Forecasting Time Series using Neural Network Model: Application for Korean Stock Market

Young Geun Shin, Sang Sung Park, Dong Sik Jang

Division of Information Management Engineering, Korea University
(toctop@korea.ac.kr, hanyul@korea.ac.kr, jang@korea.ac.kr)

Abstract

This study examines the performance of forecasting model established by using neural network. The neural network model proposed in this study is implemented to forecast the stock price stream of Korean stock market. Factors for forecasting that influence Korean stock market are obtained by the way based on multiple regression analysis. As a result of the analysis, Korean stock market is greatly influenced by the stock market of the USA. It is enough to be called comovement phenomenon. In addition, the interest rate of Korea has a great effect on Korean stock market as well. So, we established the forecasting model of Korea Composite Stock Price Index (KOSPI) using neural network with Back-Propagation (BP) algorithm. Experimental results show that the proposed model has a good performance in forecasting KOSPI index.

1. Introduction

According to globalization of all the world, the economic situation of one country is influenced strongly by the economy of other countries. Korean stock market is receiving a lot of effects of oil price, interest rate of Korea and interest rate of the USA as well as the stock market of the USA, Japan, Europe etc.. Especially the stock market of the USA and Japan has a great effect on Korean stock market. This present state is more notable after foreign exchange crisis of Korea. That is, the motion of Korea Composite Stock Price Index (KOSPI) was shown excessive comovement phenomenon about the stock index of the USA, Japan. It is to be excessively to speak that is phenomenon by globalization of financial market simply. Also, sensitive response about stock index of countries that give great influences to domestic industry of Korea is serious. Interest rate fluctuation of the early 2000s for an economic turnaround of the USA had considerable effects on Korean stock market as well as the stock market of the USA. According to the progress of the opening of financial market after foreign exchange crisis of Korea, Korean stock market came in direct influence of the stock market of the USA. Thus, there are some causes that Korean stock market is seriously affected by the stock market of the USA. First, because Korean economy was linked closely with the U.S. economy through trade and investment, information connected with the U.S. economy has effects on Korean economy justly. Also, the progress of the opening of financial market makes the investor of the stock market respond more sensitively to the U.S. market change. Then correlation of per share earnings ratio between Korean stock market and the stock market of the USA rises. If foreign investors who influence significantly to Korean stock market trade stocks in Korean stock market similar dealing pattern doing in the stock market of the USA, Possibility to happen comovement phenomenon that the motion of Korean stock market follows the motion of the stock market of the USA will increases. Second, according to globalization of all the world industry such as FTA, a market infection phenomenon that Price fluctuation in one market is influenced over the acceptable level as a link through basic economic condition by price fluctuation in other markets presents. There is a long-term connection among the stock markets of the USA, Japan and Korea due to above-mentioned effects. Korean stock market shows a phenomenon that is following the stock market of the USA and Japan in the long term.[1] Various statistics techniques are used extensively to forecast a stream of data. There are MDA, Logit Analysis, Regression Analysis etc. in these statistical technique. Data that has changes with the passage of time can be modeled by using Stochastic Process and predicted. Data that have changes with the passage of time can be divided into stationary, nonstationary time series according to the feature of data and different estimate method is used in each case. Data are considered as stationary time series data if it satisfies the following contents. Otherwise it is considered as nonstationary time series data.[2]

a. When variable has the property which intends to recur to poise with the passage of time

b. The case that satisfies under the three conditions

- When average is regular
- When variance exists and is regular
- When covariance between variable that exist between two visual points depends on a time difference than a point of time.

Model of White Noise (WN), Autoregressive process (AR), Moving Average process (MA), AutoRegressive Moving Average process (ARMA) etc. are used for prediction which considers stationary time series data. But most time series data are nonstationary. nonstationary time series data have a probability structure which changes with the passage of time. Especially this structure has a feature that a shape of average and covariance changes with the passage of time. It is difficult to forecast future value using observed value for nonstationary of time series data. Because the prediction model is developed on the assumption that data are stationary. Therefore, more generalized model that can explain nonstationary which average or variance has is required for the analysis of nonstationary time series data. Most times series data includes some Regression Trend or some Seasonal Variation from seasonal factors. For normalization of these nonstationary time series data, regression trend or seasonal variation is removed through difference equation or seasonal difference equation. But it can be more effective to include these difference process and effect in the prediction model. A model used for this method is AutoRegressive Integrated Moving Average process (ARIMA) etc.. But this model has to be complemented in detail to apply the prediction of stock market because of the lacks of analysis for assumption and constraint. Because above-mentioned various statistics techniques express the prediction model through the linear combination of independent variables, it is difficult to express and predict the data which happen in the real world. To complement this weak point, various prediction cases that use Neural Network technique are been researching recently. Neural Network is a inference model that link each other of the digital neurons similar to the neurons that compose a human's brain.[3] Neural network can improve its own function according to change of data and circumstance using own training function. And because neural network is not assuming the linear combination of variables or certain probability distribution, adaptation ability about actual data is greater than other prediction methods. So this study concerns on predicting time series data using neural network. The stock market index of the USA, Korea and the interest rate of Korea, the USA are used as the time series data in this paper.

The remainder of this paper is organized as follows. Section 2 describes regression analysis and neural network used for predicting the stock index. The process of selecting factors that have an effect on the index is described in Section 3. Section 4 describes neural network model that forecast KOSPI index using factors that have an effect on the index. Results of empirical investigation through the proposed method are described in Section 5. Finally, Section 6 offers concluding remarks.

2. Methodology

2.1 Regression analysis

Regression analysis is a statistical technique to investigate relationships between variables. it is used to search the influencing power that independent variable has an effect on dependent variable or to predict change of dependent variable accompanied by change of independent variable. Among variables that have a causal relationship mutually, independent variable that influences any other variable is called explanatory variable or predictor variable. And dependent variable that is influenced by other variable is called effect variable.

Regression analysis is a statistical analytic method that can make out a causal relationship between two variables, and then forecast change of other variable from one variable. Most the structure of causal relationship has been related sophisticatedly for several factors. In this paper, we execute a multiple regression analysis because of the being of various factors that influence KOSPI index. Multiple regression analysis is a method that allows additional factors to input the analysis separately so as to estimate the effect of each factor. To quantify the impact of various simultaneous effects on a single dependent variable is important and useful. A structure of general multiple regression may be written in Eq.(1).

$$y = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \tag{1}$$

The variable y is termed the dependent variable, and the variable $x_1, x_2, \cdots, x_p$ are termed the independent variable; α is the constant term and $\beta_1, \beta_{2,} \cdots, \beta_p$ are the coefficient of the variable x_i. There are some assumptions in using multiple

regression analysis

(1) The dependent variable is assumed to be a random variable.

(2) The independent variables $x_1, x_2, \cdots, x_p$ are a fixed constant.

(3) There should be linear relationship between the independent variable and the dependent variable.

(4) Error term follows normal distribution.

2.2 Neural network model

Neural network model is being used in the analysis of financial time series as they move from simple pattern recognition to a diverse range of application areas.[4] Neural network model is a information processing system that is composed of node, link, activity function and training algorithm etc.. Usually, the structure of neural network model is built of input node, hidden node and output node. Input node receives data from the outside of system and transfers these data into system. Hidden node in the system has produced quickly the result value after receiving transferred value from input node and treating. Then output node computes the result value of system based on both input value and system state. The computed result values are compared to target values. Then error values are determined. Error values are used as inputs to modify the weights between layers in a backward direction. That is, the link in a forward direction is used for the operational phase and the link in a backward direction is used for the learning phase. The process of neural network with BP algorithm is illustrated in figure 1.

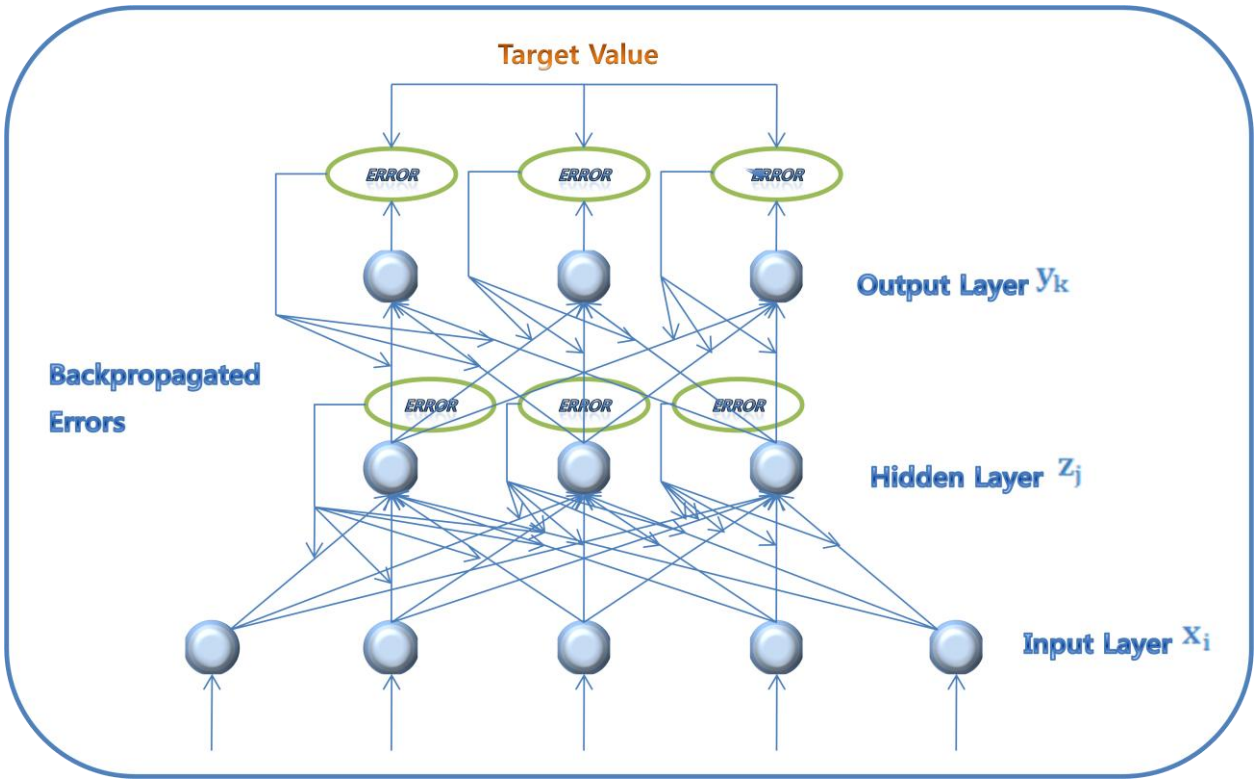

Fig. 1 Process of neural network with BP

(1) Structure of neural network model

The basic structure of neural network consists of a layer of input nodes $x = (x_1, x_2, \cdots, x_i)$ linked to a layer or more of hidden nodes $z = (z_1, z_2, \cdots, z_j)$, which are linked to a layer of output nodes $y = (y_1, y_2, \cdots, y_k)$ in turn. Then neural network model is formularized as follows.

$$z_j = f_j(\alpha_j + \sum_{n=1}^{i} w_{nj} x_n) \tag{2}$$

$$y_k = f_k(\alpha_k + \sum_{n=1}^{j} w_{nk} z_n) \tag{3}$$

α_j, α_k are the bias parameter and w_{nj}, w_{nk} are the weight. The transferring function maps from input node to hidden node. Hidden node estimates weighted sum of the transferring value from input node and the weight. Then it converts the weighted value into nonlinear value using function f_j and transfers to output node. The activation function f_j used at this time is the

hyperbolic tangent function. The most popular sigmoid function for neural network is the hyperbolic tangent function. The tanh() function is a sigmoidal-shaped function and has the following symmetric shape.

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{4}$$

(2) **Training of neural network model**

It is essential to train a network for the success of neural network. Process that is converting initial weight value into the value which is suitable to data is proceeding in neural network model. Eq.(5) is used to calculate the error between output value calculated using activation function and target value.

$$e = \frac{\sum_{n=1}^{k} (y_n - t_n)^2}{2} \tag{5}$$

The BP algorithm is used in this paper. BP algorithm is composed of two phases. the forward phase where activations are propagated from input node to output node and the backward phase where the error between output value and target value in a layer of output node are propagated backwards to modify the weight and the bias parameter.

3. Analysis of influencing factor to KOSPI index

Korean stock market is influenced by various factors. Therefore KOSPI index has fairly a wide range of volatility. Factors that affect KOSPI index are NYSE composite index of the USA, interest rate of the USA and Korea. Figure 2, 3, 4, 5 shows an undulating trend of the data used in multiple regression analysis

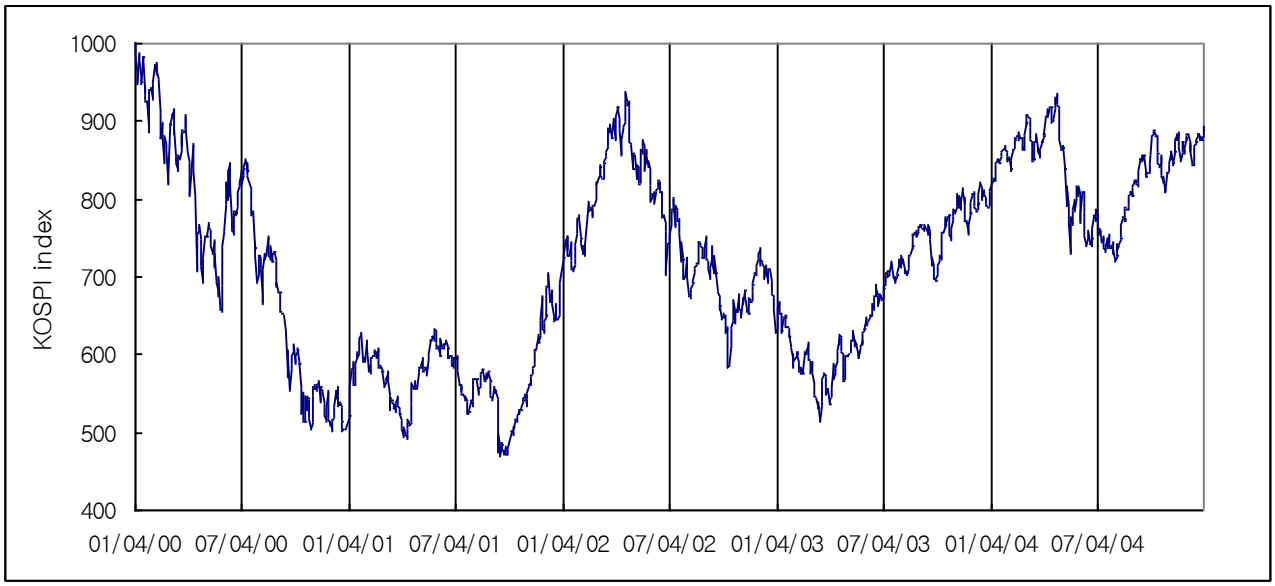

Fig. 2 KOSPI index

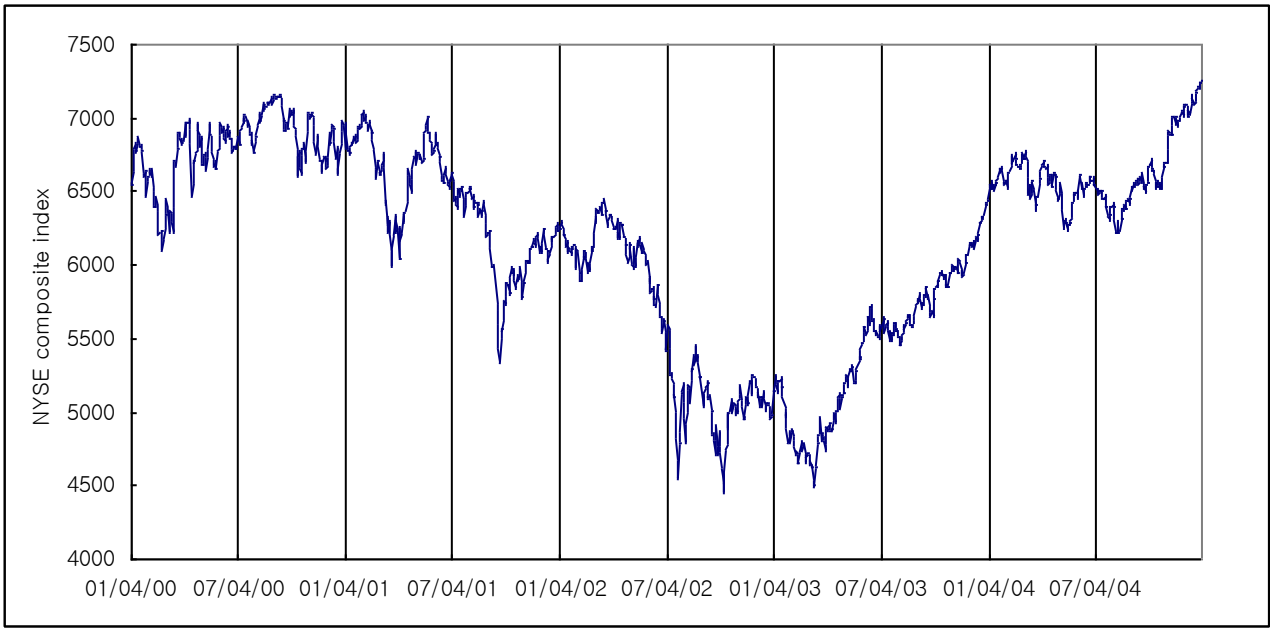

Fig. 3 NYSE composite index

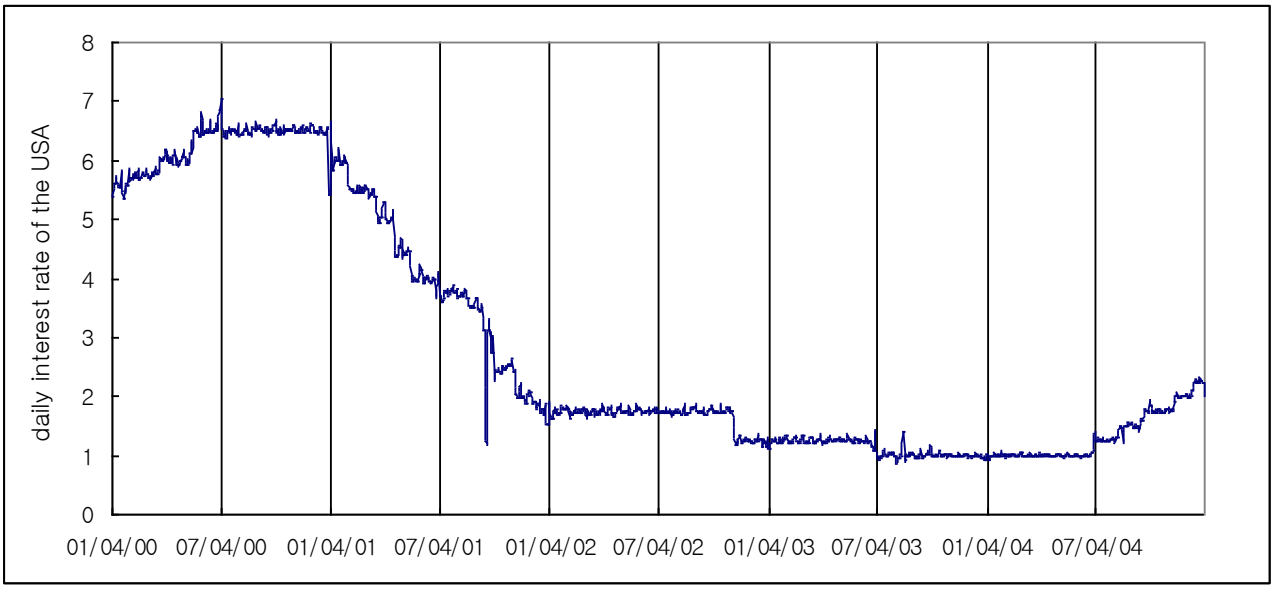

Fig. 4 daily interest rate of the USA

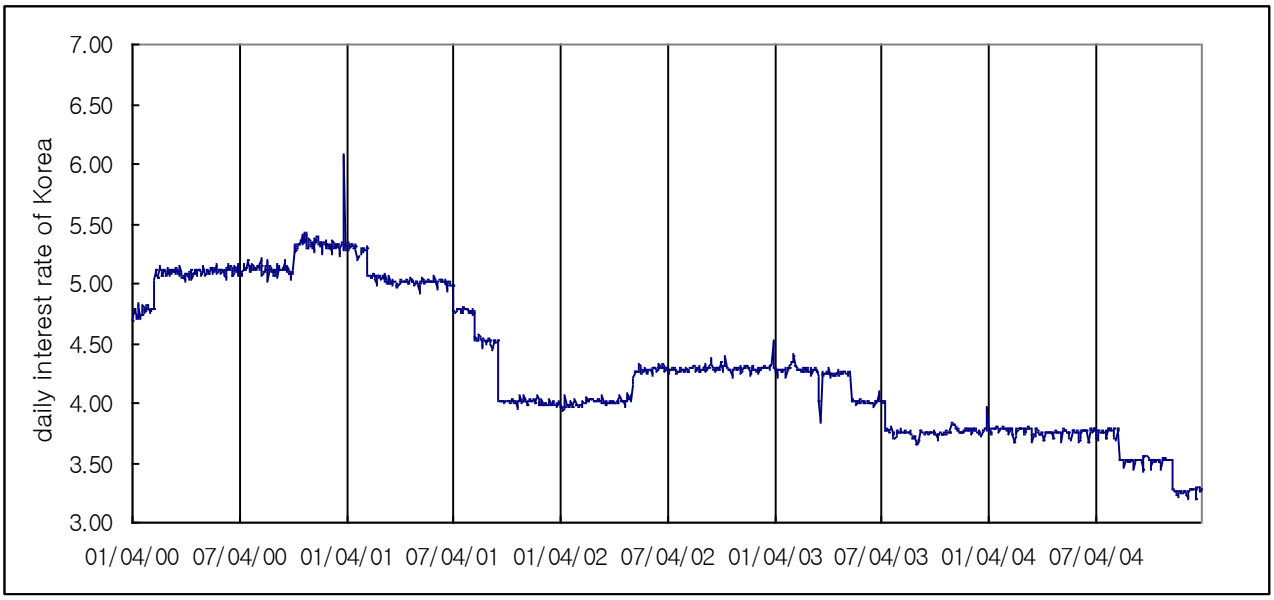

Fig. 5 daily interest rate of Korea

We analyzed using multiple regression model whether these factors that influence Korean stock market are significant, Data are taken from daily value in the period of Jan 1, 2000 until Dec 31, 2004 in considering that Korean stock market has been influencing by various factors for changing of financial market circumstance from the progress of the opening of financial market after foreign exchange crisis of Korea. We use NYSE Composite Index as Data of the stock market of the USA and FRB's Federal funds (effective) as data of the interest rate of the USA. Call rate (value tomorrow) data are used as the interest of Korea and closing KOSPI indices are used as data of Korean stock market. Experiment result is shown in table 1.

Table 1 Effect of each factor

Predictor variable	Coefficient	SE Coefficient	T	P	VIF
Constant	984.68	88.09	11.18	0.000	
interest rate of Korea	-157.47	15.11	-10.42	0.000	8.9
NYSE composite index	0.059311	0.007543	7.86	0.000	2.6
interest rate of the USA	15.798	5.127	3.08	0.002	12.3
S = 100.869, R-Sq = 35.6%, R-Sq(adj) = 35.4%					

When independent variables in multiple regression model are more highly correlated with other independent variables than with the dependent variable, multicollinearity occurs. If it is important to estimate the contributions of individual independent variable, multicollinearity can be a problem in multiple regression model. Because multicollinearity inflates the variances of the parameter estimated. This may lead to a lack of statistical significance of individual independent variables even though the overall model may be significant. If VIF (Variance Inflation Factor) value is more than 10 generally, it is often regarded as indicating the feature of multicollinearity.[5] VIF value of the interest rate of the USA is 12.6 as we may see in table 1. For this reason, the interest rate of the USA is removed as the independent data of this multiple regression model. Then regression analysis is performed after removing the outliers of independent variables. Experiment result is shown in table 2.

Table 2 Effect of each factor (after removing the outliers)

Predictor variable	Coefficient	SE Coefficient	T	P	VIF
Constant	1025.36	12.66	80.99	0.000	
interest rate of Korea	-160.252	1.901	-84.29	0.000	1.0
NYSE composite index	0.054661	0.001851	29.54	0.000	1.0
S = 30.4708 R-Sq = 92.4% R-Sq(adj) = 92.4%					

VIF value of the interest rate of Korea and NYSE composite index are 1.0 as we may see in table 2. So these variables are regarded as not having multicollinearity. R-square (adj.) value of this regression model is 92.4%. This value indicates that explanation ability of the proposed regression model is fairly high.

4. Forecasting KOSPI index using neural network model

Through above-mentioned experimental results of multiple regression analysis, the factors that influences KOSPI index significantly are selected. Prediction model is developed by using the factors as variables of neural network model. The stock market of the USA opens at 9:30 A.M. and closes at 4:00 P.M. in local time. It is 11:30 P.M. – 6:00 A.M. for Korean standard time. So it is possible that Korean stock market is influenced by the stock market of the USA of previous day. Also it is possible that the interest rate of Korea affects Korean stock market. Because call rate is determined once a month, neural network model works best when all the input and output values are between 0 and 1.[6] So all data are converted to the values between 0 and 1. Data of the period Jan 1, 2000-Dec 31, 2003 is used as the training data for the neural network model. And data of the period Jan 1, 2004-Dec 31, 2004 is used as the testing data. Then accurate rate for forecasting KOSPI index of the proposed model is evaluated. Generally two kinds of statistical error terms have been used for measuring result of neural network model. These are Mean Absolute Error (MAE), Root Mean Square Error (RMSE). MAE expresses the linear relationship between the errors of different samples. It is defined as following Eq.(6).

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}}|\mathrm{y_k} - \mathrm{t_k}| \quad (6)$$

y_k are the output values and t_k are the target values. N is the total number of samples. RMSE is defined as the value that expresses accurate rate of model and as following Eq.(7)

$$\mathrm{RMS} = \frac{1}{\mathrm{N}}\sqrt{\sum_{\mathrm{k}=1}^{\mathrm{N}}(\mathrm{y_k} - \mathrm{t_k})^2} \quad (7)$$

These two criterions are used for evaluating the proposed model in this paper.

5. Results

In forecasting time series data, a major problem is establishing the optimal number of nodes on each layer. The most common method in determining the number of hidden nodes is by trial-and-error or by means of experiments. So the optimal number of hidden nodes in this proposed model is selected by means of empirical testing. Usually, it is recommended to start with one hidden layer.[7] If the results are not good, the number of hidden layers would rise. Also it is difficult and important to determine the number of hidden node. If too many hidden nodes are used, the number of connections grows up too big. Then the neural network model can fit to the training data and the accurate rate of the training data will be good. But the neural network model to the test data will be poor. So to gauge the best number of hidden node is important. In this paper, we performed experiments to find the number of hidden node appropriate to the proposed model. Through performing the experiments repeatedly, the number of hidden node: 5 and the learning rate coefficient η: 0.3 is the best value to the proposed model. The results that are performed for 100 epochs are shown in table3.

Table 3 the appropriate number of hidden nodes

The number of hidden nodes	Learning rate	MAE	RMS
3	0.3	0.0982	0.0878
	0.5	0.0935	0.0884
5	0.3	0.0700	0.0671
	0.5	0.0830	0.0811
7	0.3	0.0929	0.0900
	0.5	0.0926	0.0915

When the model that have both 5 hidden nodes and 0.3 learning rate is performed, the accurate rate was best. RMS error of the model is 0.0671 at 0.3 learning rate and 0.0811 at 0.5 learning rate. These results indicate that the accurate rate of the proposed model is approximately 92%.

6. Conclusion and future work

The paper examines the factors that influences KOSPI index significantly and Evaluates neural network model that forecast KOSPI index using these factors. As a result of multiple regression analysis, KOSPI index is influenced significantly by call rate of Korea and NYSE composite index of the USA. R-square (adj.) 92.4% calculated from this regression model indicates that explanation ability of the proposed regression model is fairly high. Then KOSPI index is forecasted by applying neural network model that uses data extracted through multiple regression analysis as the variables of model. To demonstrate predictive power of the proposed model, we use the daily data in the period of Jan 1, 2000 until Dec 31, 2004. As a result of Experiments, the best accurate rate is shown in the case of including hidden nodes 5.

But the Durbin-Watson statistics used to detect the presence of auto correlation in the residuals from multiple regression analysis is 0.214141. Therefore selected variables as factors that influences KOSPI index in this paper are insufficient to explain the variation of KOSPI index totally. Future studies that forecast time series data, especially the price index of stocks, may use more influencing factors. These studies could lead to the developing of more effective

neural network model.

Acknowledgement

This work was supported by the Brain Korea 21 Project in 2007.

References

[1] Yoo,T.W., Kim,C.H.; Positive Analysis of the influence the stock market of USA and Japan on Korean stock market, Researches in stock and finance, Vol.3, No.1, 1997

[2] Lee,S.Y.; Principle on Analysis of time series, Freedom Academy, 2001

[3] Lee,G.C., Kim,M.J., and Kim,H.; An Inductive Learning-Assisted Neural Network Approach to Bankruptcy Prediction: Comparison with MDA, Inductive Learning, and Neural Network Model, Researches in business administration, Vol.23, No.4, 1994

[4] Patterson,D.W.; Artificial Neural Networks, Prentice-Hall, 1996

[5] Kang,K.S., Kim,C.R.; Regression Analysis, Kyowoo Publishing, 1999

[6] Berry,M., Linoff,G.; Data Mining Techniques, Wiley Computer Publish, 1997

[7] Hornik,K., Stinchombe,M.,White,H.,; Multilayer feedforward networks are universal approximators, Neural Networks , Vol. 2, 1990