Research Journal of Applied Sciences, Engineering and Technology 7(3): 527-532, 2014
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: February 11, 2013 Accepted: March 14, 2013 Published: January 20, 2014

# Forecasting Bank Deposits Rate: Application of ARIMA and Artificial Neural Networks

[1]Morteza Cheshti, [2]Mohammad Taher Ahmadi Shadmehri and [3]Hamid Safaye Nikoo
[1]Department of Economics, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran
[2]Department of Administrative and Economic Sciences,
[3]Department of Economics, Ferdowsi University of Mashhad, Iran

**Abstract:** In this study application of ARIMA and Artificial Neural Networks for Forecasting Bank Deposits Rate is investigated. As it's observed nowadays, banking industry is faced with great competition. The number of banks and use of new tools especially electronic banking and development of Islamic banking have maximized this competition and turned intelligent management of banks into a critical issue. Foundation of banks is based on attracting deposits; hence, forecasting the deposits has a great importance for banks. This study seeks to forecast the bank deposits. To do this, we have used the ARIMA methods with emphasis on the Box-Jenkins method as well as the Artificial Neural Network. The monthly data of different branches was used in this study for an eight-year period. This study examined the hypothesis that neural networks are more accurate than ARIMA models in forecasting the bank deposits. Research results indicate that although both models have a high capacity to forecast the variables, generally the neural network models present better results and it is better to use this method for forecasting. The neural network method has a relative advantage as $R^2$ is 16% in ARIMA Method and 99% in Neural network Method. Also RMSE is 170985 and 176960 for ARIMA Method and Neural network Method respectively.

**Keywords:** ARIMA models, deposit, forecasting, neural networks

## INTRODUCTION

In today economy, the financial institutions and organizations have a position under which no economy can survive without these institutions. Performance of these institutions and organizations is completely significant in the slowdown or growth of an economic and the efficiency of these institutions will lead to the economic growth and their inefficiencies lead to the adverse economic consequences (Goodarzvand, 2009).

Financial organizations are increasing in general and the number of banks is enhancing in particular and the competition among them is being more intense. This competition will lead to the use of new management methods and use of appropriate tools. Meanwhile, the issue of forecasting is among the major categories of each financial institution including the bank (Kia, 2010).

Basis and continuation of bank activities is based on the attraction of financial resources and the life of a bank is based on the increased attraction of resources and their optimal management. Given the random nature of bank activities in attracting the deposits, forecasting this issue is among the important issues for banks. By several ways, the banks are forecasting in different sections of bank activity. The rate of attracting the deposits is among these forecasts.

Abounoori and Sepanlou (2005) have sought to investigate in a study the Intra-Organizational Factors such as the level of banking facilities, number of employees, number of booths, coefficient of salary variation, location and facilities of branches in attracting the deposits in Bank Mellat. The average Account Balance of four main deposits in three years (current loan, loan savings, short-term and long-term investment deposits) is the dependent variable of model and the results of study indicate that the independent variables of model explain more than 99% of fluctuations in the dependent variable.

Babaei (2001) has investigated in a study the factors affecting the Account Balance of deposits in the commercial banks with emphasis on deposits of Bank Melli Iran. The result suggests that remained bank deposit has the reverse relationship with the price of Paykan car and has the direct relationship with the index of housing price and the exchange rate.

Azar and Afsar (2006) studied in a study entitled as "Modeling the forecast of stock price with the fuzzy neural network approach" the issue of forecasting the stock price. Results of research indicate this fact that the fuzzy neural networks have been superior to

**Corresponding Author:** Morteza Cheshti, Department of Economics, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran

ARIMA method in six criteria of performance evaluation.

In a study entitled as "The comparative study of linear methods of ARIMA and nonlinear methods of fuzzy neural networks in forecasting the demand for town gas subscription. Pourkazemi *et al*. (2005) indicated that the fuzzy neural network approach is superior to ARIMA method in terms of all performance criteria.

For optimizing the supply of cash, Kumar and Walia (2006) have forecasted its demand by using the artificial neural network and time series models based on the real data of cash in one of the bank branches in India from the time period, 2nd April to 30th June 2004 and their results have shown that the artificial neural networks have the superior performance than the time series methods.

In general, it seems that the findings of research indicate that the results of estimating the fuzzy neural networks have been successful and have reduced the forecasting error significantly and they have had the significant features in fast convergence, high accuracy and ability to estimate the strong function (Momeni, 2008; Moradi, 2009).

Generally, the parametric and non-parametric forecasting methods are used for future forecasting of economic variables. Using the estimated parameters, the parametric models predict the future values of variables and the non-parametric models have been formed based on this theory that the behavior of an economic variable is repeated over time and the future behavior of a variable can be found according to its past behavioral characteristics. Both methods have been used in this study for better comparison and it is sought to compare the obtained results with each other. In this study, the research background is first studied and then different methods of forecasting such as the models, ARMA and artificial networks, have been discussed. Thereupon, the models are estimated and the results of estimations are provided for various types of estimates. Then the obtained results of two methods, ARIMA and neural networks, have been compared.

## MATERIALS AND METHODS

Forecasting is among the major issues of most of the science branches, thus numerous methods have been created for this matter. In this study, two model groups, one of which is in the form of statistical and econometrics model and the other in the form of mathematical programming approaches, have been investigated and the applied models in this study in the econometric models is the ARIMA Model and the neural network model in the mathematical programming model; these two models have been studied in more detail as follows.

**Evaluation of statistical models and econometric:**

**Autoregressive Integrated Moving Average process (ARIMA):** $Y$ is a process of ARMA when its model is as follows:

$$Y_t = \xi + \alpha_1 Y_{t-1} + \beta_0 U_t + \beta_1 U_{t-1}$$

In this case, this model includes a first order autoregressive process and a first-order moving average process and $\xi$ represents the constant term. In general, $ARMA(p, q)$ is a process which includes p order autoregressive term and q order of moving average term and its model will be as follows:

$$Y_t = \phi + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots \ldots \ldots + \alpha_p Y_{t-p}$$
$$+\beta_0 U_t + \beta_1 U_{t-1} + \beta_2 U_{t-2} +$$
$$\cdots \ldots \ldots . + \beta_q U_{t-q}$$

**Autoregressive Integrated Moving Average process (ARIMA):** If a time series becomes stationary after d time of first order difference and it is then modeled by the process $\mathrm{ARMA(p, q)}$, then the main time series is the time series of Autoregressive Integrated Moving Average Process ARIMA$(p, d, q)$ in which p is the number of autoregressive terms, d: the number of first-order differencing to become stationary time series; and q: is the number of moving average terms.

It should be noted that a stationary time series or the time series, which becomes stationary after a time (or more) first order difference, should be existed for using the methodology Box-Jenkins (BJ), because the aim of BJ is the to identify and determine a statistical model which can be interpreted as the model of producing the actual sample data from the random process.

In this way, modeling includes three stages as follows:

**Identification:** At this stage, one or more models are selected from the group of ARMA models. In other words, the values are determined for p and q by investigating the autocorrelation functions.

**Estimation:** At this stage, the model (s), which have been selected in the first stage experimentally, are fitted on the data and its parameters are estimated.

**Diagnostic checking:** At this stage, the satisfactory of model (s), selected at the first stage and estimated at the second stage, are evaluated.

**Artificial neural networks:** Artificial neural networks are dynamic systems which transfer the knowledge or law hidden behind the data to the network structure by processing the experimental data. Based on the calculations on the data or examples, such these systems get the general rules, thus they are known as intelligent systems.

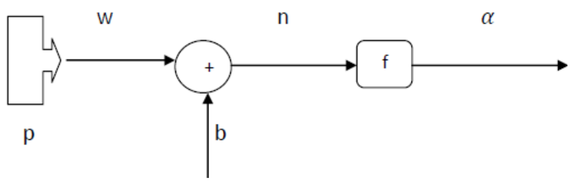


Fig. 1: The single-input neural network model

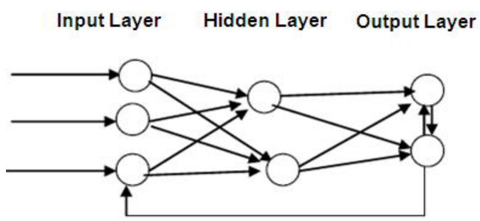


Fig. 2: Neural network structure

Artificial Neural Networks (ANN) models can be examined in single and multiple inputs modes which are as follows, respectively:

**Single-input model:** Neuron is the smallest unit of information processing and the basis of neural networks. Neuron is an analysis unit which receives and combines the signals through numerous neurons and input paths called dendrites. When the neuron is stimulated higher than a certain level (threshold), it is excited and directs an electrical signal along a single path to axon. Synapse is the contact between the axons of a neuron and dendrite of other neuron. A single neuron can have 100 to 1000 synapses and contact with 10000 neurons. Figure 1 shows the structure of a single input neuron. Scalars of $p$ and $a$ are the input and output, respectively

The effect of P on $a$ is determined by the scalar value w. Other input, which is the constant value 1, is multiplied by the term of bias 5 (b) and then is added to wp. This sum will be the net input n for the transfer function $f$. Therefore, the output of neuron is determined by the following equation:

$$\alpha = f(wp^T + b)$$

**Multiple input model:** Generally, a neuron has more than one input. Input vector is displayed with P. The scalars of $P_i$ (i = 1, 2, 3, … … ..., R) are the vector elements of P. The sum of synapses $W_{1i}$ forms the elements of matrix W. In this case, W is a row vector with the elements j = 1, 2, 2, … … .., R and $w_{1j}$. Each element of input vector, P, is multiplied by the corresponding element W. Neuron has a bias term, which is added to the multiplication of weight matrix (w) by the input P. In the multi-input mode, the net input, n, is calculated as follows:

$$n = \sum_{i=1}^{p} p_i w_{1i} + b = WP + b$$

where,

$P = [p_1, p_2, \ldots \ldots .. p_R]^T$, $W = [w_{1i}, \ldots \ldots \ldots .., w_{1R}]$

Furthermore, the neurons are connected with each other through a specific way in order to form an artificial neural network. Through the connection of neurons, the single or multiple-layer networks can be created. Multilayer network consists of an input layer, output layer and one or more hidden layers, which are between the input and output layers and connect them with each other. The number of neurons and layers, their arrangement and size form the neural network structure which is shown in Fig. 2.

In general, the neural networks can be classified according to the direction of information flow. If the data is flowed in a path from input to output, it is called the feedforward neural networks and if the data is flowed in the network in both directions by loops, it is called the recurrent neural network; these are the active networks and their conditions is constantly changing in order to reach an equilibrium point.

The general form of feedforward neural network, used in this study, is as follows:

$$F = F[\beta_0 + \sum_{j=1}^{J} \beta_j G[\sum_{k=1}^{K} \gamma_{kj} X_j]]$$

where,

J = The number of neurons in the hidden layer
K = The number of input neurons
$\beta_0$ = The bias term
$\beta_j$ = The connecting weights between hidden and output neurons
$\gamma_{kj}$ = The connecting weights between hidden and input neurons
G = The activation function of hidden layers
F = The activation function of output layer of network

Generally, the whole data in the neural networks is classified into two Training and Testing sets. Training set is used by learning algorithm in order to estimate the weights of network and the testing set is applied in order to evaluate the forecasting accuracy of training network (Abrishami and Mehrara, 2002).

## RESULTS AND DISCUSSION

For this purpose, we consider the symbols for different types of deposits. The symbol $X_1$ is considered for the long-term investment deposits, $X_2$ for the short-term investment deposits, $X_3$ for saving loan deposits, $X_4$ for current loan deposits and X for the total deposits or the sum of four types of deposits.

We consider the changes of variables or their first order difference with the high index of star. For instance, the first order difference of total deposits is displayed by X* and the first order difference of long-term investment deposits displayed as $X^*_1$.

**ARIMA model:** The software EVEIWS is used in order to estimate the parameters of model in this method.

Given the lack of reliability of all variables and the reliability of their first order difference, done by Augmented Dickey-Fuller test, we estimate the ARIMA model for their first-order difference. We obtain the optimal interval in each of the deposits by the review of diagnostic model and other ARIMA models and also different criteria especially the low and Schwartz and Akaike criteria. Moreover, the numbers in parentheses are the t-statistics, $R^2$ (coefficient of determination) and Camera-Durbin statistics and we will have respectively: The model ARIMA (12, 1, 12) for $X^*_{1t}$ is as follows:

$$X^*_{1t} = 88907/49 + 0/764085\, X^*_{1t-12} - 0/945841 e_{1t-12}$$

(2/65) (9/1) (-24)

$R^2 = 0/13$ DW = 1/99

The model ARIMA (2, 1, 12) for $X^*_{2t}$ is as follows:

$$X^*_{2t} = 24899/03 + 0/58028 X^*_{2t-2} - 0/519841 e_{2t-2} + 0/471503 e_{2t-12}$$

(2/70) (4/77) (-5/99) (8/85)

$R^2 = 0/24$ DW = 1/9

The model ARIMA (1, 1, 4) for $X^*_{3t}$ is as follows:

$$X^*_{3t} = 12621/28 - 0/729829 X^*_{3t-1} + 0/506572 e_{3-1} + 0/623832 e_{3t-4}$$

(2/17) (-8/9) (16/6) (32)

$R^2 = 0/33$ DW = 2/22

The model ARIMA (2, 1, 2) for $X^*_{4t}$ is as follows:

$$X^*_{4t} = + 21846/96 + 1/110393 X^*_{4t-1} - 0/776635 X^*_{4t-2} - 1/3675 e_{4t-1} + 0/968217 e_{4t-2}$$

(3/458) (16/8) (-11/7) (-101/9) (55/9)

$R^2 = 0/122$ DW = 2/03

The model ARIMA (1, 1, 12) for $X^*_t$ is as follows:

$$X^*_t = 100160/4 - 0/22587\, X^*_{t-1} + 0/300385 e_{t-4} + 0/25555\, e_{t-12}$$

(4/466) (-2/067) (2/698) (2/19)

$R^2 = 0/165$ DW = 1/95

Finally, the forecasted results are included in Table 1.

**Artificial neural network model:** In this section, the type of neural network, number of input layer neurons, number of hidden layer neurons, number of output layer neurons and the type of input and output transfer function are determined. Furthermore, a percentage of data, which should be assigned to the network training, validation and testing, is determined. The number of iterations is also specified. The software MATLAB has been used in order to design and forecast the target models.

The process variable as the input has been used in order to design the neural network model for forecasting the deposits and the deposit variable has been applied as the target variable and the output of model is compared to it. The purpose of network is to reduce the difference between these two variables (output and target).

The ratio of training data, validation, testing, number of hidden layer neurons and $R^2$ are presented in the following table for each of the deposits. Moreover, the network is trained through Levenberg-Marquardt algorithm and the activation function is considered in the hidden layer of Sigmoid and in the linear output layer. Input ratio of each of the deposits in the neural network is shown in Table 2:

Finally, the results of forecasting the deposits have been included in the following Table 3.

**Comparison between ARIMA and artificial neural network:** In this chapter, the monthly statistics of different types of banking deposits has been used during an 8-year period (totally 96 observations) for short-term forecast.

Given that the accuracy evaluation in the neural network, trained by Levenberg-Marquardt method, has been done through $R^2$ and RMSE (Root Mean Square Error) and these two indexes can be calculated in ARIMA method, we will compare these two as follows:

Since the important issue of reliability of variables should be considered in estimating the statistical models while using the time series data, the reliability of variables in this study has been studied based on the graph of correlation and partial correlation and also Augmented Dickey-Fuller reliability test (ADF).

The models of ARIMA have been used for their first order difference. It should be noted that the value

Table 1: Forecasting different types of bank deposits through ARIMA model for the year 2011

| | April | May | June | July | August | September |
|---|---|---|---|---|---|---|
| $X_{1t}^*$ | 4484065 | 4571162 | 4655684 | 4750327 | 4859304 | 4944956 |
| $X_{2t}^*$ | 2853323 | 2842859 | 2897365 | 2940812 | 2978992 | 3036271 |
| $X_{3t}^*$ | 1470501 | 1422241 | 1572539 | 1458066 | 1563444 | 1508369 |
| $X_{4t}^*$ | 2739520 | 2729657 | 2764164 | 2824695 | 2879665 | 2908247 |
| $X_t^*$ | 11552000 | 11650076 | 11966943 | 11922530 | 12092452 | 12238232 |
| | October | November | December | January | February | March |
| $X_{1t}^*$ | 5031284 | 5129197 | 5215817 | 5262460 | 5367618 | 5483950 |
| $X_{2t}^*$ | 3040093 | 3035728 | 3060781 | 3069683 | 3101590 | 3031322 |
| $X_{3t}^*$ | 1570397 | 1543960 | 1585898 | 1579312 | 1605951 | 1608342 |
| $X_{4t}^*$ | 2911849 | 2908206 | 2915918 | 2941867 | 2979246 | 3015154 |
| $X_t^*$ | 12388770 | 12451710 | 12512367 | 12603916 | 12851815 | 12818466 |

Table 2: Input ratio of each of the deposits in the neural network software

| | Training data (%) | Validation data (%) | Testing data (%) | Hidden layer neurons | $R^2$ (%) |
|---|---|---|---|---|---|
| Long term deposits | 70 | 15 | 15 | 0 | 98 |
| Short-term deposits | 70 | 15 | 15 | 10 | 99 |
| Savings deposits | 65 | 20 | 15 | 5 | 99 |
| Current deposit | 75 | 10 | 15 | 10 | 99 |
| Total deposits | 65 | 20 | 15 | 8 | 99 |

As it's shown in Table 2 $R^2$ is 98 and 99% in every deposit

Table 3: Results of forecasting by the above networks for different types of deposits during a 12-month period

| | April | May | June | July | August | September |
|---|---|---|---|---|---|---|
| $X_1$ | 4376400 | 4455800 | 4474000 | 4485800 | 4493700 | 4449100 |
| $X_2$ | 2931500 | 2936600 | 2941600 | 2946300 | 2950900 | 2955300 |
| $X_3$ | 1375600 | 1389100 | 1399700 | 1408000 | 1414300 | 1419200 |
| $X_4$ | 2734200 | 2751900 | 2766300 | 2777700 | 2786700 | 2793500 |
| X | 11484000 | 11712000 | 11938000 | 12158000 | 12371000 | 12577000 |
| | October | November | December | January | February | March |
| $X_1$ | 4502600 | 4505000 | 4506600 | 4507600 | 4508300 | 4508700 |
| $X_2$ | 2959600 | 2963700 | 2967700 | 2971500 | 2975200 | 2978700 |
| $X_3$ | 1422900 | 1425600 | 1428700 | 1429300 | 1430500 | 1431400 |
| $X_4$ | 2798600 | 2802300 | 2804900 | 2806600 | 2807600 | 2807900 |
| X | 12773000 | 12961000 | 13140000 | 13308000 | 13467000 | 13616000 |

As it's shown in Table 3 different types of deposits during a 12-month period forecast logical estimation in neural network software

Table 4: Comparison of indexes $R^2$ and RMSE in results obtained from forecasting two methods

| | ARIMA method | | Neural network method | |
|---|---|---|---|---|
| | $R^2$ (%) | RMSE | $R^2$ (%) | RMSE |
| Long term | 13 | 130934 | 98 | 61880 |
| Short term | 23 | 56729 | 99 | 70396 |
| Savings | 33 | 54962 | 99 | 45566 |
| Current | 12 | 65895 | 99 | 73364 |
| Total deposits | 16 | 170985 | 99 | 176960 |

of $R^2$ will be higher if the parameters are estimated through the level of variables while being unreliable; however, these estimates will have no enough credits while the software will be reporting the unreliability. Therefore, the value of $R^2$ will be low when the parameters are estimated by the first order difference and this is evident in the whole conducted study.

It should be noted about the forecasts, done by the neural network model, that these models require no evaluation of the reliability of variables, thus the level of variables have been tested; therefore the values of their $R^2$ are very high, thus two methods should not be compared through this criterion.

However, through the second criterion, evaluating the RMSE and due to the lack of data as the result of bank constraints, it can be concluded that those two methods have a little difference with each other. Table 4 also indicates this point. Generally, the neural network method has a relative advantage because the major point, which should be noted, is that the neural network models need more data than the statistical models in order to provide the most appropriate responses; in this study we are faced with the data limitations due to the lack of banking resources statistics.

## CONCLUSION

This study has been developed with the aim to forecast different types of bank deposits and this issue is important because the major activity of banks is the proper management of resources and expenditures and the rate of deposits plays the major role in lending power and investment of banks and the high rate of deposit is also so important in terms of making the money in the economy.

According to the rates of deposits and by considering the legal restrictions, the bank can offer the loan in a certain range, determined by the central bank and make the profit by this way. On the other hand, future is uncertain and the bank management should forecast the future by a way for better decisions.

The results indicate that both methods have a high potential in forecasting the variables because in most of the cases the forecasts were close to the reality in the periods with real observations; however, it should be noted that the neural network models have had the most appropriate results and according to the numerous studies, conducted in this regard, it can be concluded that by increasing the statistical data or more observations, the better and more accurate results can be expected from using these models.

On the other hand, if the deposits are also available daily and given the more available observations, the forecasts will be more accurate and it will be possible to forecast the periods shorter than 10 days or one month. Moreover, the banks, themselves, can forecast the deposits through the current software or designing the specific software in order to design the system which always updates this information, but it should be noted that the existence of necessary statistics and information is the most important tool in this regard.

## REFERENCES

Abounoori, E. and H. Sepanlou, 2005. Analyzing the effects of Intra-Organizational Factors on attracting the banking deposit (Case Study: Bank Mellat Tehran). Daneshvar Raftar, October 2005; 12.

Abrishami, H. and M. Mehrara, 2002. Applied Econometrics (new approaches). Publications of University of Tehran, Tehran.

Azar, A. and A. Afsar, 2006. Modeling the stock price forecast with the approach of fuzzy neural networks. Q. J. Bus. Res., No. 40, Fall.

Babaei, E., 2001. Investigating the factors affecting the account balance of deposits in commercial banks with emphasizing on the deposits of Bank Melli Iran. Publications: Economy, Econ. Res. J., No. 2

Goodarzvand, C.A., 2009. Tool boxes of MATLAB. Naghoos Press, Tehran.

Kia, M., 2010. Neural Networks in MATLAB. Kian Rayaneh Sabz Publications, pp: 229.

Kumar, P. and E. Walia 2006. Cash forecasting: An application of artificial neural networks in finance. Int. J. Comput. Sci. Appl., 3: 61-77.

Momeni, M., 2008. Modern Principles of Research in Operations. Publications of University of Tehran, Iran.

Moradi, A., 2009. Application of Eviews in the Econometrics. Publications of Jihad University of Tehran, Iran.

Pourkazemi, M.H., A. Afsar and B. Nahavandi, 2005. A comparative study of linear methods of ARIMA and non-linear methods of fuzzy neural networks in forecasting the demand for town subscription. J. Econ. Res., No. 71.