STATISTICS IN TRANSITION new series, Spring 2015
Vol. 16, No. 1, pp. 83–96

APPLICATION OF BOX-JENKINS METHOD AND ARTIFICIAL NEURAL NETWORK PROCEDURE FOR TIME SERIES FORECASTING OF PRICES

Abhishek Singh[1], **G. C. Mishra**[2]

ABSTRACT

Forecasting of prices of commodities, especially those of agricultural commodities, is very difficult because they are not only governed by demand and supply but also by so many other factors which are beyond control, such as weather vagaries, storage capacity, transportation, etc. In this paper time series models namely ARIMA (Autoregressive Integrated Moving Average) methodology given by Box and Jenkins has been used for forecasting prices of Groundnut oil in Mumbai. This approach has been compared with ANN (Artificial Neural Network) methodology. The results showed that ANN performed better than the ARIMA models in forecasting the prices.

Key words: forecasting, feed forward network, ARIMA, ANN.

1. Introduction

Price forecasting is very essential for planning and development. Therefore, it has become pertinent to develop methods which help the policy makers to have some idea about the prices of commodities in the future. There are various approaches to forecast prices such as using econometric methods which use economic theory and cause and effect relationships to forecast prices of essential commodities. These approaches require a large amount of information regarding different variables which may lead to various types of errors. The time series approach to forecasting is an approach which relies on the assumption that the past pattern in a time series will be repeated in the future and this information can be used to forecast prices. There are many methods for analyzing a time series but one of the most simple and benchmark method is that of Box and Jenkins (1970) which is popularly known as ARIMA methodology. De Gooijer and Hyndman

[1] Asstt. Prof., Department of Farm Engineering; Institute of Agricultural Sciences; Banaras Hindu University, Varanasi, India-221005. E-mail: asbhu2006@gmail.com.
[2] Prof., Department of Farm Engineering; Institute of Agricultural Sciences; Banaras Hindu University, Varanasi, India-221005. E-mail: gcmishrabhu@gmail.com.

(2006) provided an excellent review of time series methods in forecasting. Numerous studies have shown that this univariate method is very effective when compared to some other multivariate methods like linear regression and vector autoregressive models. The problem with ARIMA methodology is that it assumes a linear structure of the process the realization of which is a particular times series, which is often not correct. The other important aspect is that ARIMA methodology is only suitable under the assumption that the time series is stationary. To overcome this limitation of the ARIMA methodology, Artificial Neural Networks (ANN) have also been used to forecast the prices as shown by Kohzadi Nowrouz et al. (1996), Tang et al. (1991) and Zoua et al. (2007). This is because Artificial Neural Networks do not make any assumption about the process from which a particular time series has generated. Therefore, Artificial Neural Networks effectively cover both linear and non-linear processes, stationary as well as non-stationary time series. Neural Networks are now being used in wide domain of studies in areas as diverse as finance, medicine, engineering, geology and physics. This tremendous success of the Artificial Neural Networks can be attributed to some of its distinct character such as its power to model extremely complex function, in particular the non-linear functions. They can also handle the problem of parsimony in linear models. Combination of forecasts also increases the forecasting abilities of different methods as suggested in studies by Newbold et al. (1974), Zhang (2003). With the availability of sophisticated software, fitting of non-linear equations with the help of non-parametric methods has evolved to a new level. Neural Networks have been effective at forecasting and prediction in a variety of scenarios, Adya et al. (1998). Chen et al. (1992) and Park et al. (1991) found that for forecasting electric load ANN was better than traditional approaches. Tang et al. (1991) used ANN for forecasting car sales and airline passenger data and reported that ANN outperformed Box-Jenkins approach, both for short-term and long-term forecasting. Agricultural processes are affected by typical factors which are unique to this sector and prediction of prices of agricultural commodities is very difficult because they are not only governed by demand and supply also by so many other factors which are beyond control such as weather vagaries, storage capacity, transportation, etc. The performance of ANN in agricultural scenario is relatively less explored.

Therefore, in this paper time series of prices of Groundnut oil in Mumbai from January 1994 to July 2010 has been analyzed with both the ARIMA methodology and artificial neural networks and the forecasting abilities of both the models have been compared.

Rest of the paper is organized as follows - in Section 2 the traditional univariate time series approach to forecasting is described. In Section 3 the neural network architecture that is designed for this study is discussed. Section 4 discusses the evaluation methods for comparing the two forecasting approaches. Data and forecast procedure are discussed in Section 5. Moreover, section 5 shows the results obtained from the ARIMA and the Artificial Neural Network estimations. Section 6 contains a comparison of applied statistical measures and conclusions.

2. Auto Regressive Integrated Moving Average (ARIMA) time series model

Introduced by Box and Jenkins (1970), the ARIMA model has been one of the most popular approaches for forecasting. In the ARIMA model, the estimated value of a variable is supposed to be a linear combination of the past values and the past errors. Generally, a non-seasonal time series can be modelled as a combination of past values and errors, which can be denoted as ARIMA (p,d,q) which is expressed in the following form:

$$X_t = \theta_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots\ldots\ldots\ldots + \Phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots\ldots\ldots - \theta e_{t-q} \tag{1}$$

where X_t and e_t are the actual values and random error at time t, respectively, Φ_i (i = 1,2,…….,p) and θ_j (j = 1,2,……,q) are model parameters. p and q are integers and often referred to as orders of autoregressive and moving average polynomials respectively. Random errors e_t are assumed to be independently and identically distributed with mean zero and the constant variance, σ_e^2. Similarly, a seasonal model is represented by **ARIMA (p,d,q) x (P,D,Q)** model, where P denotes number of seasonal autoregressive (SAR) terms, D denotes number of seasonal differences, Q denotes number of seasonal moving average (SMA) terms. Basically, this method has three phases: model identification, parameters estimation and diagnostic checking.

The ARIMA model is basically a data oriented approach that is adapted from the structure of the data itself.

3. Artificial Neural Network (ANN) model

Neural Networks are simulated networks with interconnected simple processing neurons which aim to mimic the function of the brain central nervous system. ANN closely mimics functioning of the brain so its architecture is similar to that of the brain. A biological neuron has three types of components, namely dendrite, soma and axon. The dendrite accepts signals from other neurons which are electrical impulses transmitted through a synaptic gap with the help of certain chemical processes. A biological network is a collection of many biological neurons. Similarly, ANN is characterized by its architecture, i.e. the pattern of connections between the neurons, the method of determining the weights of the connections i.e. training or learning algorithm and its activation function. Mcculloch and Pitts (1943) for the first time proposed the idea of the artificial

neural network but because of the lack of computing facilities they were not in much use until the back propagation algorithm was discovered by Rumelhart et al. in 1986. Neural networks are good at input and output relationship modelling even for noisy data. The greatest advantage of a neural network is its ability to model complex non-linear relationship without a priori assumptions of the nature of the relationship. Apart from this, artificial neural networks can also be used for classification problems as was shown by Ripley (1994).
The ANN model performs a non-linear functional mapping from the past observations (X_{t-1} , X_{t-2},.……….., X_{t-p}) to the future value X_t i.e.

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}, w) + e_t \tag{2}$$

where w is a vector of all parameters and f is a function determined by the network structure and connection weights.

Training of the Neural Network is an essential factor for the success of the neural networks and among the several learning algorithms available, back propagation has been the most popular and most widely implemented learning algorithm of all neural networks paradigms. The important task of the ANN modelling for a time series is to choose an appropriate number of hidden nodes, q, as well as the dimensions of the input vector p (the lagged observations). However, in practice the choices of q and p are difficult.

4. Criteria for comparing the prediction accuracy of ARIMA and ANN procedures

Different criteria will be used to make comparisons between the forecasting ability of the ARIMA time series models and the neural network models. The first criterion is the absolute mean error (AME). It is a measure of average error for each point forecast made by the two methods. AME is given by

$$AME = \left(1/_T\right) \sum |P_t - A_t| \tag{3}$$

The second criterion is the mean absolute percent error (MAPE). It is similar to AME except that the error is measured in percentage terms, and therefore allows comparisons in units which are different.

The third criterion is the mean square error (MSE) which measures the overall performance of a model. The formula for MSE is

$$\text{MSE} = \left(1/_T\right) \sum (P_t - A_t)^2 \tag{4}$$

where P_t is the predicted value for time t, A_t is the actual value at time t and T is the number of predictions and the fourth criterion is RMSE which is the square root of MSE.

5. Results

Monthly cash prices of groundnut oil in Mumbai from April 1994 to July 2010 are used to test the prediction power of the two approaches. Data are obtained from the official Website of Ministry of Consumer Affairs. An ARIMA model was estimated. The model was then used to forecast on its respective three month out-of-sample set.

In the case of the neural networks, the time series was divided into a training, testing, and a validation (out-of-sample) set. The out-of-sample period was identical to the ARIMA model. SPSS (Statistical package for social sciences) was used to analyze the data and to carry out the calculations.

5.1. ARIMA time series results

Data is first differenced in order to remove the trend and the ARIMA estimated. For estimating the ARIMA model the three stages of modelling as suggested by Box and Jenkins namely identification, estimation and diagnostic checking were undertaken. Identification was done after examining the autocorrelation function and the partial autocorrelation function. After that, estimation of the model was done by the least square method. In the diagnostic checking phase the model residual analysis was performed.

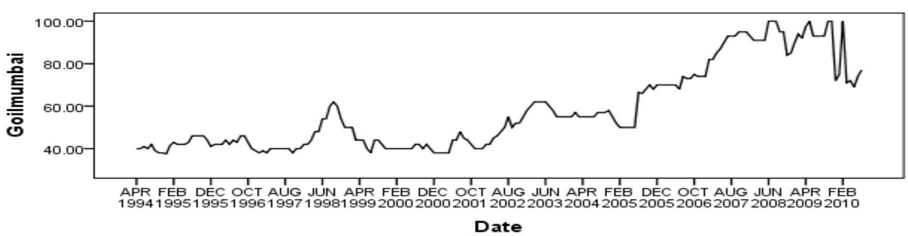

Figure 1. The time plot of prices of the Groundnut oil in Mumbai

In Figure 1 the time plot prices of the Groundnut oil in Mumbai is given. By looking at the graph it can be inferred that the series is not stationary because the mean of the time series is increasing with the increase in time. However, to confirm this autocorrelation function was also observed.

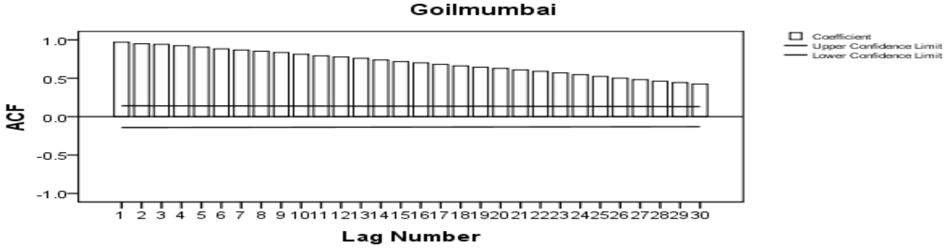

Figure 2. The autocorrelation function of the time series

In Figure 2 the autocorrelation function of the time series is shown. It certainly shows that the series is not stationary because autocorrelation coefficient does not cut off to statistical insignificance enough quickly which is caused by the fact that autocorrelations are significantly greater than the $\pm\ 2/\sqrt{N}$ confidence limits at 5% level of significance up to the 30^{th} lag. To make the series stationary it was differenced.

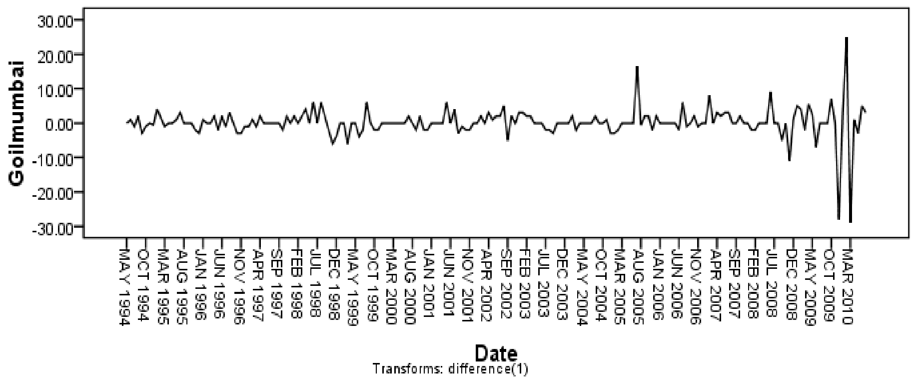

Figure 3. The time plot of the differenced series

In Figure 3 the time plot of the differenced series is given. It clearly shows that the series has now become mean stationary. However, it is not a variance stationary since the variance of the data around the mean of the differenced series in the end is greater than the rest of the series. Therefore, log transformation of the data was done.

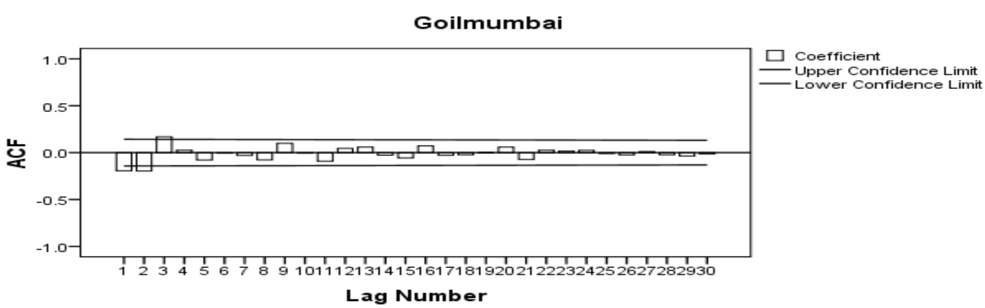

Figure 4. Autocorrelation function (ACF) of the differenced series

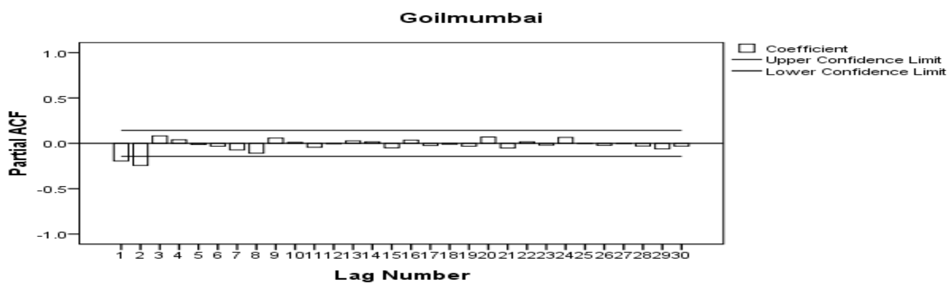

Figure 5. Partial autocorrelation function (PACF) of the differenced series

In Figure 4 and Figure 5 autocorrelation function (ACF) and partial autocorrelation function (PACF) of the differenced series are shown. Various ARIMA models were fitted using the expert modeler option in the SPSS software and after going through these stages the ARIMA (0,1,0) (1,0,1) model was found to be the best among the family of ARIMA models. ARIMA model parameters and model fit statistics are given in Table 1. The estimates of both the AR seasonal Lag 1 and MA seasonal Lag 1 were found to be statistically significant.

Table 1. ARIMA model parameters and model fit statistics

	Estimate	SE	t	Sig	Model Fit Statistics	
Differencing	1				Stationary R Squared	0.041
					R Squared	0.951
AR Seasonal Lag 1	0.990	0.091	10.841	0.00	RMSE	4.327
					MAPE	3.707
MA Seasonal Lag 1	0.953	0.231	4.127	0.00	MAE	2.215
					Normalized BIC	2.985

At the diagnostic checking stage residuals were examined and the autocorrelation coefficients were found to be non-significant (Figure 6).This shows that the model is satisfactory.

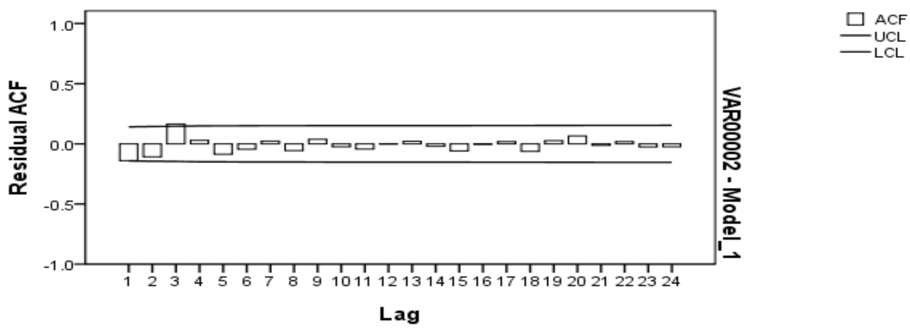

Figure 6. Autocorrelation function of the residuals

5.2. Neural network results

A feed forward neural network was fitted to the data, where values of the time series at first, second and third lag were taken as independent variables and the value to be forecasted was the dependent variable. The data was divided into 3 sets training, testing and hold out. In Table 2 it is shown that 81.6% observations were used for training, 16.8% for testing and 1.5% for forecasting. The training set was used for the estimation of the weights in the neural network and then predictions were made in the testing set. On the basis of the error in the testing set, the weights of the neural network were again adjusted to minimize the errors in the testing set.

Table 2. Case processing summary

		N	Percent
Sample	Training	160	81.6%
	Testing	33	16.8%
	Holdout	3	1.5%
Valid		196	100.0%
Excluded		0	
Total		196	

The information about the neural network architecture is given in Table 3. It shows that the network has an input layer, a single hidden layer and an output layer. In the hidden layer there is 1 unit and the activation function used is the hyperbolic tangent.

Table 3. Network architecture

Input layer	Covariates No. of units Rescaling methods of covariates	Lag1, Lag2, Lag3 3 Standardized
Hidden Layers	No. of hidden layers No. of units in hidden layers Activation Function	1 1 Hyperbolic tangent
Output Layer	Dependent variables Number of units Rescaling methods for scale dependents Activation function Error function	1 1 Standardized Identity Sum of squares

The architecture of the network has been shown in Figure 7. Light colour lines show weights greater than zero and the dark colour lines show weight less than zero.

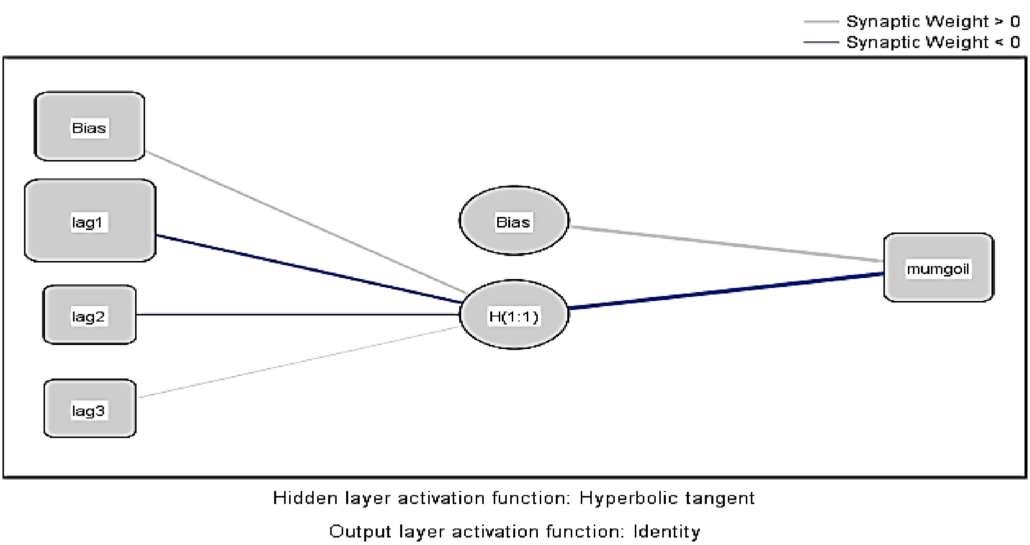

Figure 7. The architecture of the network

The training summary and the fit statistics for the training, testing and the holdout sets are given in Table 4.

Table 4. Model summary

Training	Sum of squares error Relative error Stopping rule used	3.480 0.044 Maximum number of epochs(100000) Exceeded
Testing	Sum of squares error Relative error	7.253 1.048
Holdout	Relative error	0.291

The estimates of the weights and bias are given in Table 5. The results in this table show the value of weights from input to the hidden layer and from the hidden layer to the output layer. H (1:1) means hidden layer 1 and 1St neuron. The weight attached to the neuron from bias is .354, from lag 1 is -.356 from lag 2 is -.045 and from lag 3 is .042.

The weights from the hidden layer to the output layer for bias 1.024 and from 1st neuron in the hidden layer to the output is -3.188.

Table 5. Parameter estimates

Predictor		Predicted	
		Hidden Layer 1	Output Layer
		H(1:1)	mumgoil
Input Layer	(Bias)	.354	
	lag1	-.356	
	lag2	-.045	
	lag3	.042	
Hidden Layer 1	(Bias)		1.024
	H(1:1)		-3.188

The observed values and the predicted graph in the Figure 8 show that except for few outliers it is a straight line.

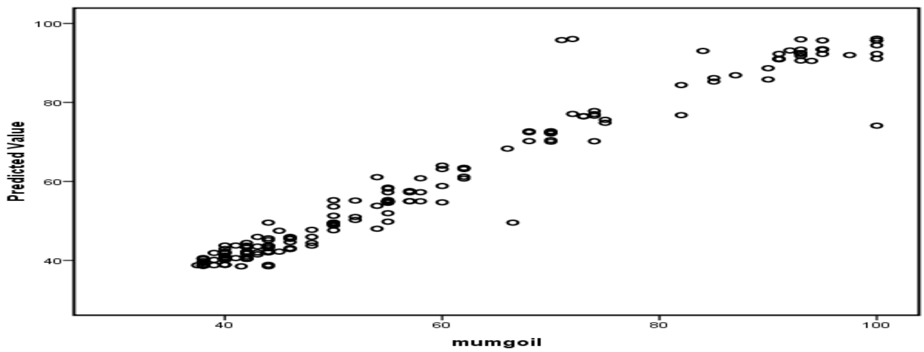

Figure 8. The observed values and the predicted values

It indicates almost one to one correspondence among the observed and predicted values. Hence, it can be inferred that the performance of ANN is satisfactory.

The residual and predicted chart (Figure 9) also shows that the residual does not follow a definite pattern and therefore is not correlated. If there is no dependence among the residuals then they can be regarded as observations of independent random variables and show that the ANN is satisfactory.

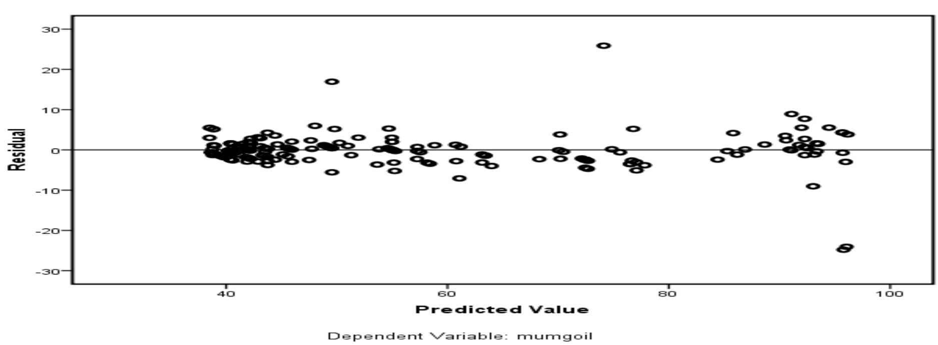

Figure 9. The residual and the predicted

6. Comparison of the accuracy of models and conclusions

The ARIMA and ANN models were compared for their forecasting capabilities with the help of RMSE and MSE. The results are shown below in Table 6.

The one step ahead forecast for May 2010 (69) was best predicted by ANN model (70.63) followed by the forecast of the ARIMA model (72.63).

The two step ahead forecast for June 2010 (74) was best predicted by ARIMA model (73.24) followed by the forecast of the ANN model (71.46).

The three steps ahead forecast for July 2010 (77) was best predicted by ANN model (76.36) followed by the forecast of the ARIMA model (74.91)

Overall, the forecast by ANN model was found to be the best predicted with MAPE (2.21), RMSE (3.09), MSE (9.52) followed by the forecast by the ARIMA model with MAPE (3.00), RMSE (4.26), MSE (18.12).

Table 6. Comparison of the accuracy of models

Observed		Predicted	
		ARIMA	ANN
May 2010	69	72.63	70.63
June 2010	74	73.24	71.46
July2010	77	74.91	76.36
	MSE	18.12	9.52
	RMSE	4.26	3.09
	MAPE	3.00	2.21

Artificial neural networks performed considerably better than the ARIMA models showing the forecasting ability and accuracy of this approach. The mean squared error (MSE), root mean square error (RMSE) and mean absolute percent error (MAPE) were all lower on average for the neural network forecast than for the ARIMA. The reason the neural network model performed better than the ARIMA may be because the data shows chaotic behaviour, which cannot be fully captured by the linear ARIMA model. Finally, the neural network results conform to the theoretical proofs that a feed forward neural network with only one hidden layer can precisely and satisfactorily approximate any continuous function.

REFERENCES

ADYA, M., COLLOPY, F., (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. Journal of Forecasting, 17, 481−495.

BOX, G. E. P., JENKINS, G. M., (1976). Time Series Analysis: Forecasting and Control. revised ed. Holden-Day, San Francisco.

CHEN, S. T., YU, D. C., MOGHADDAMJO, A. R., (1992). Weather sensitive short-term load forecasting using nonfully connected artificial neural network. IEEE Transactions on Power Systems, 7, 3,1098−1105.

DE GOOIJER, J. G., HYNDMAN, R. J., (2006). 25 years of time series forecasting. International Journal of Forecasting. Elsevier, Vol. 22(3), 443−473.

KOHZADI, N., BOYD, M. S., KAASTRA, I., KERMANSHAHI, B. S., (1996). A comparison of artificial neural network and time series models for forecasting commodity price Neurocomputing, 10, 169−18.

MCCULLOCH, W. S., PITTS, W., (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5, 115−133.

NEWBOLD, P., GRANGER C. W. J., (1974). Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society, A 137, 131−165.

PARK, D. C., EL-SHARKAWI, M. A., MARKS II, R. J., ATLAS, L. E., DAMBORG, M. J., (1991). Electric load forecasting using an artificial neural network, IEEE Transactions on Power Systems, 62, 442−449.

RIPLEY, B., (1994). Neural Networks and Related Methods for Classification (with discussion). Journal of the Royal Statistical Society, B, 56, 409−456.

RUMELHART, D. E., HINTON, G. E., WILLIAMS, R. J., (1986). Learning Internal Representations by Error Propagation, in Parallel Distributed Processing: Exploration in the Microstructure of Cognition. Cambridge, MA: MIT Press., Vol. 1, 318−362.

TANG, Z., DE ALMEIDA, C., FISHCWICK, P. A., (1991). Time series forecasting using neural networks vs. Box Jenkins methodology. Simulation,57, 5, 303−310.

Website of Ministry of Consumer Affairs:
http://fcainfoweb.nic.in/PMSver2/Reports/Report_Menu_web.aspx

ZHANG, G. P., (2003). Times series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159−75.

ZOUA, H. F., XIAA, G. P., YANGC, F. T., WANGA, H. Y., (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. Neurocomputing, 70, 2913−2923.