FORECASTING THE TOTAL INDEX OF TEHRAN STOCK EXCHANGE

Ghodratollah EMAMVERDI[*]
Mohammad Sharif KARIMI[**]
Sima KHAKIE[***]
Mojtaba KARIMI[****]

Abstract

Development of financial markets and financial instruments is one of the pillars of economic growth and capital accumulation in any countries especially in the developing ones. Capital market works as a substantial tool for development and financing which attracts many investors for purchasing and selling capital. Hence they are seeking for forecasting events and situations of future till can invest with a lower uncertainty. On the other hand total index of stock exchanges accounted for a precursor one in economics which we expect following improvement in index of capital market, the economic growth will improve and we face condition of recession after its decrease. Thus the importance of exact prediction of this index is redoubled. In this paper we forecast the index by applying ARIMA and Neural Network and then we achieve the minimum and maximum of total index by using FARIMA. Indeed the main aim of this research is to imply this point that these three methods are complementary to each other in decision making of investors and economic policy makers.

Keywords: Forecasting, Stock price index, Tehran Stock Exchange, Fuzzy ARIMA, Neural Network.

JEL Classification: C45, C53, C58, G17.

[*] *Assistant professor of Economics, Islamic Azad University, Central Tehran Branch,* ghemamverdi@gmail.com

[**] *Assistant Professor, Department of Economics, Faculty of Social Sciences, Razi University, Kermanshah, Iran, s.karimi@razi.ac.ir*

[***] *Master of Science of Planning of Economic Systems, Islamic Azad University, Central Tehran Branch,* khakieh.sima@gmail.com

[****] *Master of Science of Planning of Economic Systems, Islamic Azad University, Central Tehran Branch,* tabakarimi62@gmail.com

1. Introduction

Physical capital accumulation is one of the necessary requirements for economic growth and through financial markets can accelerate the capital accumulation. Financial markets are important for every country especially the developing ones because of their contribution in gathering wanderer resources through savings in national economy, optimization and redirecting the flow of financial resources.

Stock market is one of the most important kinds of financial markets. Various factors such as political and economy conditions of the countries influence on the progress of the stock market. One of the critical issues for growing Stock market is to attract more capital and for absorbing more capital, sufficient conditions should be provided to encourage micro and macro investors to invest on Stock Exchange and one of these conditions is reducing the risk of investing. A main criterion for evaluation of the growth of Stock Exchange is the total index, and in case of prediction of total index the risk of buying and selling of shares will reduce and minimize for shareholders. Due to the importance of the prediction of total index which is accounted for as a leading indicator in economy, different methods are proposed that in this research three recognized of them are investigated. One of the prevalent methods is the prediction of time series model introduced by Box-Jenkins which is usually showed as ARIMA and this method is applied when definite, adequate and at least 100 or 150 data exist. A main hypothesis behind ARIMA is that future values of time series have a clear and specific functional relationship with the past and current values of time series and net errors of the model. Another method investigated in this study is Artificial Neural Network (ANN) Model which is a non-linear method and is appropriate for identifying complicated relationships between variables. Neurons are the smallest units of information processor that they are the performance basis of Neural Network. Finally the third method is Fuzzy Autoregressive Integrated Moving Average (FARIMA) used in uncertainty conditions when definite and adequate data does not exist. Indeed fuzzy method expresses the complexities and ambiguities in a mathematical form and provides the basis for an analysis to indefinite (fuzzy) variables of the environment. In many studies about prediction always comparison has been made between two models that one of them was ARIMA method; however, the aim of this research is to present a suitable prediction model for the total index of Tehran stock Exchange and also a maximum and minimum value of intended variable in considered period. According to this the concepts of ARIMA time series, Fuzzy ARIMA and Artificial Neural Network have been explained in next section the return on share price index of Stock Exchange predicted by these methods and finally results of the models have been analyzed.

2. Review of research

Time series forecasting is an important area of forecasting in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future. This modeling approach is particularly useful when little knowledge is available on the underlying data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables several models have been suggested for time series forecasting, that are generally divided to linear and nonlinear models.

One of the most important and widely used linear time series models is the Auto Regressive Integrated Moving Average (ARIMA) model that has enjoyed fruitful applications in fore casting social, economic, engineering, foreign exchange, and stock problems. Second class of time series forecasting is nonlinear time series models. Artificial neural networks are one of these models that are able to approximate various nonlinearities in the data and are flexible computing frameworks for modeling a broad range of nonlinear problems One significant advantage of the ANN models over other classes of nonlinear model is that ANNs are universal approximates, which can approximate a large class of functions with a high degree of accuracy. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data. Commonly used network models include multi-layer perceptron (MLP), Radial Basis Function (RBF), Probabilistic Neural Networks(PNNs) and General Regression Neural Networks(GRNNs) Single hidden layer feed forward network is the most widely used model form for time series modeling and forecasting accuracy is one of the most important factors to choose the forecasting method, and regardless numerous time series forecasting models, the accuracy of time series forecasting is fundamental to many decision processes and hence the research for improving and diagnosing the effectiveness of forecasting models has been never stopped.

First time Tseng (2001) applied the combination of Fuzzy and ARIMA models for prediction of dollar price against Taiwan currency. He could show this new method not only presents more exact prediction but also determines the best and the worst condition.

Shabani (2010) in a thesis with the subject of forecasting the price index of Tehran Stock Exchange has analyzed and compared the various linear and non-linear methods in prediction of share price index of stock exchange. In this research two patterns of using past values of share price and applying variables such as exchange rate and oil price accompanied by lagged terms of exchange index. The results indicate neural network model based on lagged total index has lower errors relative to GARCH.

Moreover neural network model with the inputs of lagged share price index, exchange rate and oil price had a better prediction in comparison to a neural network model with the only lagged share price index as an input. Monajjemi et al. (2009) in a paper with the title share price forecasting in stock exchange market by fuzzy neural network and genetic algorithms and its comparison to artificial neural network achieved to this result that share price prediction by fuzzy neural network and genetic algorithms reduce estimation error of share price relative to artificial neural network technique and it is a suitable method for forecasting. Mehrara et al. (2009) in a research with the subject of modeling and prediction of Tehran Stock Exchange Index and determining effective variables by neural network model obtained to this result that among macroeconomic variables in Iran, the variables such as cost of land, housing cost with two months lag, consumer price index, monetary base with a lag, lagged house rental rate and finally global price of crude oil with a lag are the effective and important variables in modeling of price index and cash return of Tehran Stock Exchange.

Thus among alternative assets in Iran's economy, it seems housing is a serious competitor for share in assets market. In other words transformations in the housing market could have an impact on stock market. In Relevance to FARIMA limited studies have been conducted in Iran, and among them we can point to Khashe'i and Bijari (2010) and Maleki (2010) on Exchange rate and global gold respectively.

3. Theoretical Foundations of ARIMA, Neural Network and FARIMA

In this research the prediction of stock price index will be performed by three methods of ARIMA, FARIMA and Neural Network and in this section advantages and disadvantages of each method will be reviewed.

Box-Jenkins (1976) is a conventional method in forecasting of economic variables. The basic hypothesis in this method is that current values of the variable is related to the past values and behavior of the variable in future can be predicted by its behavioral pattern in past. Indeed time series method is a basic method for other ones; however, this method has some defects too amongst its need to large number of data, about 100 to 150. Box-Jenkins method includes four steps:

1. Experimental identification of the model structure
2. Estimation of the unknown parameters of the model
3. Identifying the precision of fitting of the model
4. Forecasting by selected model (Johnston, 2009).

In ARIMA model it is assumed that the net error term is a stochastic variable with normal distribution of zero mean and δ^2variance and independent of the observations.

In this method by combination of two phenomena of Autoregressive and Moving Average, the ideal model is made. The purpose of autoregressive is dependence of the variable to its past values in previous periods. In Moving Average section, studied variable is a function of the moving average of the values of prediction errors in current and past periods which it is called white noise too. (Gujarati, 2009).

In classical methods of time series study like Box-Jenkins method and ARIMA it is assumed that the studied time series is produced through a linear process. (Houshmand et al., 2008).

From the advantages of this common method can point to its analysis capability with great details, high explaining power and easy deployment. ARIMA time series models are suitable for forecasting in a condition that data are sufficient and exact and they can present fine predictions; although, this method has some defects amongst the need to large number of data and also since in real environment information are always changing and considering uncertainty and Imprecise in real and unknown environments, it is not simple to predict with much data thus applying fuzzy method can solve this problem to some extent. Another method is the Artificial Neural Network which is based on determined lagged by ARIMA model and are estimated by using activity functions and the number of neurons in hidden layer and the error of prediction will be calculated. In fact this method is an appropriate supplementary for forecasting financial variables.

Artificial Neural Network models have different kinds; however, their total structure is similar and all have some components which are the same in all of them and it is pointed to as follows:

1. Layers

Input layers:

There are units in the number of explaining variables in this layer which everyone includes the data of one explaining variable.

Hidden layers:

There are units of information processor in this layer that have a very effective role on correct learning (estimation) of the model.

Output layer:

The processing of sent information from hidden layer will be performed in this layer. The number of units of this layer is as number of endogenous variables.

2. Activity (Transfer) Function

Every processor unit in hidden and output layers uses determined mathematic function for information processing who are called activity or transfer function.

3. Connectors

Different layers and neurons of neural network will be connected in order to send received information and signals to each other. Each of the connectors that connect the neurons has special weight indicating relative importance of the output of every neuron in calculating the value of next neuron output. Actually these weights form network memory and the main aim of designing neural network is to estimate them which prediction is made based on these weights (Kia, 1983).

One of the important characteristic of neural networks is their ability in learning and generalization. Learning process (training) of a neural network is achieved through adjustment of weights according to the connections among neurons and learning process is performed in two ways of supervised and unsupervised (AsghariOskoi, 2002).

Although neural networks are not comparable with natural neural system, they have some features which it distinguishes them where they need to learn a linear or nonlinear mapping. These capabilities include: learning, capability to use as a corporate memory, storage and addressable memory, the ability to generalize and training, and the impairment, resilience and the tolerance of errors (Monajemi et al., 2009).

To resolve some other defects of ARIMA, Fuzzy ARIMA is proposed. This method is suitable for situations that data are not sufficient and also for understanding the minimum and maximum of the variable. Two time series models of ARIMA and Fuzzy ARIMA are complementary to each other and since results of FARIMA are presented in a range around real data, prediction errors are lower than ARIMA which the results are given as definitive numbers, because in definitive condition if the result was wrong, the error would be 100% while the error in fuzzy condition is a number between 0 and 1. Two basic parameters in FARIMA are *P* and *C* which are the values of parameters and their developed around the center respectively and generally model takes advantage of the minimizing total ambiguities (which is equal to the sum of single extensions of any fuzzy parameters of the model).

In this regression model, the target is to find fuzzy coefficients which minimize fuzzy output extent for all data set. Equation (1) represents the target function of regression model which should be minimized

$$Minimize \quad S=\sum_{i=1}^{p}\sum_{t=1}^{k}c_i\,|\phi_{ii}|\,|W_{t-i}|+\sum_{i=p+1}^{p+q}\sum_{t=1}^{k}c_i\,|\rho_{i-p}|\,|a_{t+p-i}| \tag{1}$$

$$subject\,to \quad \sum_{i=1}^{p}\alpha_i W_{t-i}+a_t-\sum_{i=p+1}^{p+q}\alpha_i a_{t+p-i}+(1+h)\left(\sum_{i=1}^{p}c_i|W_{t-i}|+\sum_{i=p+1}^{p+q}c_i|a_{t+p-i}|\right)\geq W_t \qquad t=1,2,..k$$

$$\sum_{i=1}^{p}\alpha_i W_{t-i}+a_t-\sum_{i=p+1}^{p+q}\alpha_i a_{t+p-i}-(1+h)\left(\sum_{i=1}^{p}c_i|W_{t-i}|+\sum_{i=p+1}^{p+q}c_i|a_{t+p-i}|\right)\leq W_t \qquad t=1,2,..k$$

$$c_i\geq 0 \qquad for\; i=1,2,...,p+q$$

Since there are two restrictions for every data set, while there are *m* numbers of data sets, there are *2m* restrictions (Khashei and Bijari, 2007). From the advantages of Fuzzy ARIMA is that we can use the benefits of both methods of ARIMA and Regression Fuzzy simultaneously that include:

1) Determine the best and worst possible positions for correct and more accurate decisions.

2) The need to fewer observations in comparison to ARIMA model.

4. Data and empirical results

The applied variable for prediction of the return on share price index, is the return on share price index in Tehran Stock Exchange over the weekly period of March 22, 2004 to January 18, 2012 and it is totally 147 data.

ARIMA model:

Firstly the ARIMA model is identified as ARIMA (1,1,2) that the coefficients in this model are as below:

$$D\left(LOG\ (y_t)\right) = 0.008101 - 0.861672y_{t-1} + 1.160113\varepsilon_{t-1} + 0.344865\varepsilon_{t-2} \quad (2)$$

Artificial Neural Network model:

According to estimated ARIMA model in previous section and significance of the coefficients of first order autoregressive, network input is the return on share price index and network output is the return on share price index.

In neural network model, part of the data is considered as training set and others as results test. If performance in test set was weak, network combination or training parameters should change and as long as performance is satisfactory, the network is trained. In this research 80 percentages of data are used for training and the rest for testing the results.

The number of unit of information processors usually is determined regarding to try and by error method and the status leads to the minimum value in evaluation criteria such as Mean Square Error (MSE) and Root Mean Square Error (RMSE). In this research, after testing various numbers of units of information processors and network training, finally the number of units of information processors in considered.

Following the network structure design, input and output variables are introduced to the network and 80 percentages of data (118 data) are used in test section. Finally prediction of share price index is done. Results of models of time series and Artificial Neural Network are as following. Table (1) shows that every four evaluation index in Artificial Neural Network are lower than ARIMA(1,1,2);

therefore prediction power of ANN method is more than ARIMA(1,1,2) in this research.

Table 1

Results of Predictions by Models of Time Series and Artificial Neural Network

U-THEIL	*MAE*	*MSE*	*RMSE*	Model
0.591748	0.016221	0.000428	0.020684	*ARIMA(1,1,2)*
0.529194	0.013210	0.000304	0.017436	*ANN*

Source: finding of study

In addition to the above four criteria, to ensure correct selection of model, two tests of Granger-Newbold (1976) and Diebold-Mariano (1995) are performed. Since it is important to assess the accuracy of economic forecasts, spread of statistical methods for comparing different methods of prediction accuracy is essential. The zero hypotheses is based on the equality of forecast accuracy based on two different methods to predict one variable (Hearvy, 1997). The base of these statistical methods is on statistical prediction errors and fundamental assumptions of this approach include:

a) Forecast errors are normally distributed with zero mean

b) Forecast errors are not correlated

c) Forecast errors are contemporaneously uncorrelated (Anders, 2010)

The first step for Granger-Newbold test is to form the sequence of forecast errors of ARIMA and ANN models which e_{1t} is the prediction error of ARIMA and e_{2t} is that of neural network model.

Now by using sequences of forecast errors of two models, variables of $x_t = e_{1t} + e_{2t}$ and $z_t = e_{1t} - e_{2t}$ are formed. Statistic of this test is as below:

$$t = \frac{r_{xz}}{\sqrt{\frac{(1-r_{xz}^2}{(H-1)}}} \quad (3)$$

In equation (3) r_{xz} indicates correlation between z_t and x_t H is the number of observations. The value of test statistic obtained from

Granger-Newbold test equals to 1.11 which in confidence level of $\alpha = 0.05$ is not significant; hence H_0 hypothesis cannot be rejected.

Diebold-Mariano test is another method for comparing the prediction power of models and is performed when 3 hypotheses of three above hypotheses are violated. For calculating statistic of Diebold-Mariano test initially with assumption that by increasing the size of error, losses resulting from the prediction error grows at a rapid pace, so we consider it once in second and once in fourth power of prediction errors and calculate the sequence of d_t.

$$d_t = e_{1t}^2 - e_{2t}^2 \tag{4}$$

In this step we test autocorrelation among sequence components of d_t using Liang-Box statistic and because there is no autocorrelation among components of d_t sequence, DM-statistic is calculated through $t = \frac{\bar{d}}{\sqrt{\operatorname{var}(\bar{d})}}$

The result of Liang-Box test shows that d_t sequence or the same squared forecast errors of neural network and ARIMA does not have autocorrelation. The mean of d_t sequence or in other words

$\bar{d}$ equals to 6.54803×10^{-5} and variance estimation is 7.76442×10^{-8}.

The value of statistic is 1.265 which at the confidence level of 0.05 is not significant, thus H_0 hypothesis cannot be rejected.

Now we consider d_t sequence in fourth power of prediction errors that is calculated as following:

$$d_t = e_{1t}^4 - e_{2t}^4 \tag{5}$$

Here also Liang-Box test shows d_t sequence has no autocorrelation, thus we calculated mean, variance and DM-statistic as before.

The mean of d_t sequence is 1.87946×10^{-7}, variance is 8.461×10^{-13} and DM is 1.1 which it is not significant at the confidence level of 0.05 therefore H_0 hypothesis cannot be rejected. Thus the prediction power of both models is the same and does not have any difference.

Results of FARIMA forecast that is explained in previous section, is as follows for share price index of Tehran Stock Exchange over the considered period:

Step 1: for creating FARIMA model, at first the last 50 data among 150 data which are applied in time series and neural network prediction are separated and then the coefficients of this model comes in the form of fuzzy. ARIMA model for 50 data is as following:

$$r_t = -0.0005 + 0.798018 r_{t-1} - 0.9612\varepsilon_{t-1} \tag{6}$$

Step 2: we acquire the domain of coefficients by writing the limitations (restrictions) and minimizing the target function as:

$$MinS = \sum_{i=1}^{p}\sum_{t=1}^{k} C_i\,|w_{t-i}||\varphi_{ii}| + \sum_{i=p+1}^{p+q}\sum_{t=1}^{k} C_i|\rho_{i-p}||\,\alpha_{t+p-i}| \tag{7}$$

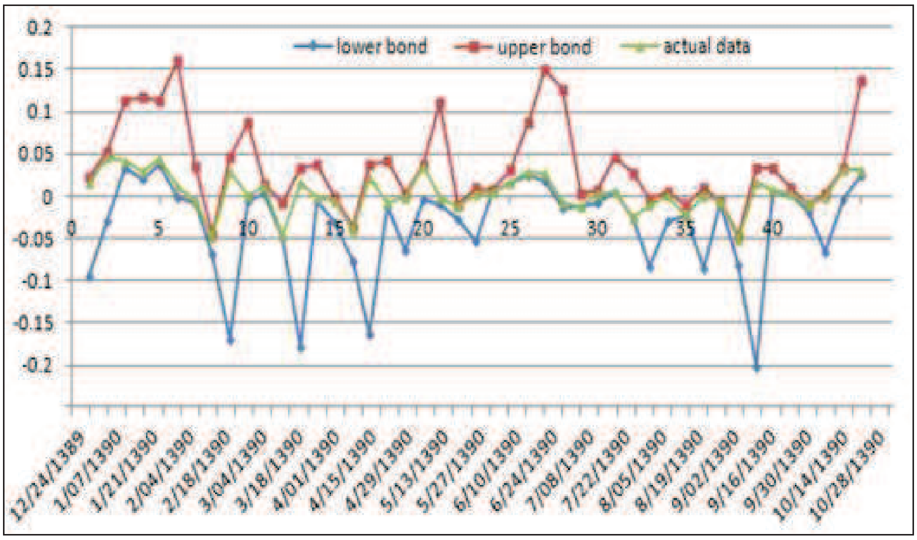
$$MinS = 0.115C_1 + 0.104C_2 \tag{8}$$

$$D\big(LOG(INDEX)\big) = -0.0005 + 0.798{,}0.18y_{t-1}(-0.962{,}2)\varepsilon_{t-1} \tag{9}$$

FARIMA model produces a regression upper than real data regression and a regression lower than that which they are known as upper bound and lower bound. These two bound create a domain that the real data regression will place between them. Upper and lower bound of FARIMA and real data models have been plotted.

Figure 1

Real Data and Upper and Lower Bound of FARIMA



Source: finding of study

Advantage of using this model on determining these two bounds is to specify the best and worst possible conditions for decision makers in financial markets. Also more reliable decision field would be created rather than non-fuzzy regressions, because prediction is fuzzy instead of zero and one prediction.

Step 3: neglecting those data that have placed on upper and lower bounds, 24^{th}, 25^{th} and 41^{th} data had this condition and by removing them, below results obtained:

$$r_t = -0.0005 + 0.798r_{t-1} - 0.9622\varepsilon_{t-1} \tag{10}$$

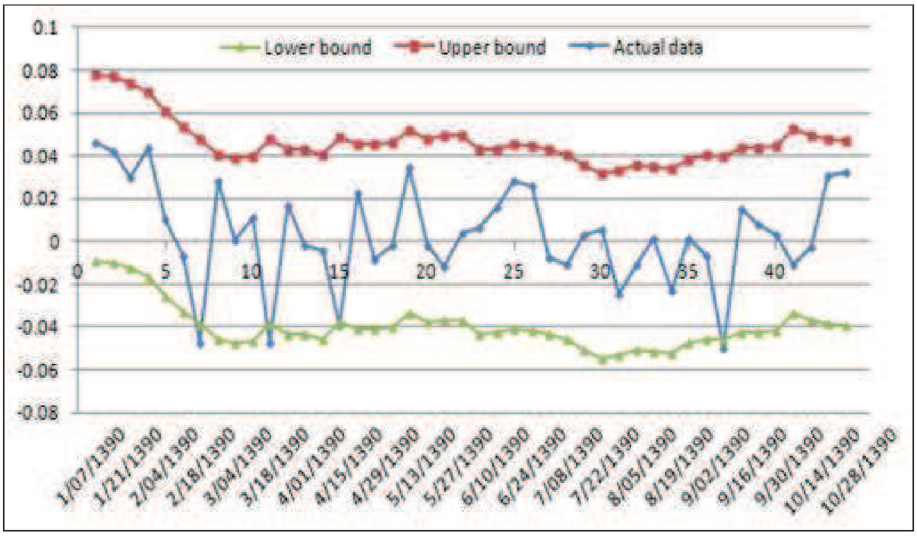
Comparison of predicted results from ARIMA and FARIMA models:

In below figure the domain of ARIMA model which is plotted with actual data has been shown:

As can be observed in four points real values are out of this domain while if we draw the domain of FARIMA model with real values, all values would be between two bounds.

Figure 2

The Domain of ARIMA Model with Real Data



Source: finding of research

5. Conclusion

According to RMSE, U-THEIL, MAE and MS E criteria, results of different methods for forecasting of share price index of Tehran Stock Exchange indicate Neural Network model has a better performance in predictions with abundant numbers of data rather than ARIMA model; however, based on Granger-Newbold and Diebold-Mariano that have a test statistic with t-distribution, prediction power of ARIMA and Neural Network does not have any significant difference together and since statistical methods have more accuracy than mathematical formulas, the end result is the prediction power parity of these two models in forecasting return on share price index.

FARIMA results consider ARIMA errors as fuzzy and advantages in predictions with limited data in FARIMA model caused this model presented an appropriate tool to forecast. Although it should be noted that FARIMA model is very sensitive to the fluctuations of data and this subject is more probable in large data which their changes are higher. Indeed in this paper we tried by hybrid methods of time series, neural network and fuzzy, improve the forecasting of time series model. Actually investors and economic policy makers will achieve the prediction of share price index by using ARIMA and Neural Network models and they obtain maximum and minimum return in Stock market through FARIMA method.

References

1. AsghariOskoi, M. (2002), *"Neural network applications in time series prediction",Iranian Journal of Economic Research*, No. XII, 97-70.
2. Chen,.C, Hae-Ching,.C, Chin Shan,H. (2009), "Forecasting value at risk (VaR) in the futures market using Hybrid method of Neural Networks and GARCH model", http://www.kyu.edu.tw/teacpage/teacpage97[27 Mar 2015].
3. Enders, W. (2010), "Applied econometric time series", Third Edition, Washington, Wiley.
4. Giovanis, E. (2009), "Application of ARCH-GARCH models and feed- forward neural networks with Bayesian regularization in Capital Asset Pricing Model: The case of two stocks in Athens exchange stock market",http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1325842.[16 Dec 2015].
5. Hooshmand, M., Shabani, M., and Zabihi, A. (2007). "The role of human capital in economic growth of Iran using Autoregressive Distributed Lag model", *Journal of Quantitative Economics (former Economic Review),* No. Two, 63-83.
6. Johnston, Jack.,Dynardv, John. (2009). "Econometric methods", New York, Springer.
7. Khashei, M., Mahmod, B. (2007). "Applying a hybrid method of artificial neural networks and fuzzy regression to predict price of gold", *Special Issue on Numerical Methods in Engineering*, No.26, 67-75.
8. Kia, M. (1983), "Neural networks in MATLAB",*Tehran, KhadamateNashre Kian RayaneSabz Publication.*
9. Maleki, G. (2011). "Prediction of gold price using fuzzy methods",*Master's Thesis*, Department of Economics and Accounting, Islamic Azad University, Central Tehran Branch.
10. Mehrara, M., Moini, A., Ahray, M., and Hamooni, A. (2009). "Modeling and forecasting Tehran Stock Exchange index and determining effective variables",*Journal of Economic Research and Policies*, 7th year, 31-51.

11. Monajemi, S. A., Abzari, M. and RaiyatiShavazi, A. (2009). “Predicting stock prices In stock market using fuzzy neural networks and genetic algorithms and their comparison with artificial neural network”,*Journal of Quantitative Economics (former Economic Review*), No. III, 1-26.
12. Pandey, A., Srivastava,V.(2008). “A Comparative study of neural-network & fuzzy time series forecasting techniques – case study: Wheat production forecasting”, *International Journal of Computer Science and Network Security*, 8, 382-387.
13. Shabani, M. (2011). “Predicting Tehran stock exchange index using time series neural networks methods”, Master's Thesis, Department of Economics and Accounting, Islamic Azad University, Central Tehran Branch.
14. Tseng, F., Tzeng, G. (2002). “Fuzzy ARIMA model for forecasting the foreign exchange market”.*Fuzzy Sets and Systems*,126 (2002) 367-376.
15. Wong, Ch. (2010). “Comparison of ARIMA time series and fuzzy time series”,*Expert Systems with Applications*, 38(8), 9296-9304.