Total supplier purchases}]}{[\text{Average accounts payable}]}$
6	X24	Fixed assets turnover	$\frac{[\text{Net sales}]}{[\text{Average net fixed assets}]}$
7	X25	Total assets turnover	$\frac{[\text{Net sales revenue}]}{[\text{Average total assets}]}$
8	X28	Pre-tax profit to paid-in capital	$\frac{[\text{Pre} - \text{tax profit}]}{[\text{paid} - \text{in capital}]}$
9	X30	Operating Expense	$\frac{[\text{Operating Expense}]}{[\text{Net sales}]}$
10	X33	Net income before tax	$\frac{[\text{Net income before tax}]}{[\text{Net sales}]}$
11	X35	Pre-tax rate of return on equity	$\frac{[\text{Pre} - \text{tax income}]}{[\text{shareholders equity}]}$
12	X37	Pre-tax rate of return on assets	$\frac{[\text{Earnings Before Interest and Tax (EBIT)}]}{[\text{Total assets}]}$
13	X39	Return on Fixed Assets	$\frac{\text{Net profit}}{\text{Fixed assets}}$
14	X40	Interest expenses to operating income	$\frac{[\text{Interest expense}]}{[\text{Operation Income}]}$
15	X41	Interest Protection Multiples	$\frac{([\text{EBIT}] - [\text{Interest expense}])}{[\text{Interest expense}]}$
16	X54	Price-to-book	$\frac{[\text{Stock price}]}{[\text{Net asset value}]}$

Tab. I *Financial variables.*

This research referring to the related literature [44,45,46], and applied the 54 commonly used financial indicators as the independent variables, and further applied Eq. (3) to find the variables that have major impacts on the stock price with stepwise regression method as the input variable of back propagation neural network. In Eq. (3), y refers to the stock price on the 5th business day after the seasonal financial statement publication date, X refers to the 54 financial ratios that measure the asset structure, asset management ability, operation ability, market value and liquidity, and t is the publication date of the quarterly financial statement. To measure the medium-term effect after the disclosure of the quarterly financial statement, this research applied the stock price on the 5th business day after the publication date as the output variable.

$$y_{t+5} = f(\ln X_{1t}, \ln X_{2t}, \ldots, \ln X_{54t}) \tag{3}$$

Tab. I shows the 16 significant output variables screened through stepwise regression. These 16 financial ratios include financial leverage, asset management, short-term repayment ability, earning power and market value, which is applicable for the follow-up data analysis and handling for the study.

3.3   BPN Model Construction

This research applied steepest descent method to look for the weight changes and allow a minimum error. According to previous studies [47, 48, 49], there are many conditions that converge the neural network: (1) when the root mean square error (RMSE) of the actual value and the prediction value is closed to a certain default value [42]; (2) when the default training iteration is achieved. This research applied the above-mentioned methods to gradually increase the ANN training time to further reduce RMSE value until RMSE is steady and appropriate. The definition of RMSE is shown as Eq. (4):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - y_i)^2} \tag{4}$$

In which, $N, d_{i,}$ and y_i respectively represent training sample amount, the actual value of the i-th training sample and the prediction value of the i-th training sample.

If more training datasets can be provided and more detail classification can be processed, then the prediction result will be more consistent. For the back propagation neural network, the training and validation datasets were limited to the functional value. Therefore, this study adopted the min-max scaling method to handle the data normalization process and the data have to be standardized according to Eq. (5):

$$PN = \frac{P - P_{min}}{P_{max} - P_{min}} \times (D_{max} - D_{min}) + D_{min} \tag{5}$$

where PN is the standardized data, P is the former data, P_{max} is the former data with maximum value, P_{min} is the former data with minimum value, D_{max} is the standardized data with maximum value, D_{min} is the standardized data with minimum value, and the value of the standardized data is in a range of [-1, 1].

3.4 DOE 2^{k-p} Model Construction

Design of experiment is usually processed in two stages. The first stage will find the significant factor and further effectively control it. The second stage will focus on the significant factors to find out the working interval. Finally, the experiment data will apply the ANOVA to calculate the contribution of the interaction effect between each factor, and the factor contribution will further be used to identify the significant factors.

When applying full factorial design of experiment, if there are a lot of factors, the computation cost and execution time of the experiment conductor may all be impacted. The 2^{k-p} fractional factorial experiment is a high performance screening experiment method. The experimental frequencies of Montgomery's 2^k full factorial design will increase significantly according to the increasing factor number, K, such as $2^3 = 8$, $2^4 = 16$, and $2^6 = 64$ [50]. In the other way, 2^{k-p} fractional factorial experiment has k factors, each factor has two levels, and finally has 2^{k-p} experimental frequencies. To simply the experiment combination, the full factorial design is designed to the 2^{k-p} fractional factorial design of experiment. However, taking 2^6 as an example, 64 experiments will make 63 total degree of freedom, in which, there are only 6 degree of freedom are main factor effect, 15 degree of freedom is the interaction effect of the two factors. However, there are 42 degree of freedom are under the interactions of 3 and more factors. Therefore, it can assume that multiple factor interaction effect is not significant, thus only partial experiment of the 2^k experiments need to be done to understand the main factor effect and low level factor interaction effect. 2^{k-p} fractional factorial experiment will then be created. It can be seen that 2^{k-p} fractional factorial experiment is currently the most commonly used statistical sampling method.

To conclude, to save the time and cost of the experiment, and obtain better data with less experiment, this research applied the two level fractional factorial (2^{k-p}) design of experiment to explore the interaction effect relation between the structural parameters, learning parameters, and reaction variables during ANN training.

4. Experimental Results and Discussion

4.1 Experimental Results

In comparison with the traditional regression analysis, to effectively improve the prediction accuracy, this research has combined design of experiment and back propagation neural network to construct stock price prediction model. However, previous studies mostly applied trial and error method to process manual adjustment setting on the parameter setting of ANN. To lower the time cost created by trial and error and further improve the prediction rate of ANN, this research applies design of experiment method instead of the traditional trial and error method to process ANN parameter optimal design. In this section, we used the Taiwan 50 Index in TWSE stock market as an example to illustrate the experimental process.

- Step 1: Select the financial ratio of the Taiwan 50 listed companies in TWSE stock market as the sample

The Taiwan 50 Index is applied as the analysis dataset, and the financial ratio of Taiwan 50 Index listed companies in between the first season of 2007 and the first season of 2008 is used as the output variable. The 54 financial ratio is used as the research variables, and the stock price on the 5th business day after the financial statement publication date is the output variable.

- Step 2: Screen variations with stepwise regression

  54 variables are applied through stepwise regression to find out the 16 variables that has major impacts on the stock prices as the output variables of back propagation neural network. In this research, the statistical significance level is 0.05 and the variables obtained are shown as Tab. I.

- Step 3: Validate experiment factor and quality characteristic

  This research applies the four parameters of hidden layer neurons number, learning rate, momentum, and Epoch in the ANN learning process as the design of experiment factors according to literature exploration [42], and every factor has 2 levels, as shown in Tab. II. It is explained in Tab. II that level is the quantitative state of a factor considered in the experiment, such as the factor of the hidden layer neurons number, N considers to process experiments respectively in low level (N = 8) and high level (N = 20), or observe the convergence effect of Epoch in low level (Epoch = 20,000) and high level (Epoch = 40,000).

  Too few hidden neurons in the neural network model could result in a high training error and high generalization error due to underfitting [50]. Too many hidden neurons could results in a low training error, but produce a high generalization error due to overfitting [50]. In this study, the number of neurons in the hidden layer is decided on a rule of thumb [51,52], which holds that (a) be between the input and output layer size, (b) is set to something near (inputs+outputs) * 2/3, or (c) is never larger than twice the size of the input layer. Therefore, this parameter is set to between 8 and 20.

  In the back propagation neural network, a lower learning rate would result in overly slow network learning, but a higher learning rate could cause the objective function to diverge. In the traditional approach, the neural network is trained at a constant learning rate, a tedious process of trial and error. With DOE training, there is no need to use a constant learning rate. In this study, this parameter is set to between 0.3 and 0.9 following Saarinen et al. (1993) [53], Sarle (1999) [54], and Hsieh et al. (2011) [42].

  The BPNN model suffers from two main disadvantages a longer training time and slow convergence. The momentum is thus used to speed up the training process of neural network model. In this study, this parameter is set to between 0.5 and 0.9, following Hsieh et al. (2011) [42], Sheela and Deepa (2011) [55].

  The number of training cycles depends on the learning rate and momentum, making it necessary to choose the most suitable values for the parameter. In the traditional approach, the maximum number of epochs is always very high, the termination of the network learning process is set according to the

validation set or the MSE. In this study, this parameter is set to between 0.5 and 0.9, following Sarle (1999) [54], Hsieh et al. (2011) [42], Sheela and Deepa (2011) [55].

Item	Control factor	Level	
		1 (Low)	2 (High)
N	Number of neurons in the hidden layer	8	20
LR	Learning rate	0.3	0.9
M	Momentum	0.5	0.9
Epoch	Number of epochs	20,000	40,000

Tab. II *The level settings for each control factor.*

- Step 4: Establish stock price prediction model

  This research applies 2^{4-1} fractional factorial experiment, and processes 3 replications to lower the variation of the experiment and increase the accuracy of the experiment. It adds a central point in the high and low experiment level to explore whether there is a non-linear reaction value between the high and low levels. The design of the two levels is often assumed to be a linear relation, whether there is a curvilinear relation between the two points can be seen by adding a central point. Therefore, this research has added 6 central points to process the high and low level curvature inspection. The major reason for this research chooses to process three replications are if only the single replicate is processed then the variation of the experiment will be severe. Therefore, to maintain the stability of the experiment, more replication is chosen to be done instead of processing more experiment treatments. Then an experiment is processed with the planned factor level parameter value to cope with fractional factorial design of experiment. In order to evaluate the prediction performance, the proposed approach could be expressed with two performance indicators: testing R (TR, Pearson correlation), and the RMSE. The result of the experimental factor combination and reaction value is shown as Tab. III.

  In statistics, the Pearson correlation is defined as the strength and direction of the relationship between the measured and predicted output values of the output variable. A correlation value over 0.8 is regarded as very strong, while 0.6 to 0.8 is regarded as moderately strong, 0.3 to 0.5 is fair, and below 0.3 is poor [56]. To verify the feasibility of using the neural network to predict the parameters for stock prices, this study adopted Pearson correlation coefficients to evaluate the quality of the trained network. As shown in Tab. III, the Pearson correlation coefficient reached an average value at 80.21% with a comparatively low standard deviation of 0.1066. This represents a high correlation between the input variables and output variable, presenting a good representation of the estimated properties.

- Step 5: The main effect and the interaction effect are analyzed with design of experiment

Std Order	Run Order	Cener Pt	Blocks	N	LR	M	Epoch	RMSE	R
6	1	1	1	20	0.3	0.9	20000	0.0270554	0.802397
23	2	1	1	8	0.9	0.9	20000	0.0470643	0.916068
28	3	0	1	14	0.6	0.7	30000	0.0312505	0.772462
27	4	0	1	14	0.6	0.7	30000	0.0209704	0.766201
18	5	1	1	20	0.3	0.5	40000	0.0291753	0.750341
14	6	1	1	20	0.3	0.9	20000	0.0206281	0.782977
16	7	1	1	20	0.9	0.9	40000	0.0166585	0.864046
26	8	0	1	14	0.6	0.7	30000	0.0207872	0.727248
13	9	1	1	8	0.3	0.9	40000	0.0295201	0.597599
25	10	0	1	14	0.6	0.7	30000	0.0241962	0.896121
30	11	0	1	14	0.6	0.7	30000	0.026363	0.855076
8	12	1	1	20	0.9	0.9	40000	0.0138229	0.941246
4	13	1	1	20	0.9	0.5	20000	0.0276226	0.84149
1	14	1	1	8	0.3	0.5	20000	0.0388042	0.982435
21	15	1	1	8	0.3	0.9	40000	0.0333401	0.853172
17	*16*	*1*	*1*	*8*	*0.3*	*0.5*	*20000*	*0.0361917*	*0.962001*
24	17	1	1	20	0.9	0.9	40000	0.0279477	0.841218
20	18	1	1	20	0.9	0.5	20000	0.0254498	0.748895
7	19	1	1	8	0.9	0.9	20000	0.033244	0.570916
19	20	1	1	8	0.9	0.5	40000	0.0279061	0.745501
11	21	1	1	8	0.9	0.5	40000	0.0302484	0.659614
15	22	1	1	8	0.9	0.9	20000	0.0320179	0.70771
5	23	1	1	8	0.3	0.9	40000	0.0222027	0.868459
9	24	1	1	8	0.3	0.5	20000	0.0399339	0.962057
29	25	0	1	14	0.6	0.7	30000	0.0248234	0.62418
2	26	1	1	20	0.3	0.5	40000	0.0283103	0.799933
12	27	1	1	20	0.9	0.5	20000	0.027048	0.754672
22	28	1	1	20	0.3	0.9	20000	0.0236742	0.745429
10	29	1	1	20	0.3	0.5	40000	0.0289298	0.898542
3	30	1	1	8	0.9	0.5	40000	0.0307752	0.826263

Tab. III *Design-of-experiment results for RMSE and testing R.*

After ANN training and testing experiment are completed, a factor analysis is processed on the experimental factor combination and reaction value in Tab. III. As shown in Fig. 1, factors with greater reaction effects will be identified from this Pareto chart illustration. There are four factors used in the standardized effects analysis, including A, B, C, D and each factor means neurons numbers, learning rate, momentum, and number of Epochs, respectively. It can be seen that the AB experimental factor combination created by the interaction between A (N) and B (LR) factors has a significant impact on reaction value R and has exceeded the critical value. For reaction value R, the main effect and the interaction effect between the factors of A (N), B (LR), C (M), D (Epoch) are respectively drawn into Fig. 2 and Fig. 3. Due to a central point is added, there is an extra central point between the

high and low levels of the reaction plot. At this time, we can test the effect of different factors at the same time, not having to process experiment on each factors respectively, and can also observe the interaction effect created between factors in a single and complicated experiment. Fig. 2 is the main effect reaction plot of the four factors, which is the difference condition of every level between the located treatment reaction value means.
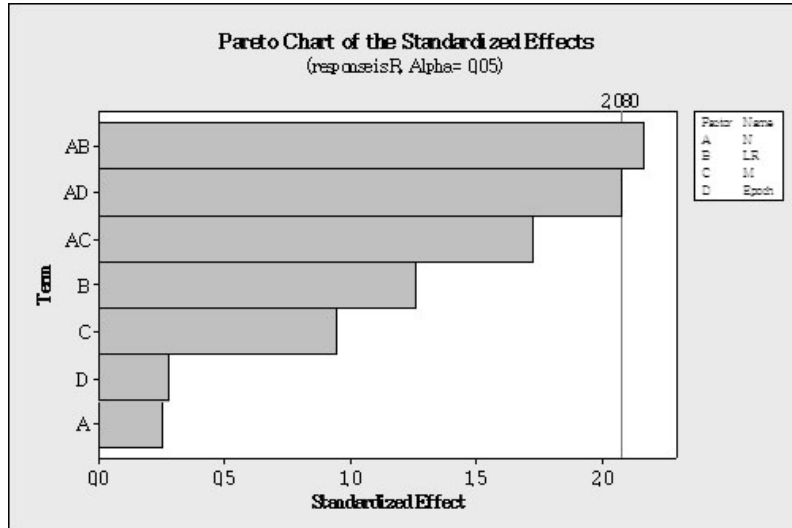

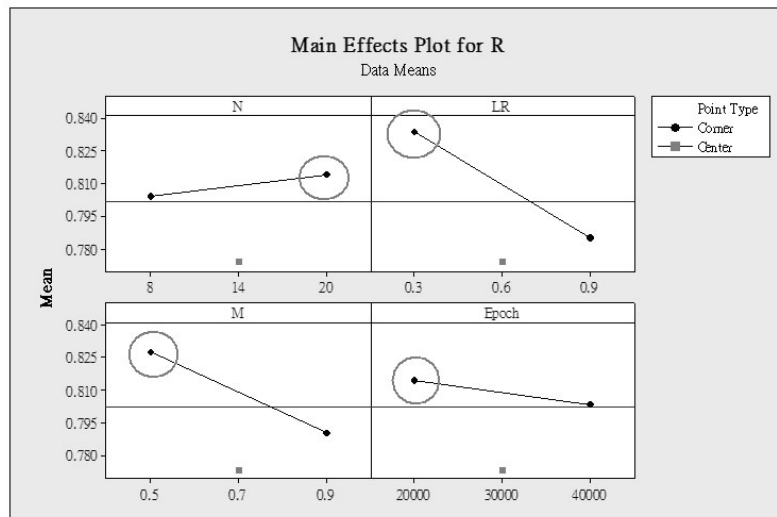**Fig. 1** *Standardized effects of the DOE-based BPNN factors.*

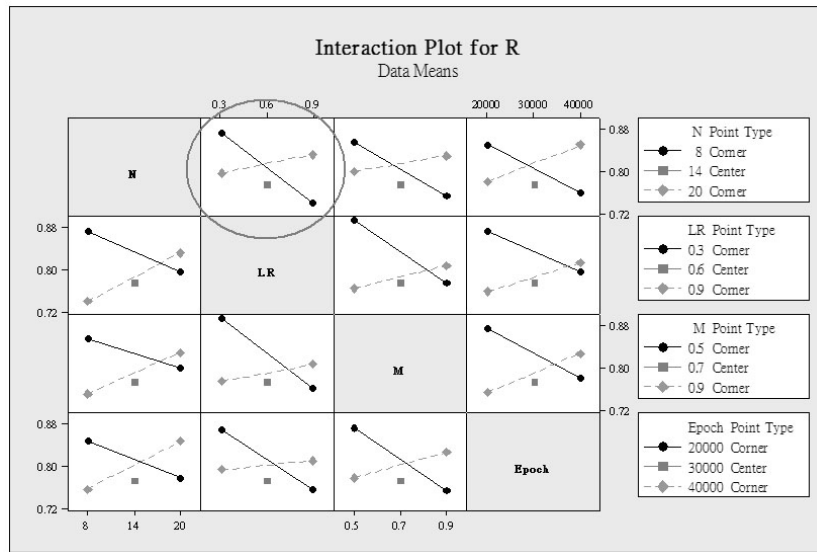**Fig. 2** *Main effects of the DOE-based BPNN factors.*

Fig. 3 *Interactive relationship of the DOE-based BPNN factors.*

Fig. 2 illustrates the number of neurons at low level then the experimental results would get higher values of R. In addition, the same situation with learning rate, Fig. 2 also illustrates the lowest levels of momentum and Epochs will match with the higher values of R. However, there exist some interactions among number of neurons and learning rate as shown in Fig. 1. Hence, the optimization parameter setting cannot be judged from the main effect plot in Fig. 2.

It shows in Fig. 3 that when the effect of a factor under a different level of another factor is different, then it is called that these two factors have interactions. If the effect of a factor under a different level of another factor is the same in an experiment, then it is called that these two factors have no interactions. In the interaction plot, the results indicated the number of neurons in the hidden layer should be 8 and differences with main effects plot.

- Step 6: Decide the optimal parameter setting

  An optimal design is implemented through the analysis of main effects and factor interaction effects, the result is shown as Fig. 4, in the condition of the optimal combination of the four factors, reaction value y can reach 0.9617.

- Step 7: Experiment validation

  After the optimal parameter is obtained, verification must be processed to verify whether the optimal parameter setting is correct. Therefore, this research uses the *k*-fold approach with an optimal parameter combination according to the optimal parameters in Fig. 4. The *k*-fold approach [57] is a cross-validation method used to evaluate the classification accuracy rate. For the purposes of this research, *K* is set as 5 and the data was divided into five
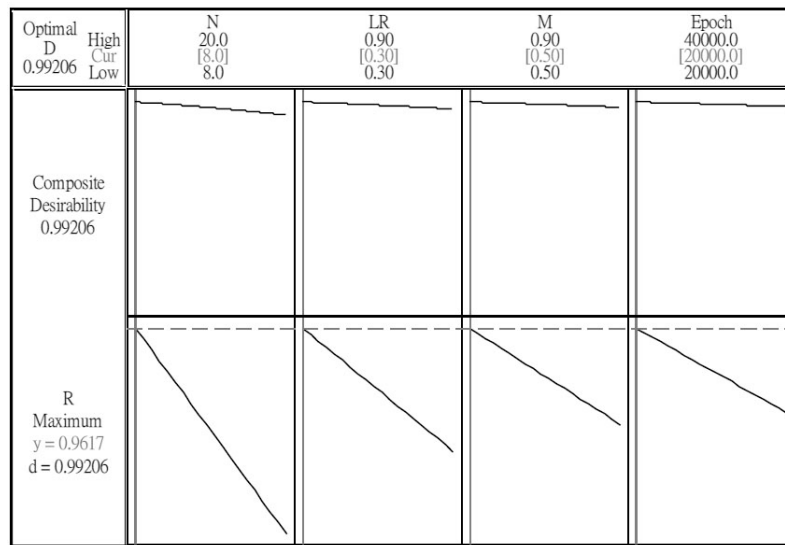
Fig. 4 *Optimal parameter setting for DOE-based BPNN with 1st sub-dataset.*

portions. The final average accuracy rate was the average of the five accuracy rates. Therefore, the hidden layer of neurons number is set to 8, learning rate is set to 0.3, momentum is set to 0.5 and numbers of Epoch is set to 20,000. Finally, the experiment validation processed is the optimal parameter combination value and its experimental result as shown in Tab. IV. Therefore, the hidden layer of neurons number is set to 8, learning rate is set to 0.3, momentum is set to 0.5 and numbers of Epoch is set to 20,000. Finally, the experiment validation processed is the optimal parameter combination value and its experimental result as shown in Tab. IV.

Epoch	N	LR	M	Y1	Y2	Y3	Y4	Y5	$\bar{Y}$
20,000	8	0.3	0.5	0.9103	0.9493	0.9436	0.9614	0.8729	0.9275

Tab. IV *Experiment results with 1st sub-dataset.*

The second sub-dataset is from the first season of 2010 to the fourth season of 2011, the experimental process is processed by repeating step 4 to step 7. After summarizing the optimal result of step 4 to 7, the parameter combination is shown as Fig. 5, the neurons number of 8, learning rate of 0.3, momentum of 0.9 and Epochs of 20,000. In the optimal combination condition of the four factors, reaction value y can reach 0.9166, then apply this optimal parameter combination to respectively process 5-fold cross validation to obtain the stability, finally process the experiment validation. The optimal parameter combination and the experimental result are shown as Tab. V.
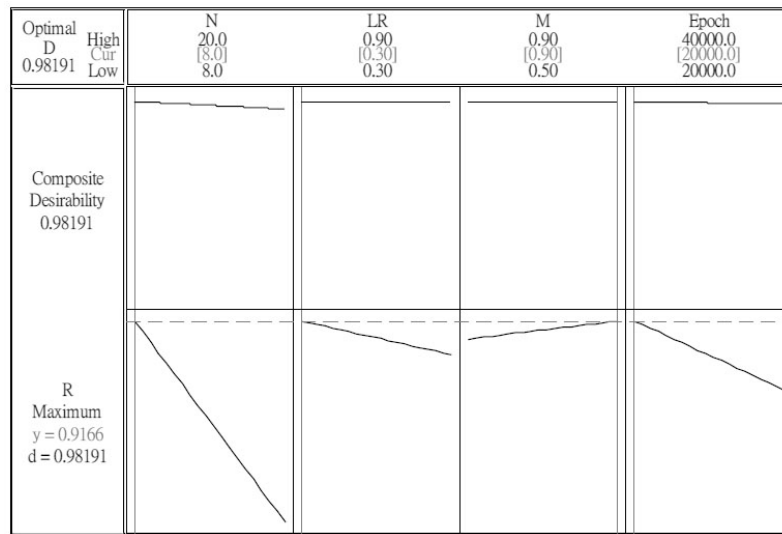
Fig. 5 *Optimal parameter setting for DOE-based BPNN with 1st sub-dataset.*

Epoch	N	LR	M	Y1	Y2	Y3	Y4	Y5	Ȳ
20,000	8	0.3	0.9	0.8509	0.9107	0.9089	0.8455	0.8433	0.8719

Tab. V *Experiment results with 2nd sub-dataset.*

4.2　Comparative Research and Discussion

The platform adopted to develop the proposed DOE-based BPNN model and other AI approaches uses a PC with the following attributes: Intel I7 eight-core CPU, 16G RAM, Windows 7 OS and the MATLAB R2013 development environment. For the purposes of discussion, the experimental parameters settings were identical to those noted in Section 4.1 and 5-fold cross validation was used to evaluate forecasting performance.

This research has proposed the method that integrates traditional design of experiment and BPNN to improve the performance of stock price prediction. To compare the performance of the proposed model, the Pearson correlation, RMSE, and CPU computing time processed on the testing dataset with linear regression, least median squares, BPNN, radial basis function networks (RBFN), support vector regression (SVR) are summarized as Tab. VI. The Pearson correlation and RMSE of the proposed method are all better than traditional statistical analysis and conventional neural networks. The proposed method is verified to effectively improve the prediction performance. In addition, this research has divided the dataset into two sub-datasets based on the Subprime Mortgage Crisis in the U.S. The experiments are processed according to the design of experiment steps, and the result shows the outcome obtained from both first sub-dataset and second sub-dataset are better than the other models, as shown in Tab. VI.

In the 1st sub-dataset experiment, the proposed DOE-based BPNN model only needs 36 seconds of CPU time with 9.27 RMSE, as opposed to the statistical methods such as the linear regression approach which required about 286 seconds with 23.791 RMSE, while least median squares approach required 232 seconds with 17.159 RMSE. In addition, the proposed DOE-based BPNN model outperformed the conventional BPNN, RBFN, and SVR approaches, which respectively needed 384, 618, and 125 seconds of CPU time with 15.544, 42.628, and 12.537 RMSE. The same results were found in the 2nd sub-dataset experiment, where the proposed DOE-based BPNN model needed the least CPU time and lowest RMSE value to predict the stock price variation.

Prediction Models	1st sub-dataset			2nd sub-dataset		
	Pearson correlation	RMSE	Computing Times (s)	Pearson correlation	RMSE	Computing Times (s)
Linear Regression	0.8845	23.791	286	0.8296	20.528	247
Least Median Squares	0.9039	17.159	232	0.5342	30.776	492
BPNN	0.8894	15.544	384	0.8481	21.483	461
RBFN	0.4057	42.628	618	0.1315	34.218	532
SVR	0.9095	12.537	125	0.8206	20.884	203
DOE-based BPNN	0.9275	9.27	36	0.8719	14.45	54

Tab. VI *Performance comparison with other models.*

This research presents several key findings regarding the implications and determinants of time series predictions of stock price:

(1) Our approach requires 80% fewer financial ratios than other methods but still presents highly-accurate stock price predictions with lowly RMSE. This also proves the stepwise regression methodology could be a suitable approach to process feature selection.

(2) This research found that the Pearson correlation would significantly influence the prediction accuracy, especially in time series datasets. This issue has been seldom explored in prior researches on stock price prediction.

(3) This research empirically determined the DOE-based optimization approach achieved better forecasting accuracy than other statistical approaches, such as linear regression, least median squares and SVR. Besides, it also has significant performance than conventional neural networks, such as BPNN and RBFN. Furthermore, the DOE-based optimization approach has rarely been used for forecasting, especially for financial problems.

4.3 America Banking Datasets

To validate the general application of the proposed model in practice, this study collected the history financial data from 40 American banks listed on the New York

Stock Exchange (NYSE). In addition, these 40 American banks were including 20 bankruptcy banks and 20 healthy banks. The collection period was mainly from January 2006 to December 2009. All the financial indicators adopted in this study were gathered from balance sheets, cash flow statements, and income statements. Detailed information for the 40 American banks is provided in "Appendix B".

This experiment directly adopted the 16 commonly used financial indicators from Table 1 as the input variables for the back propagation neural network, and then predicted stock prices with the DOE-based BPNN model. As shown in Table VII, in the non-distressed sub-dataset experiment, the proposed DOE-based BPNN model only needed 62 seconds of CPU time with 8.15 RMSE, while the statistical methods such as the linear regression approach which required about 142 seconds with 15.623 RMSE, and the least median squares approach which required 166 seconds with 16.267 RMSE. In addition, the proposed DOE-based BPNN model also outperformed the conventional BPNN, RBFN, and SVR approaches, which respectively required 213, 365, and 94 seconds of CPU time with 12.783, 23.892, and 10.126 RMSE. The same results were found for the financially distressed sub-dataset experiment, where the proposed DOE-based BPNN model needed the least CPU time and the lowest RMSE value to predict stock price variations. Thus, the findings of this experiment are valuable and indicate that the proposed model outperforms other approaches in practice.

Prediction Models	Non-Financial Distress Dataset			Financial Distress Dataset		
	Pearson correlation	RMSE	Computing Times (s)	Pearson correlation	RMSE	Computing Times (s)
Linear Regression	0.9022	15.623	142	0.8096	21.436	184
Least Median Squares	0.8824	16.267	166	0.7133	26.528	203
BPNN	0.9108	12.783	213	0.8326	18.693	248
RBFN	0.6212	23.892	365	0.6034	28.126	356
SVR	0.9216	10.126	94	0.8594	15.768	98
DOE-based BPNN	0.9432	8.15	62	0.9234	10.28	66

Tab. VII *Performance comparison with other models of NYSE Datasets.*

5. Conclusions

Stock investment is an investment toll which is popular and commonly used in the world. Applying ANN on the prediction of stock price has been verified as an effective prediction tool. Most academic researchers have applied trial and error according to their experiences to get the optimization parameters' setting of ANN. This research can successfully find out the key factors through applying design of experiment systematically, and further find the optimal parameter combination. The empirical result shows a significant improvement for the correlation coefficient

of ANN. In addition, the application of fractional factorial experiment will also allow a significant decrease of experiment frequency, and further save time and cost. The difference is that it can explore many factors during one experiment without going through the trial and error. Therefore, it can be seen that the proposed methodology is feasible.

There are some researches can be worth to investigate. First, bio-inspired computing algorithms can be considered to integrating with SVM model. Second, the parameter optimization method proposed by this research only applies BPNN as the study object, future studies will be able to verify the predication performance of other neural network models. Third, this research defines the stock price on the 5th business day after the publication date of the financial statement as the output variable. Future studies can explore towards a mid-term or long-term financial statement disclosure. Finally, this research recommends adding more experimental factors on design of experiment to improve the interpretability of reaction value.

References

[1] Altman, E.I., Brenner, M: Information Effects and Stock Market Responseto Signs of Firm Deterioration. Journal Financial and Quantitative Analysis, **16(1)**, 1981, pp. 35-51.

[2] Sarkar, A., Schwartz, R.A.: Market Sidedness: Insights into Motives for Trade Initiation. Journal Finance, **64(1)**, 2009, pp. 375-423.

[3] Latan, H.A., Jones, C.P.: Standardized Unexpected Earnings–1971-77. Journal Finance, **34**, 1979, pp. 717-724.

[4] Abarbanell, J.S., Bushee, B.J.: Fundamental Analysis, Future Earnings, and Stock Prices. Journal Accounting Research, **35(1)**,1997, pp. 1-24.

[5] Abarbanell, J., Lehavy, R: Can Stock Recommendations Predict Earnings Management and Analysts' Earnings Forecast Errors. Journal Accounting Research, **41(1)**, 2003, pp. 1-31.

[6] Ou, J.A., Penman, S.H.: Accounting Measurement, Price-Earnings Ratio, and the Information Content of Security Prices. Journal Accounting Research, **27**, 1989, pp. 111-144.

[7] Shin, H.S.: Disclosure Risk and Price Drift. Journal Accounting Research, **44(2)**, 2006, pp. 351-379.

[8] Hopwood, W.S., Schaefer, T.F.: Incremental information content of earnings and nonearnings-based financial ratios. Contemporary Accounting Research, **5**, 1988, pp. 318-342.

[9] Mensah, Y., Werner, R.: The capital market implications of the frequency of interim financial reporting: an international analysis. Review Quantitative Finance and Accounting, **31**, 2008, pp. 71-104.

[10] Alon, I., Qi, M., Sadowski, R.J.: Forecasting aggregate retail sales:: a comparison of artificial neural networks and traditional methods. Journal Retailing Consumer Services, **8**, 2001, pp. 147-156.

[11] Chiang, W.K., Zhang, D., Zhou, L: Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression. Decision Support System, **41**, 2006, pp. 514-531.

[12] Chen. M.Y.: A Hybrid ANFIS Model for Business Failure Prediction - Utilization of Particle Swarm Optimization and Subtractive Clustering. Information Sciences, **220**, 2013, pp. 180-195.

[13] Chen, M.Y.: Visualization and Dynamic Evaluation Model of Corporate Financial Structure with Self-Organizing Map and Support Vector Regression. Applied Soft Computing, **12(8)**, 2012, pp. 2274-2288.

[14] Chang, P.C., Liu, C.H., Lin, J.L., Fan, C.Y., Ng, C.S.P.: A neural network with a case based dynamic window for stock trading prediction. Expert System with Applications, **36**, 2009, pp. 6889-6898.

[15] Zemke, S.: Nonlinear index prediction," Physica A: Statistical Mechanics and its Applications, **269(1)**, 1999, pp. 177-183.

[16] Chen, M.Y.: Comparing Traditional Statistics, Decision Tree Classification and Support Vector Machines Techniques for Bankruptcy Predicting. Intelligent Automation & Soft Computing, 18(1), 2012, pp. 65-73.

[17] Shie, F.S., Chen, M.Y., Liu, Y.S.: Prediction of Corporate Financial Distress: An Application of the America Banking Industry. Neural Computing and Applications, **21(7)**, 2012, pp. 1687-1696.

[18] Dunis Ch. L., Jalilov, J.: Neural network regression and alternative forecasting techniques for predicting financial variables. Neural Network World, **12(2)**, 2002, pp. 113-139.

[19] Enke, D., Thawornwong, S.: The use of data mining and neural networks for forecasting stock market returns". Expert Systems with Applications, **29(4)**, 2005, pp. 927-940.

[20] Lipinski P.: Clustering of large number of stock market trading rules. Neural Network World, **15(4)**, 2005, pp.351-357.

[21] Zhu, X., Wang, H., Xu, L., Li, H.: Predicting stock index increments by neural networks: The role of trading volume under different horizons. Expert Systems with Applications, **34(4)**, 2008, pp. 3043-3054.

[22] Al-Qaheri H., Hassanien A. E., Abraham A.: Discovering stock price prediction rules using rough sets. Neural Network World, **18(3)**, 2008, pp. 181-198.

[23] Martinez, E. E., Smith, A. E., Idanda, B.: Reducing Waste in Casting with a Predictive Neural Model. Journal of Intelligent Manufacturing, **5(4)**, 1994, pp. 277-286.

[24] Werbos, P.: Beyond regression: new tools for prediction and analysis in the behavior sciences. Ph.D. dissertation, Harvard University, Cambridge, MA, August 1974.

[25] Yao, J., Poh, H.L.: Forecasting the KLSE Index Using Neural Networks. Proceeding, IEEE International Conference on Neural Network, **2**, 1995, pp. 1012-1017.

[26] Gencay, R.: Non-linear prediction of security returns with moving average rules. Journal of Forecasting, **15**, 1996, pp. 165-174.

[27] Fu, K., Xu, W.: Training neural network with genetic algorithms of forecasting the stock price index. IEEE International Conference on Intelligent Processing Systems, 1997, pp. 401-403.

[28] Taguchi, G.: Taguchi Methods, Research and Development. American Suppliers Institute Press, 1991.

[29] Taguchi, G.: Taguchi Methods, Signal-to-Noise Ratio for Quality Evaluation. American Suppliers Institute Press, 1991.

[30] Box, G.E.P., Meyer, R.D.: Dispersion Effects from Fractional Designs. Technometrics, **28(1)**, 1986, pp. 19-27.

[31] Nair, V.N.: Testing in Industrial Experiments with Ordered Categorical Data. Technometrics, **28(4)**, 1986, pp. 283-291.

[32] Leon, R.V., Shoemaker, A.C., Kacker, R.N.: Performance Measures Independent of Adjustment. Technometrics, **29**, 1987, pp. 253-265.

[33] Montgomery, D.C.: Design and analysis of experiments. New York: John Wiley, 2005.

[34] Berger, P.D., Maurer, R.E.: Experimental Design with Applications in Management, Engineering, and the Sciences, 2002.

[35] Box, G.E.P, Jenkins, G.M.: Time series analysis: forecasting and control. San Francisco: Holden-Day, 1976.

[36] Hwarng, H.B., Ang, H.T.: A simple neural network for ARMA(p,q) time series. Omega, **29(4)**, 2001, pp. 319–333.

389

[37] Kaastra, I, Boyd M.: Designing a neural network for forecasting financial and economic time-series. Neurocomputing, **10**, 1996, pp. 215–36.

[38] Zhang, G., Hu M.Y.: Neural network forecasting of the British pound/US dollar exchange rate. Omega, The International Journal of Management Science, **26**, pp. 495–506.

[39] Yao, J., Li, Y., Tan, C.L.: Option price forecasting using neural networks. Omega, The International Journal of Management Science, **28**, 2000, pp. 455–66.

[40] Chen, M.Y., Chen, D.R., Fan, M.H., Huang, T.Y.: International transmission of stock market movements: An adaptive neuro-fuzzy inference system for analysis of TAIEX forecasting. Neural Computing and Applications, DOI: 10.1007/s00521-013-1461-4.

[41] Balestrassi, P.P., Popova, E., Paiva, A.P., Marangon Lima, J.W.: Design of experiments on neural network's training for nonlinear time series forecasting. Neurocomputing, **72(4–6)**, 2009, pp. 1160–1178.

[42] Hsieh, L.F., Hsieh, S.C., Tai, P.H.: Enhanced stock price variation prediction via DOE and BPNN-based optimization. Expert Systems with Applications, **38(11)**, 2011, pp. 14178–14184.

[43] Harrell, F. E.: Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag, 2001.

[44] Chen, M.Y.: A Hybrid Model for Business Failure Prediction- Utilization of Particle Swarm Optimization and Support Vector Machines. Neural Network World, **21(2)**, 2011, pp. 129-152.

[45] Chen, M.Y.: Using a Hybrid Evolution Approach to Forecast Financial Failures for Taiwan Listed Companies. Quantitative Finance, 2011. DOI: 10.1080/14697688.2011.618458.

[46] Chen, M.Y.: Bankruptcy Prediction in Firms with Statistical and Intelligent Techniques and a comparison of Evolutionary Computation Approaches. Computers and Mathematics with Applications, **62(12)**, 2011, pp. 4514-4524.

[47] Hush, D.R., Horne, B.G.: Progress in supervised neural networks. IEEE Signal Processing Magazine, 1993, pp. 8-39.

[48] Cheng, C.S., Tseng, C.A.: Neural network in detecting the change of process mean value and variance. Journal of the Chinese Institute Industry Engineers, **12**, 1995, pp. 215-223.

[49] Haykin, S.: Neural Networks: A comprehensive foundation. Prentice Hall, Ontario, Canada, 1999.

[50] Geman, S., Bienenstock, E., Doursat, R.: Neural Networks and the Bias/Variance Dilemma. Neural Computation, **4**, 1992, pp. 1-58.

[51] Blum, A.: Neural Networks in C++. New York: Wiley, 1992.

[52] Swingler, K.: Applying Neural Networks: A Practical Guide. London: Academic Press, 1996.

[53] Saarinen, S., Bramley, R., Cybenko, G.: Ill-conditioning in neural network training problems. Siam J. of Scientific Computing, **14**, 1993, pp. 693-714.

[54] Sarle, W.S.: Ill-Conditioning in Neural Networks. USA: SAS Institute Inc., 1999.

[55] Sheela, K.G., Deepa, S.N.: Analysis of computing algorithm using momentum in neural networks. Journal of Computing, **3(6)**, 2011, pp. 163-166.

[56] Chan, Y.H.: Biostatistics 104: Correlational Analysis. Singapore Medical Journal, **44(12)**, 2003, pp. 614-619.

[57] Salzberg, S.L.: On comparing classifiers: Pitfalls to avoid and a recommended approach. Data Mining and Knowledge Discovery, **1**, 1997, pp. 317–327.

Appendix A. Financial indicators

No.	Code	Variables
1	X1	Fixed assets ratio
2	X2	Fixed assets to equity ratio
3	X3	Fixed assets to long-term debt ratio
4	X4	Fixed assets to permanent capital ratio
5	X5	Debt ratio
6	X6	Long term funds to fixed assets ratio
7	X7	Shareholders' equity to debt ratio
8	X8	Shareholders' equity to long term debt ratio
9	X9	Capital to total assets ratio
10	X10	Net worth ratio
11	X11	Long term debt to net worth ratio
12	X12	Fixed assets to net worth ratio
13	X13	Gearing ratio
14	X14	Current ratio
15	X15	Quick ratio
16	X16	Cash to current asset ratio
17	X17	Cash to current liability ratio
18	X18	Return on capital employed to current asset ratio
19	X19	Short-term borrowing to current asset ratio
20	X20	Account payables turnover rate
21	X21	Payable turnover rate
22	X22	Receivables turnover rate
23	X23	Inventory turnover rate
24	X24	Fixed assets turnover rate
25	X25	Total assets turnover rate
26	X26	Net worth turnover rate
27	X27	Operation income to capital ratio
28	X28	Pre-tax profit to paid-in capital ratio
29	X29	Gross margin ratio
30	X30	Operating expense ratio
31	X31	Operating profit margin ratio
32	X32	Operating profit margin (minus interest expenses) ratio
33	X33	Net income before tax ratio
34	X34	Net profit ratio
35	X35	Pre-tax rate of return on equity ratio
36	X36	After-tax rate of return on equity
37	X37	Pre-tax rate of return on assets
38	X38	After-tax rate of return on assets
39	X39	Return on fixed assets ratio
40	X40	Interest expenses to operating income ratio
41	X41	Interest protection multiples ratio
42	X42	Sales seasonal variation rate

43	X43	Sales growth rate
44	X44	Total assets growth rate
45	X45	Net worth growth rate
46	X46	Fix asset growth rate
47	X47	Cash flow ratio
48	X48	Operating income seasonal variation rate
49	X49	Operating income growth rate
50	X50	Pre-tax net profit margin variation rate
51	X51	Pre-tax net profit margin growth rate
52	X52	After-tax net profit margin variation rate
53	X53	After -tax net profit margin growth rate
54	X54	Price-to-book

Appendix B. America Banks Lists

Non-Financial Distress Banks Name	Banks Abbr. or Code
American National Bankshares Inc.	AMNB
Annapolis Bancorp Inc.	ANNB
Atlantic Bancshares Inc	ATBA.OB
Atlantic Southern Financial Group, Inc.	ASFN
BancorpSouth, Inc.	BXS
Bank of America Corporation	BAC
Bank of Florida Corporation	BOFL
Bridge Bancorp, Inc.	BDGE
Capital One Financial Corp.	COF
Capitol Federal Financial	CFFN
Citizens Republic Bancorp, Inc	CRBC
Comerica Incorporated	CMA
Commerce Bancshares Inc.	CBSH
East West Bancorp, Inc.	EWBC
First Bancorp	FBNC
First Citizens Bancshares Inc.	FCNCA
First Horizon National Corp.	FHN
Bank of the Ozarks, Inc.	OZRK
Webster Financial Corp.	WBS
Whitney Holding Corp.	WTNY
Financial Distress Banks Name	**Banks Abbr. or Code**
American International Group, Inc	AIG
Community Bankers Trust Corporation	BTC
CIT Group, Inc.	CIT
Dun & Bradstreet Corp	DNB
Farmers Capital Bank Corp.	FFKT
First National Bank Alaska	FNBA
First Trust Bank	FTB

Freddie Mac	FM
Horizon Financial Corp.	HFC
IndyMac Bancorp Inc	IMB
Lehman Brothers Holdings Inc	LBH
Mechanics Bank	MB
Michigan Heritage Bancorp Inc.	MHB
National Bank of Greece SA	NBG
Rainier Pacific Financial Group Inc.	RPF
Signature Bank	SBG
Silver State Bancorp	SSB
Sunwest Bank	SB
The Toronto-Dominion Bank	TTDB
Virginia National Bank	VNB