The ARIMA versus Artificial Neural Network Modeling

Motaz Khorshid
Faculty of Computer & Information. Cairo University, Egypt

Assem Tharwat,
Faculty of Computer & Information. Cairo University, Egypt

Amer Bader
Faculty of Computer & Information. Cairo University, Egypt

Ahmed Omran
Central Lab For Agricultural Expert Systems. Agriculture Research Center, Egypt

Abstract *Linear models almost reach their limitations with non-linearity in the data. This paper provides a new empirical evidence on the relative macroeconomic forecasting performance of linear and nonlinear models. The well-established and widely used univariate Auto-Regressive Integrated Moving Average (ARIMA) models are used as linear forecasting models whereas Artificial Neural Networks (ANN) are used as nonlinear forecasting models. The neural network paradigm that was selected for developing the proposed model is a Multi-layer Feedforward network based upon the Backpropagation training algorithm. ANN has been proven to be successful in handling nonlinear problem optimization and prediction. The forecasting models used to identify whether action is needed to alter the future, when such action should be taken by the decision maker in order to change the future of the bank or its environment to improve the bank's chance of achieving its targets. We applied the proposed model on a Financial Balance Sheet's data of a commercial bank in Egypt. The Results show that, the proposed model (which dependent on the ANN) is more accurate than the other models, which depend on the ARIMA model with accuracy between 8 % and 10.4 %.*

Keywords: *Economic Forecast, Neural Networks, ARIMA, Back-propagation, and Time Series Models.*

1. Introduction

This paper is mainly an empirical research effort intended to compare the forecasting performance of ANN and traditional time series methods, such as ARIMA. Recent work that has used ANN models for forecasting purposes in different applications such as Inflation forecasting, forecasting daily foreign exchange rates, etc [4, 5, 8, 9, 11, 15, 16, 20, 21, 25, 26]. In most of these experiments ANN has indicated a good forecasting accuracy.

Balance Sheet is a list that shows the financial condition of the firm at a particular date. It consists of two sides one side is called Assets and the other side is called Liabilities and Owner's Equity (the total liabilities)[17, 18,24].

This paper is concerning on the most important five items of our case study commercial bank's balance sheet, which are Cash Balances with C.B.E, banks & foreign correspond, Total Loans, Total Assets, Customer Deposits and Total Deposits. The data collected for those items was represented by five different time series (TS1, TS2, TS3, TS4, TS5) to be analysis in this paper to predict the future values for these items and perform a comparative exercise and access the relative performance of the different forecasting methods.

Commercial relations with the five previous items can calculate the all values of other items in the Balance Sheet [17, 18].

Neurosolution, commercially available neural networks simulator, was used in the training of the neural networks models and MINITAB12 was used as statistical software.

This paper includes seven sections after introduction. The next two sections after the introduction briefly discuss methodology and mathematical formulation of ARIMA models and ANN. Section four includes an overview of the software that used for applying both ARIMA and ANN. ARIMA models and the ANN models designs are given in sections five and six. In section seven the comparison between the ARIMA and ANN forecasting models is discussed, and finally some conclusions and further points for research are in the last section.

2. ARIMA Models

ARIMA models were originally proposed by Box and Jenkins (1976), and it is considered today a quite popular tool in economic forecasting. The basic idea is that a stationary time series can be modeled as having both autoregressive (AR) and moving average (MA) components. AR models are based on the application of regression analysis to lagged values of the Yt series. The Autoregressive model of order p, AR (p) is:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (1)$$

In the context of Box-Jenkins Yt = the actual value of the series at time t, Yt-1 = the actual value of the series at time t-i, ϕi = the Autoregressive parameter for Yt-i , εt =the irregular fluctuation at time t, not correlated with past values of the Yt's [19, 22].

MA models are based on the past levels of the series, Yt may be influenced by the recent"stocks" (i.e., random errors) to the series that is, the current value of a series may be the best explained by looking at the most recent q error [19, 22]. The Moving Average model of Order q, MA (q) is:

$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2}$$

θi= The Moving Average parameters for εt-i, εt-i = the error term at time t-i, εt = the error term at time t, and εt, εt-1, εt-2, …are uncorrelated with one another.

Mixing autoregressive (AR) and moving average (MA) terms in the same model called ARMA model .The forecasted values of Yt in ARMA model of order (p) and (q), ARMA (p, q) is given by:

$$Y_T = \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t + \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_P Y_{t-P} \tag{3}$$

εt, εt-1, εt-2, …are uncorrelated with one another [19, 22]. ARIMA model equations are the same ARMA equations except Yt is replaced by the differenced series Wt: Wt=Δd Yt = the Yt series differenced d times. The forecasted values of Yt in ARIMA model of order (p), (d) and (q), ARIMA (p d, q) is given by:

$$W_t = \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t + \phi_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + ... + \phi_P W_{t-P} \tag{4}$$

εt, εt-1, εt-2, …are uncorrelated with one another [19, 22].

Non-stationary integrated series can also be handled in the ARIMA framework, but it has to be reduced to be stationary.

Fig1 shows schematic representation of Box-Jenkins methodology for time series modeling that includes three phases that are Identification, Estimation, Diagnostic and Forecasting [22].

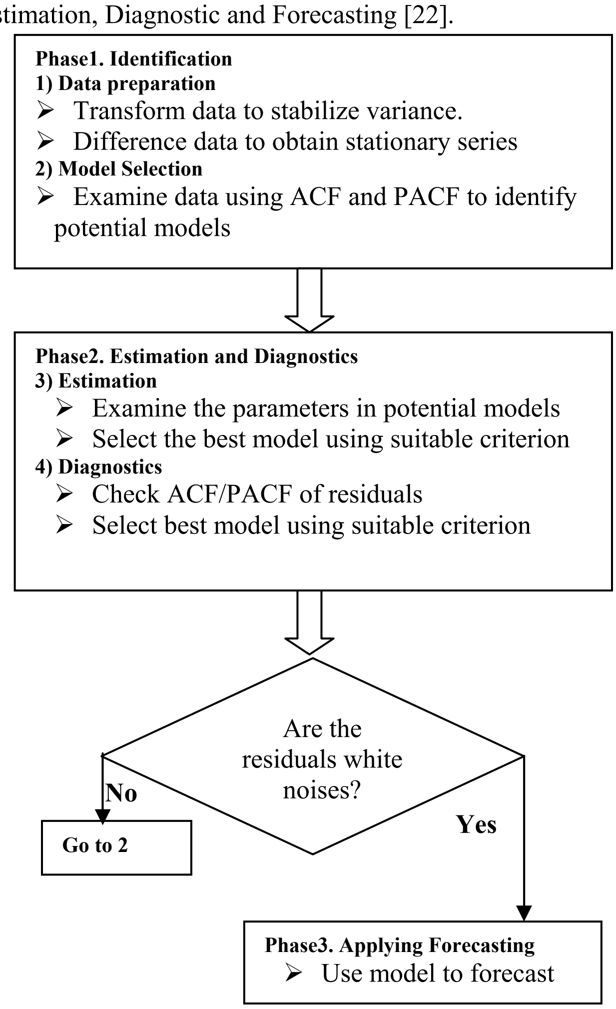


Figure 1 Schematic representation of Box-Jenkins methodology for time series modeling

Table 1 characteristic patterns in the ACF and PACF of ARIMA models

Model	Autocorrelation Function	Partial Autocorrelation Function
AR (p)	Die Out	Cuts off after lag p
MA (p)	Cuts off after lag q	Die Out
ARIMA	Die out after lag q-p	Die out after lag p-q

- Identification: the series is differenced, if necessary, to make it stationary. The data may be stationary in the viewpoint of trend through one or two differences but the time series may be non-stationary. To stabilize the variance, a nonlinear transformation such as a logarithmic or square root transformation is often performed. Then the sample ACF and PACF are calculated; the behavior of both the ACF and PACF determines the number of AR (p) and /or MA (q). Table1 shows the characteristic patterns in the ACF and PACF of ARIMA models [19, 22].
- Estimation and Diagnostics
- In Estimation, least squares estimates of the process parameters are generated. In diagnostic, checking the residuals from the estimated model should look like a random series; failing that, further analysis of the residuals leads to a re-specification of the model [19, 22].

Forecasting: - the fitted model, having first been "integrated "if necessary, is used to forecast the Yt's. (In practice, each of these stages requires the use of a scientific computer program) [19, 22].

3. Artificial Neural Networks

Neural Network is a branch of artificial intelligence. ANN act like a human brain, trying to recognize regularities and patterns in the data. They can learn from experience and generalize based on their previous knowledge. Neural networks are composed of highly interconnected processing elements (nodes) that work simultaneously to solve specific problems. In time series analysis ANN models were used as nonlinear function approximations. ANN takes in a set of inputs and produces one/a set of outputs according to some mapping rules predetermined in their structure [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13].

This paper considers the most popular form of ANN, which called the feed-forward network. The selected feed-forward neural network model can fit the financial analysis problem because of their adaptively owing to their structure [20, 23, 24]. The existence of hidden layer and nonlinear activation function models the nonlinearity of the data. This important property of feed-forward neural network models enables modeling multi-attribute, nonlinear mapping for the financial analysis problem [3, 4, 5, 6, 7, 8, 9, 14, 15, 16, 20, 24, 25].

Fig.2 depicts such a network that consists of layers of nodes; the input layer and output layer represent the input and output variables of the model. Between the input and output layers there are one or more hidden layers that progressively transform the original input stimuli to final output and enables ANN to learn nonlinear relationships [6, 10, 13].

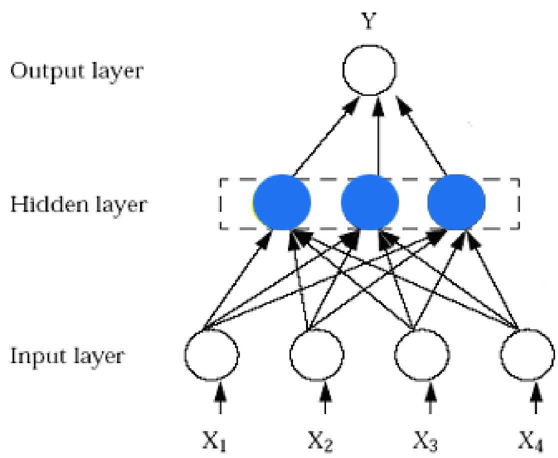


Figure 2 The feed forward neural network with a single hidden layer

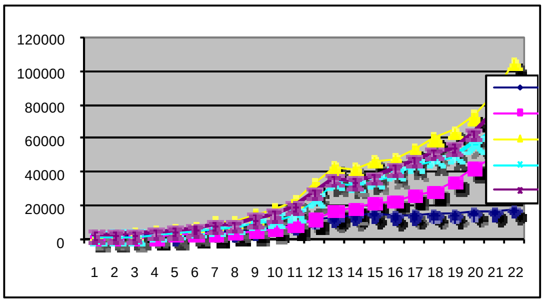


Figure 3 The plot of all time series

ANN has to be trained in order to be used to perform certain tasks like predicting a response corresponding to a new input pattern. The training procedure involves iteratively modifying the randomly initialized weights of the ANN to minimize some kind of error function usually the mean square error (MSE) [6, 10, 13].

Various standard optimization techniques such as the conjugate gradient and quasi-Newton methods exist for minimizing the error function, however, in application studies, the Back-propagation Algorithm developed by the neural network community is the most popular training algorithm used [6, 10,13, 20,24,26].

4. Implementation Tools

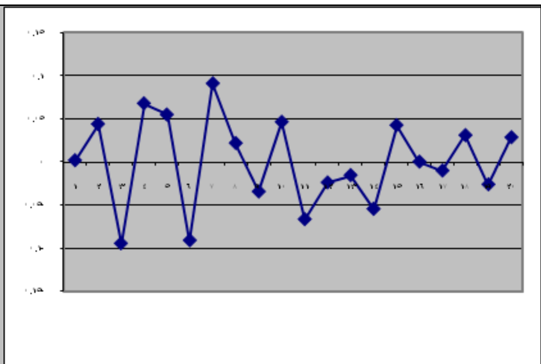
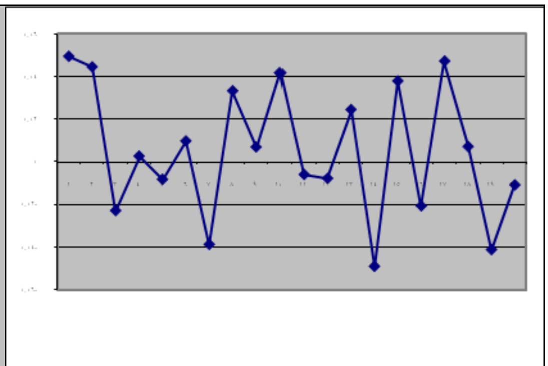
The Network simulator such as NeuroSoltion requires no programming skills and often come with special hardware to minimize the training time. The commercially available neural network simulator NeuroSoltion (NeuroDimension incorporated) was used for development of the proposed neural network application. Use of a shell program of this type is attractive for forecasting financial data environment. MINITAB is the statistical software that offers the methods that we need to implement ARIMA model.

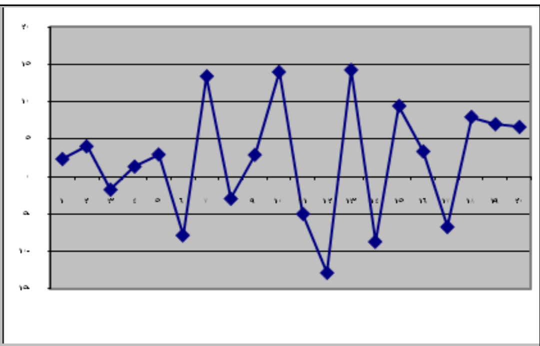
5. Designing ARIMA Models

First, each time series should be plotting to discover whether it is stationary or not. Fig3 shows that there is a trend and no seasonality in the given time series. When we plot every time series after two differences the un-stabilization in variances was appeared. To stabilize variances in every time series transformation functions are used [19, 22].

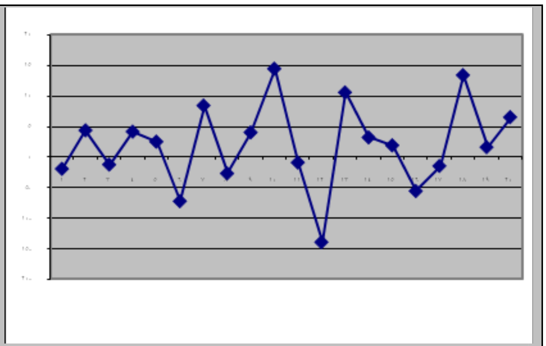
We tried two types of transformation functions, logarithmic and square root, followed by taken two differences for each time series to be stationary. In Fig4 Every time series "looks" stationary, since the time plot of the series appears "similar" at different points along the time axis [12, 14, 19, 22].

Autocorrelation and partial autocorrelation functions are used to identify an acceptable model for each time series; Fig5 shows the autocorrelation and partial autocorrelation functions for every time series. From the values of ACF & PACF we can deduce the acceptable models for each time series. Table2 states their models.

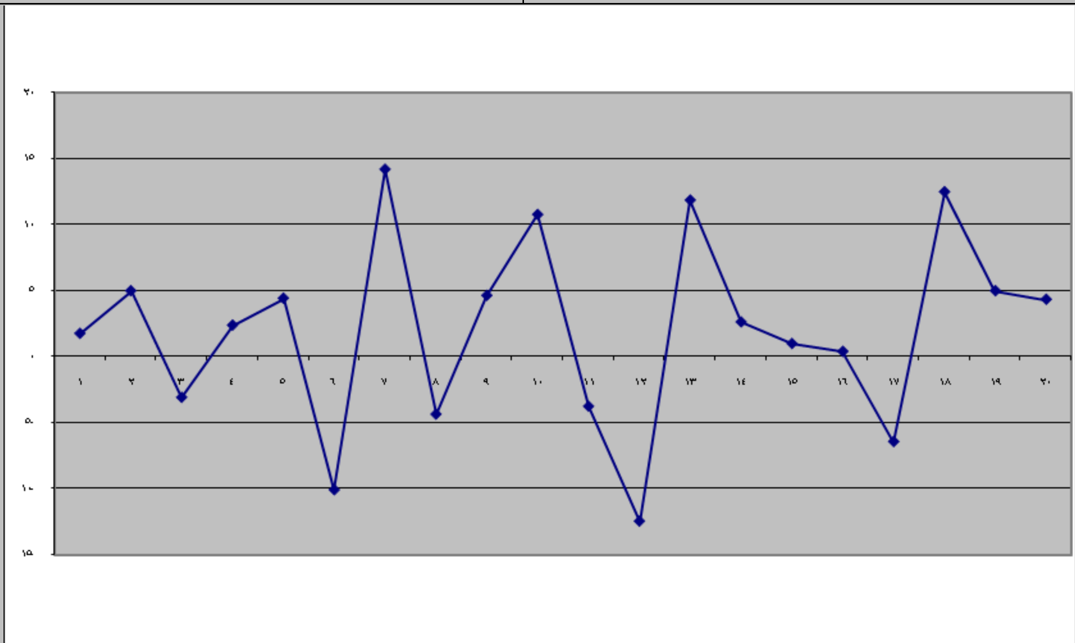
	
TS1 with “log 10” transfer function	TS2 with “log 10” transfer function



TS3 with “log 10” transfer function and two

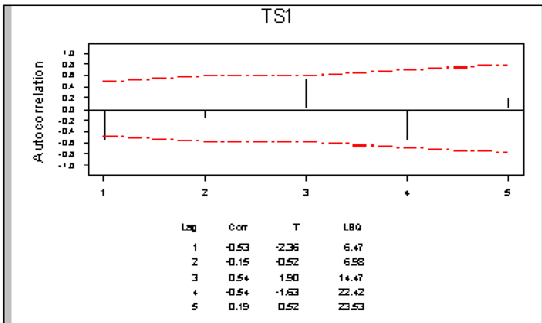


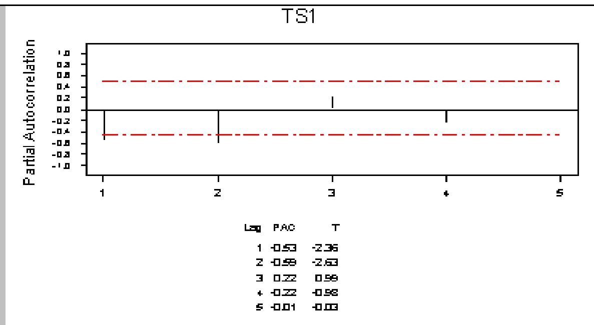
TS4 with “square root” transformation function



TS5 with “square root” transformation function and two differences

Fig4. Time series plots after be stationary





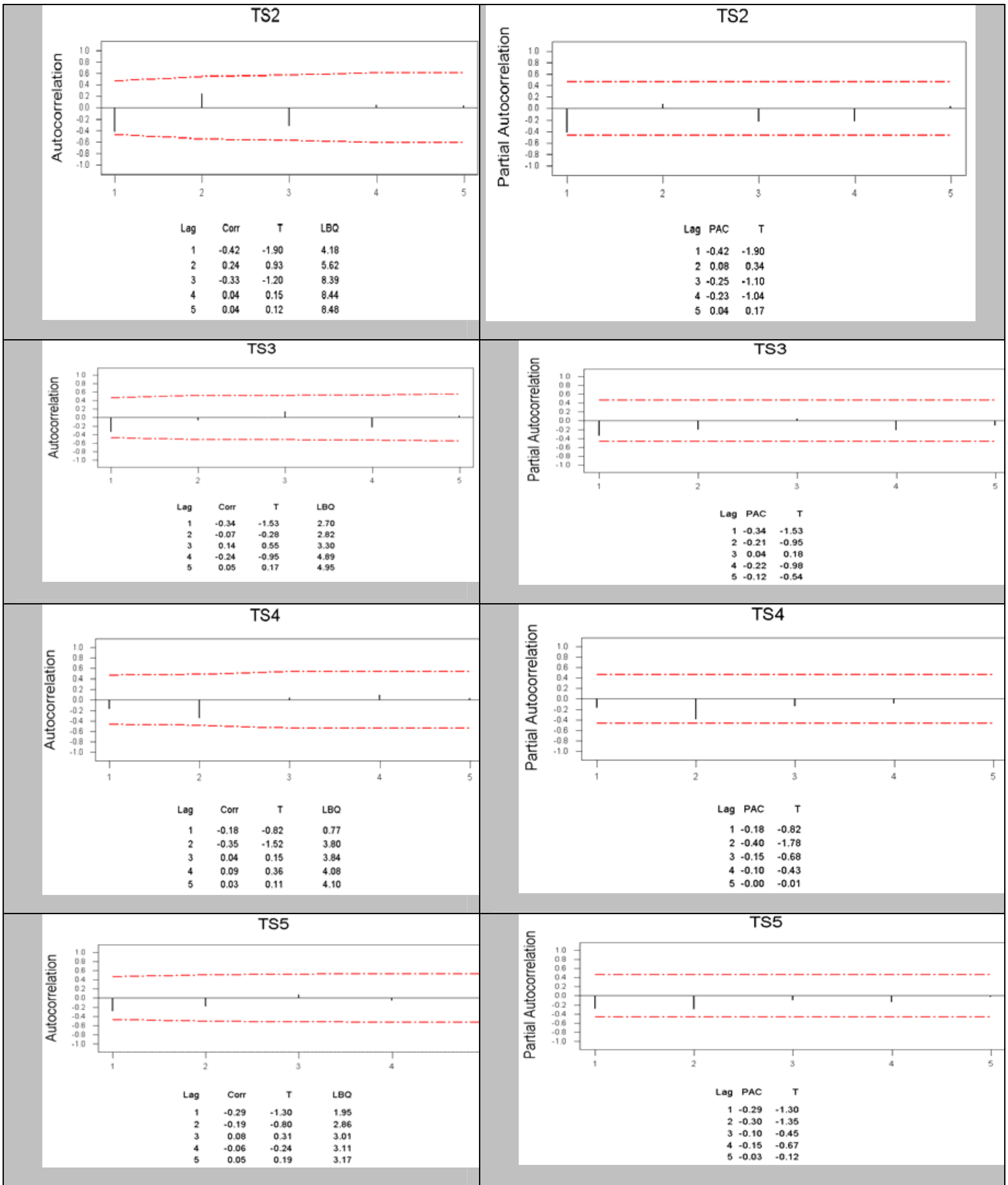


Fig5.Autocorrelation and Partial Autocorrelation functions for every time series (TS1, TS2, TS3, TS4, T5).

Time Series	ACF &PACF	Model
1	The autocorrelations die out in a damped sine-wave manner and that there are exactly **two** significant partial autocorrelation.	AR (2)
2	**One** non-zero autocorrelation at lag1 and that the partial autocorrelation decay exponential.	MA (1)

3	**One** non-zero autocorrelation at lag1 and that the partial autocorrelation decay exponential.	MA (1)
4	At the autocorrelation decay exponential after the interval [q-p] and in the partial autocorrelation decay exponential after the interval [p-q].	ARIMA (1, 2, 1)
5	A tremendous variety of patterns in the ACF and PACF.	ARIMA (1, 2, 1)

Table2. Analysis of ACF &PACF and the acceptable model for each time series

The MINITAB software is used to examine the parameters in potential models and select best models. Tables 3,4 summaries the results to select the best model.

Time Series	Model	Parameters	Corresponding Equation of the best model
1	AR (2)	$\phi_1 = -0.4399$, $\phi_2 = -0.297$ ϕ_0 = Not significant.	$\Delta^2 \text{Log } Y_t = \log [-0.4399 Y_{t-1} - 0.297 \; Y_{t-2}] + \varepsilon_t$
2	MA (1)	$\theta_1 = 0.9518$ θ_0= Not significant.	$\Delta^2 \text{Log } Y_t = \log [-0.9518 \varepsilon_{t-1}] + \varepsilon_t$
3	MA (1)	$\theta_1 = 0.9510$, $\theta_0 = 0.693$	$\Delta^2 \text{Log } Y_t = 0.693 + \log [-0.9510 \varepsilon_{t-1}] + \varepsilon_t$
4	ARIMA (1,2,1)	$\phi_1 = -0.624$, $\theta_1 = 0.9214$ Constant is not significant.	$W_t = \Delta^2 \sqrt{Y_t}$ $W_t = [-0.9214 \varepsilon_{t-1} - 0.624 \, W_{t-1}] + \varepsilon_t$
5	ARIMA (1,2,1)	$\phi_1 = -0.474$, $\theta_1 = 0.9118$ Constant is not significant	$W_t = \Delta^2 \sqrt{Y_t}$ $W_t = [-0.9118 \, \varepsilon_{t-1} - 0.474 \, W_{t-1}] + \varepsilon_t$

Table3.The best models for TS1, TS2, TS3, TS4 and TS5

Where ε_t was generated from a standard normal distribution.

Diagnostic checking was applied and indicated that all the pervious models are good forecasting models. The ACF &PACF spicks were outside the limits, which suggesting the residual series were white noise. A portmanteaus test was applied to the residuals as an additional test, in this case the value Q* was not significant, that mean the residuals was considered white noise [19, 22].

6. Designing the Neural Networks

A feed-forward neural network model was employed to mimic the complex mapping function between the inputs and output. The Back-propagation learning algorithm is used to perform the training requirements. The determination of the network structure is a very difficult task as involves many variables including learning rate parameter, number of hidden layers, and the number of hidden units per hidden layer [1, 3, 6, 10, 13, 20, 23, 25].

The learning rate parameter plays a critical role in the convergence of the network to the true solution. For a given network and an infinitesimal learning rate, the weights that yield the minimum error can be found. However, they may not be found in a reasonable span of time. Use of a large learning rate proceeds much faster but may also produces oscillations between relatively poor solutions. However, high-dimensional spaces (with many weights) have relatively few local minima [6, 10, 13].

The number of the inputs variables (number of lags) in this work was between 2 and 4 inputs, the number of hidden layers was selected as one hidden layer and the number of neurodes per hidden layer depends on the number of nodes in the input layer. With too few connections, the network cannot learn much. The net result of poor parameter settings will be slow convergence and poor performance of the model [1, 2, 3, 4, 5]. Table5 describes the different neural network structures for the different time series.

Network parameters	The Model Number				
	M1	M 2	M 3	M4	M 5
1.No. Of input units	2	3	4	3	3
3.No.Of hidden units	3	7	9	7	7
2.No. Of middle layers	1				
4.No.Of output units	1				
5.Learning rate parameter	0.7				
6.Transfer Functions	Hyperbolic tangent (*tanh*) function				

Table5. Description of the ANN Structure.

The logistic function is the most popular activation function among researchers for the hidden layer. However, we use the hyperbolic tangent (tanh) function, as it has been used very successfully in forecasting experiments. It is also generally held that (tanh) gives rise to faster convergence of training algorithms than logistic functions. For the output layer, we follow the recommendation of who suggest the use of the linear function for time series prediction with continuous output [1, 2, 6, 10].

In this paper we use one of the most common forms of preprocessing which consists of rescaling the data in the range [-1, 1] so that they have similar values [6, 10, 13].

- Training of the Neural Network Models
    - The following general guidelines were considered during the training:
    - A fully connected neural network.
    - Pattern (training examples) was presented sequentially during each training session.
    - Updating network weights was performed after each training pattern was presented to the network.
    - The stopping criteria were set such that the total number of iterations does not exceed 50,000 epochs, or a 0.0001 root-mean-square (RMS).
    - Each train – and – test experiment is repeated three times with different random starting state to make sure that the solution obtained is not a local minimum [6, 10, 13].
- **The Results of the ANN Forecasting Models**

The final results of these experiments revealed that the network with between 3-9 hidden units had a better ability to decrease the system error rate (error rate due to training is between 3.5 and 4.0%) and provided a good prediction (average error rate due to test is between 8.3 to10.6 %). Figures from Fig (6-10) show the plots of the final results of the ANN forecasting models.

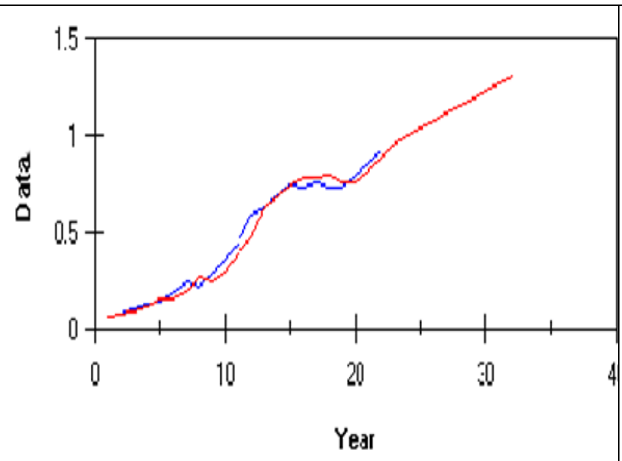


Fig6.Model1 for TS1

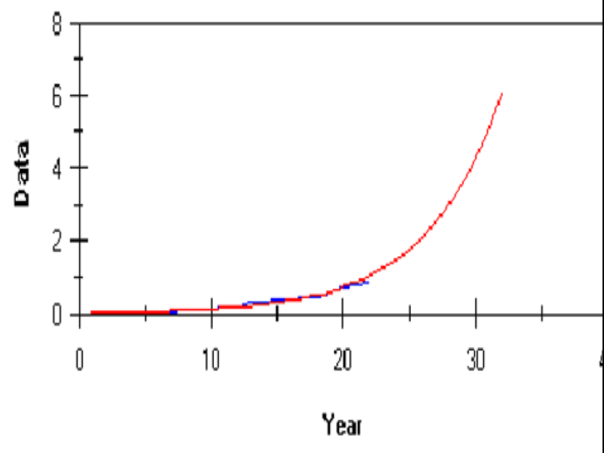


Fig7.Model2 for TS2

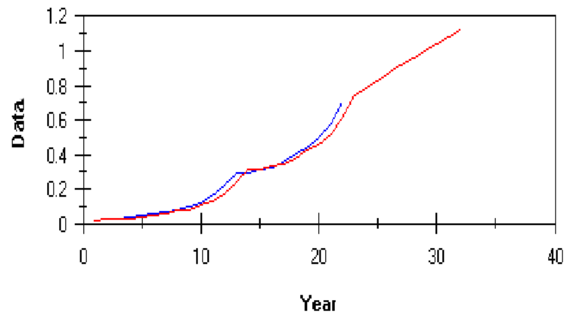


Fig8.Model3 for TS3

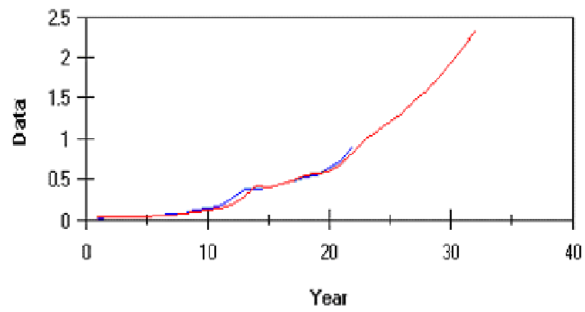


Fig9.Model4 for TS4

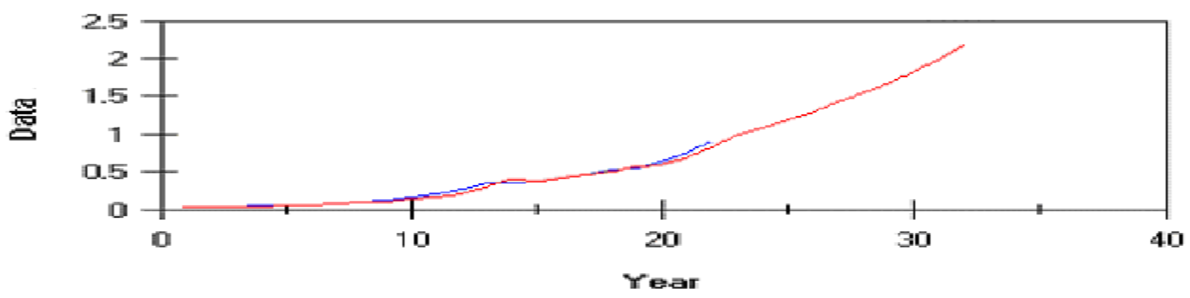


Fig10.Model5 for TS5

7. The Comparison between the Accuracy of the ARIMA and Neural Networks Forecasting Models

There are several measures of accuracy but each of them has its advantages and limitations. For this reason none of them has been accepted universally as the optimum measure of accuracy [10, 13, 19, 22]. In this study we shall use two popular types of performance measures, which are the absolute percentage error (APE) and the Root mean square error (RMSE):-

These measures can be calculated by the following formulas:

$$\textbf{APE} = \sum_{i=0}^{n-1} |(ti - oi)| \,/\, \sum_{i=0}^{n-1} |(oi)| \qquad \textbf{(5)}$$

$$\textbf{RMSE} = \sqrt{[\frac{1}{n} \sum_{i=0}^{n-1} (ti - oi)2\,]} \qquad \textbf{(6)}$$

In Eq5, 6 ti is the target output and oi is the actual output, and n is the number of test samples. The accuracy of the model calculated by: - the accuracy = 1 – (APE or RMSE) %.

The numerical results of (APE) and (RMSE) measures for both ANN and ARIMA models show in table6a and table6b.

Models	**APE values of ANN**	APE values of ARIMA
Model1	**8.30%**	**16.90%**
Model2	**8.90%**	**17.88%**
Model3	**9.30%**	**18.90%**
Model4	**10.30%**	**20.70%**
Model5	**10.60%**	**20.97%**

Table6a. The values of APE for both ANN and ARIMA models

APE values that were described in Table6a deduce that absolute percentage error of the ARIMA models in is between 16.90 % & 20.97 % and the absolute percentage error of the ANN models is between 8.30 % & 10.60%.

RMSE measure that was described in Table6b also confirmed that neural networks outperform ARIMA models.

Models	**RMSE values of ANN**	RMSE values of ARIMA
Model1	**7.60%**	**17.30%**
Model2	**8.20%**	**18.00%**
Model3	**8.60%**	**19.10%**
Mode4	**10.90%**	**20.80%**
Model5	**11.30%**	**21.00%**

Table6b. The values of RMSE for both ANN and ARIMA models

8. Conclusions and Further Points for Research

There is growing evidence that macroeconomic series contain non-linearities but linear models such as the ARIMA models are widely used for forecasting such series, despite the inability of linear models to cope with non-linearities. In this paper we provide a new empirical evidence on the relative macroeconomic forecasting performance of linear ARIMA models and the nonlinear ANN.

Results show that neural networks outperform ARIMA models in our forecasting problem. As an extension of this work, we hope to further refine the best neural network models in this paper by considering additional layers and different training periods to exploit readily available monetary and financial data in order to gauge future macroeconomic activity.

Comparison with more complicated forecasting models to prove the quality of our model is one of our main future research points. Another point is the possibility to build another approach with combing different forecasting models such as ARIMA and neural network models.

9. References

[1] Adya, M. and Calopy.(1998). How effective are neural networks at forecasting and prediction? A review and evaluation, Journal of Forecasting, 17, 481-95.

[2] Balkin, S.D. and Ord, J.K. (2000). Automatic neural networks modeling for univariate time series, International Journal of Forecasting, 16, 509-15.

[3] Barron, A.R. (1994). A comment on neural networks: a review from a statistical perspective, Statistical Science, 9, 33-5.

[4] Binner, J.M, Gazely, A.M. (1999). A neural network approach to inflation forecasting: the case of Italy, Global Business and Economics Review, 1, 76-92.

[5] Binner, J.M, Gazely. (2002). Financial innovation and Divisia indices in Taiwan: A neural network approach, European Journal of Finance, 8, 238- 47.
[6] Bishop, C.M. (1995). Neural Networks for Pattern Recognition, Oxford University, Press, NewYork.
[7] Boger.Z, Weber.R. (2000). Finding an Optimal Artificial Neural Network Topology in Real-Life Modeling. In: Proceedings of the ICSC Symposium on Neural Computation, Article No. 1403/109.
[8] Church, K.B, Curram.S. (1996). Forecasting consumers' expenditure: a comparison between econometric and neural network models, International Journal of Forecasting, 12, 255-67.
[9] Dorsey, R.E. (2000). Neural networks with Divisia money: better forecasts of future inflation?, in Divisia Monetary Aggregates: Theory and Practice (Ed.) M.T. Belongia and J.M. Binner, Palgrave, New York, pp. 28-43.
[10] Gately, E.(1996). Neural Networks for Financial Forecasting, John Wiley, and NewYork.
[11] Gazely, Binner, J.M. (2000). The application of neural networks to the Divisia index debate: evidence from three countries, Applied Economics, 32, 1607-15.
[12] Gorr, W.L, Szcypula. (1994). Comparative study of artificial neural network and statistical models, International Journal of Forecasting, 10, 17-34
[13] Hertz.J, Krogh, Palmer. (1991). Introduction to the Theory of Neural Computation. Addison-Wesley: Redwood City, California.
[14] Mareleo.S, Portugal. (1995). Neural network versus time series methods a forecasting exercise, 14th international symposium on forecasting.
[15] Moshiri, Cameron.(2000). Neural networks versus econometric models in inflation forecasting, Journal of Forecasting, 19, 201-17.
[16] Nag, A.K, Mitra. (2002). Forecasting daily foreign exchange rates using genetically optimized neural networks, Journal of Forecasting, 21, 501-11.
[17] National Bank of Egypt. (1980 to 2002). Economic Bulletins, Vol.No.4.
[18] Naylor, T. H. (1980). Corporate Planning Models, Addison & Wisley.
[19] Nicholas.R, Farnuum, LaVerne.W. (1989). Quantitative-forecasting methods, PWS-KENT publishing bank
[20] Plasmans, Verkooijen, Daniels. (1998). Estimating structural exchange rate models by artificial neural networks, Applied Financial Economics, 8, 541-51.
[21] Shazly, M.R.E, Shazly.(1999). Forecasting daily foreign exchange rates using genetically evolved neural network architecture, International Review of Financial Analysis, 8, 67-82.
[22] Spyros.M, Stephen.C, Wheelwright. (1998). Forecasting methods and applications, third Edition, John Wiley and sons Inc.
[23] Tiao, G.C, Tsay. (1994). Some advances in non-linear and adaptive modeling in time series, Journal of Forecasting, 13, 109-31.
[24] Thomas.H, Naylor. (1985). The Age of Corporate Planning Models, Addison & Wisley.
[25] Tkacz.G. (2001). Neural Network forecasting of Canadian GDP growth, International Journal of Forecasting, 17, 57-69.
[26] Zhang, H.U, Y.H. (1998). Forecasting with artificial neural networks: the state of the art, International Journal of Forecasting, 14, 17-34.