

Original Research Article

<https://doi.org/10.20546/ijcmas.2017.607.207>

## Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India

K. Naveena, Subedar Singh, Santosha Rathod\* and Abhishek Singh

Institute of Agricultural Sciences; BHU, Varanasi-221005, ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012, India

\*Corresponding author

### ABSTRACT

#### Keywords

Indian Robusta Coffee, ARIMA, ANN, Hybrid ARIMA-ANN, Forecasting.

#### Article Info

Accepted:  
19 June 2017  
Available Online:  
10 July 2017

Indian Robusta Coffee has made a slot for itself in the world market, particularly for its decent blend up quality. In India production of Robusta is more *i.e.* around 62–65%. Indian coffee prices are often random as they are largely inclined on production, demand of coffee in domestic and world level forces, etc. In this study Hybrid ARIMA-ANN models was compared with ARIMA and ANN model to evaluate the past behaviour of a time series data, in order to make inferences about its future behaviour for Robusta species of Indian coffee. Finally, the forecasting performance of these models are evaluated and compared by using common criteria's such as; Root Mean Square Error, Mean Absolute Percentage Error. Key findings reveal the superiority of Hybrid ARIMA-ANN model than in other Models, for forecasting of Indian Robusta coffee price.

### Introduction

Coffee is one of the world's most popular beverages. Among Plantation crops, Coffee has made significant contribution to Indian economy during the last 50 years. Indian Coffee has created a slot for itself in the global market; particularly Indian Robusta which is highly preferred for its good blend up quality. In India production of Robusta is more *i.e.* around 62- 65%, whereas production of Arabica is considered around 35- 38%.

Indian coffee prices are often random as they are largely influenced by production, demand of coffee in domestic and world level forces, quality of product etc. This leads to a

considerable risk and uncertainty in the process of price modelling and forecasting. Forecasting is used to provide an assistance to decision-making for the future effectively and efficiently. It is important aspect for a developing economy so that adequate planning is undertaken for sustainable growth, overall development and poverty alleviation. Statistical forecasting models are used to develop an appropriate forecast methodology by using the past data to predict the future with the help of identifying the trends and patterns within the data.

Generally agricultural data which are known to be multifaceted and often non-linear, so in

this study ARIMA, ANN time series models and hybrid of both ARIMA and ANN models were used to analyse the past behaviour of a Robusta species of Indian coffee time in order to make inferences about its future behaviour.

**The data**

The monthly wholesale price (Rs/Kg of clean coffee seeds) of Indian Robusta coffee seeds from January, 1995 to February, 2015 was recorded (Source; Bangalore market).

The Secondary data for the study were recorded from various published sources including coffee board, Bangalore. The data from March, 1995 to February, 2015 are used model building i.e. training data set and data from Dec, 2015 to February, 2016 are used for model validation i.e. testing data set.

**Materials and Methods**

Time series forecasting is a very useful technique for forecasting prices of agricultural commodities. Generally agricultural data contain both linear and nonlinear patterns, no single model is capable to identify all the characteristics of time series data on agriculture. Consequently, various types of parametric and nonparametric, linear and nonlinear time series models are used for forecasting (Fan and Yao, 2003, Ghosh et al., 2005). Time series forecasting models employed in the present study are described as below;

**Autoregressive Integrated Moving Average process (ARIMA)**

ARIMA is one of the most traditional methods of non-stationary time series analysis. In contrast to the regression models, the ARIMA model allows  $r_t$  to be explained by its past, or lagged values and stochastic error terms. An ARIMA model is usually

stated as ARIMA (p, d, and q). An autoregressive integrated moving average is expressed in the form:

If  $w_t = \nabla^d r_t = (1 - B)^d r_t$  then

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \tag{3.27}$$

Seasonal ARIMA model is denoted by (p, d, q) (P, D, Q), where p denotes the number of autoregressive terms, q, number of moving average terms and d, number of times a series must be differenced to induce stationarity. P, number of seasonal autoregressive components, Q, number of seasonal moving average terms and D denotes the number of seasonal differences required to induce stationarity (Box and Jenkins, 1994; Brockwell and Davis, 1996).

**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 (McCulloch and Pitts, 1943). The ANN structure for a particular problem in time series prediction includes determination of number of layers and total number of nodes in each layer. It is usually determined through experimentation as there is no theoretical basis for determining these parameters.

A single hidden layer feed forward ANN with one output node is most commonly used in forecasting applications [1, 2]. ANN model of  $p \times q \times 1$  is

$$y_t = \phi_0 + \sum_{j=1}^q \phi_j g + \left[ \theta_{0j} + \sum_{i=1}^p \theta_{ij} y_{t-i} \right] + \varepsilon_t$$

Here,  $j$  ( $j = 0, 1, 2, \dots, q$ ),  $ij$  ( $i = 0, 1, 2, \dots, p$ ;  $j = 1, 2, \dots, q$ )

are the weights,  $0, 0_j$  are the bias terms, and  $\varepsilon_t$  is the white noise.

### Zhang's hybrid model

When the data under consideration contains both linear and nonlinear components, neither ARIMA nor ANN is universally suitable for all types of time series.

Under such conditions model which incorporates both linear and nonlinear components are advisable.

Zhang, 2003 has pointed out this important fact and developed a hybrid approach that applies ARIMA and ANN separately for modelling linear and nonlinear components of a time series. According to Zhang, we have:

$$y_t = L_t + N_t \quad (3)$$

Where,  $y_t$  is the observation at time  $t$  and  $L_t, N_t$  denote linear and nonlinear components respectively at time  $t$ . At first, ARIMA is fitted to the linear component and the corresponding forecast  $\hat{L}_t$  at time  $t$  is obtained. So, the residual at time  $t$  is given by  $e_t = y_t - \hat{L}_t$

According to Zhang, the residuals dataset after fitting ARIMA contains only nonlinear component and so can be properly modelled through an ANN. Using  $p$  input nodes, the ANN for residuals has the following form:  $e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-p}) + \varepsilon_t$ , where  $f$  is a nonlinear function, estimated by the ANN and  $\varepsilon_t$  is the white noise. If  $\hat{N}_t$  is the forecast of this ANN, then the ultimate hybrid forecast at time  $t$  is obtained as;

$$\hat{y}_t = \hat{L}_t + \hat{N}_t$$

The most popular forecasting evaluation methods like root mean squared error (RMSE), and Mean Absolute Percentage

Error (MAPE) were used to evaluate above models.

### Results and Discussion

The price series on Robusta coffee covered monthly data from January, 1995 to February, 2016 (3 month dataset used for validation of result). These series vary from 29 to 155 Rs/Kg so it illustrates the complexity and variation of Robusta coffee price. Figure 1 shows the time series plot of average monthly price of Robusta coffee from January 1995 to February 2016. A perusal of Figure reveals a positive trend over time.

An ARIMA model was endeavored using the SPSS 16.0 statistical package. The model was then used to forecast 3 month out-of-sample set. Using Expert Modeler option in SPSS, the ARIMA model was estimated. After going through several stages ARIMA (0, 1, 1) (0, 0, 0) model was found to be the best among the family of ARIMA models. ARIMA Model parameters are given in table 1.

This model gratifies the invertibility condition and stationary condition and all the coefficients were found to be statistically significant at 1% level of significance. Also RMSE, MAE are 5.239, 3.339 respectively at the model fitting phase. The adequacy of the model was also decided based on the values of Box-Pierce Q statistics (15.143 i.e. Prob. value 0.585) it found to be non-significant. So, overall we can say ARIMA (0, 1, 1) (0, 0, 0) model shown satisfactory result, among different ARIMA models.

The information about the Neural network architecture shows that network has a input layer with two input nodes, a single hidden layer with 6 hidden node and a output layer with one output node means (2,6,1) Feed forward network. The activation function used is Sigmoidal at hidden layer and Linear

at output layer. The error is the sum-of-squares error because identity, activation function is applied to the output layer. Figure 2 represent the Actual v/s ANN fitted plot of Robusta coffee price time series. In the next step residuals are obtained from the fitted ARIMA model.

Figure reviles the ARIMA residuals plot of Robusta coffee price time series. The Brock, Dechert and Scheinkman (BDS) test Brock *et al.*, (1996) was employed to test the existence of nonlinearity. The results of the test in given in table 2 which indicate that nonlinear pattern exist in the residual data.

**Table.1** Estimate of the ARIMA Model parameter for Indian Robusta coffee price

	Estimate	SE	Test stat.	Sig.
Difference	1			
MA Lag1	-0.247	0.065	-3.767	0.000

**Table.2** Non linearity testing for ARIMA residuals of Robusta coffee price time series

Parameter	Dimension (m=2)		Dimension (m=3)	
	statistic	probability	statistic	probability
2.63	6.73	<0.001	8.20	<0.001
5.26	4.21	<0.001	54.22	<0.001
7.89	2.64	0.008	3.16	0.001
10.52	1.52	0.12	1.93	0.05

**Table.3** Forecasting performance of different models for Robusta coffee price time series in training data set

Criteria	ARIMA	ANN	ARIMA-TDNN
MAPE	4.02	3.65	2.96

**Table.4** Forecasting performance of different models for Robusta coffee price time series in testing data set

Year	Actual	Forecast		
		ARIMA	ANN	ARIMA -ANN
DEC-15	136	140.15	139.21	138.82
JAN-16	134	135.38	138.85	135.01
FEB-16	134	133.99	138.78	134.91
Criteria	MAPE	1.36	3.18	1.16

Fig.1 The time plot of Robusta price of Indian coffee

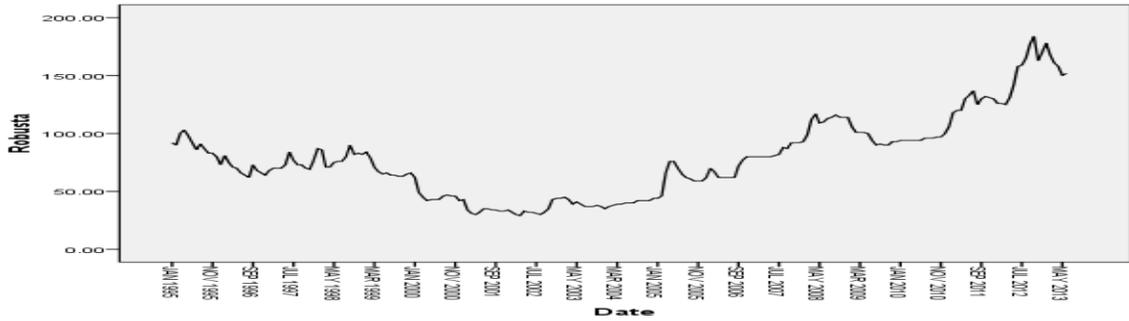


Fig.2 Actual v/s ANN fitted plot of Robusta coffee price time series

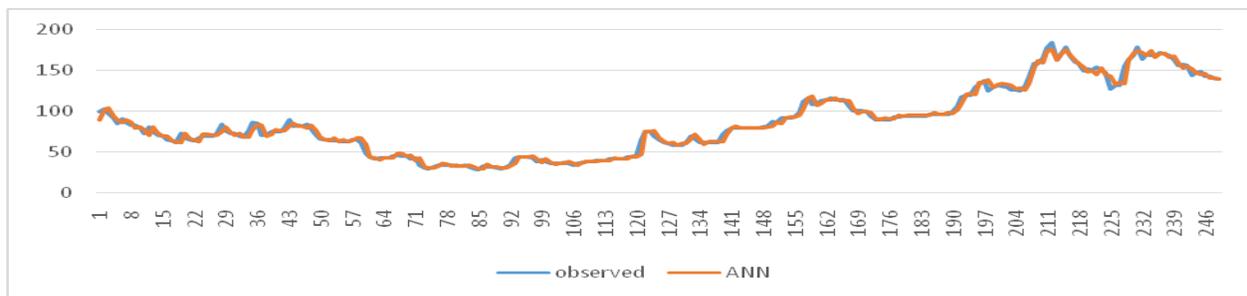


Fig.3 ARIMA residuals plot of Robusta coffee price time series

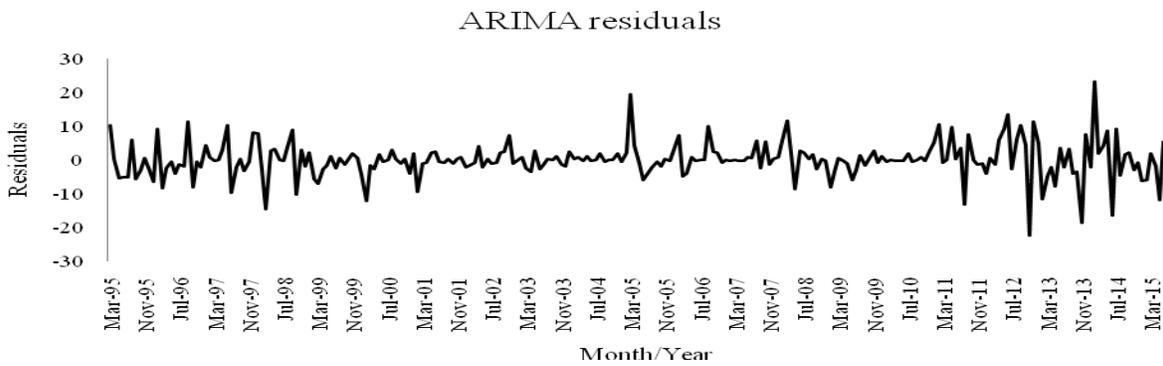
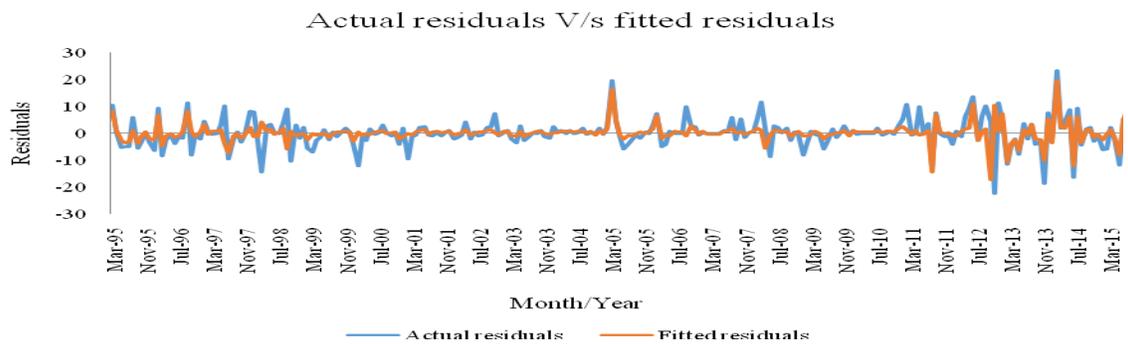


Fig.4 Actual v/s ANN fitted plot of ARIMA residuals of Robusta coffee price time series



ANN model specification for ARIMA residuals of Robusta coffee price time series, shows that network has a input layer with two input nodes, a single hidden layer with 6 hidden node and a output layer with one output node means (2,6,1) Feed forward network. The activation function used is Sigmoidal at hidden layer and Linear at output layer. The error is the sum-of-squares error because identity, activation function is applied to the output layer. Figure 3 represent the Actual v/s ANN fitted plot of ARIMA residuals of Robusta coffee price time series (Fig. 4).

Forecasting performance of different models for Robusta coffee price time series in training data set as given in table 3, shows minimum MAPE value.

To evaluate the forecasting performance last 3 observations of the considered time series was predicted employing the proposed approach. This approach was compared with the conventional ARIMA as well as Zhang hybrid approach (ARIMA-ANN). The results are given in table 4 as a test set.

The comparative results for the best ARIMA, ANN and ARIMA-ANN models are given in Table 5.0. MAPE statistic gives the indication of overall the superiority of ARIMA-ANN for forecasting of Indian Robusta coffee price.

In conclusion, the study has suggested that hybrid model of ARIMA and ANN has best model for Robusta coffee projection. The hybrid method which combines linear and non-linear models can be an effective way to improve fore-casting performance. Based on the results obtained in this work one can say a

hybrid model of ARIMA and ANN can increase forecasting accuracy plays a vital role in the adjustments of supply and demand in the future. It also help the government to make policies with regard to relative price and also to establish relations with other countries of the world by making proper export plan based on price variation in future.

## References

- Coffee Board, Indian Coffee (various issues), Bangalore.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. 1994. Time Series Analysis, Forecasting and Control, 3rd ed., Prentice Hall, Englewood Clifs.
- Brockwell, P.J. and Davis, R. A. 2002. Introduction to Time Series and Forecasting, 2nd. ed., Springer Verlag.
- Brock, W.A., Dechert, W.D., Scheinkman, J.A, lebaron, B. 1996. A test for independence based on the correlation dimension, *Econometric reviews*, 15:197-235.
- Tsay L.S 2005. Analysis of Financial time series, 2nd ed., Hoboken, N.J: Wiley
- Fan, J. and Yao, Q. 2003. Nonlinear time series: nonparametric and parametric methods, *Springer*, New York.
- Mcculloch, W.S. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophy.*, 5, 115-133.
- Zhang, G.,P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*50, 159-175.
- Ghosh, H., prajneshu and Paul, A, K. 2005. Study of nonlinear timeseries modelling in agriculture. IASRI. Project Report.

### How to cite this article:

Naveena, K., Subedar Singh, Santosha Rathod and Abhishek Singh. 2017. Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India. *Int.J.Curr.Microbiol.App.Sci.* 6(7): 1721-1726. doi: <https://doi.org/10.20546/ijemas.2017.607.207>