Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 6 Special issue [2] 2017: 433-440

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.533
Universal Impact Factor 0.9804
NAAS Rating 4.95

FULL LENGTH ARTICLE **OPEN ACCESS**

Evaluation of univariate time series models for forecasting of coffee export in India

Naveena K[1], Subedar singh and Abhishek Singh
Department of Farm Engineering; Institute of Agricultural Sciences; BHU, Varanasi, UP
[1]E-mail: naveenkagristats@gmail.com

ABSTRACT

Coffee is extremely important in the economies of many countries and in the world trade. Indian coffee is known for its quality of seeds specially Robusta coffee. It is being used for the purpose of blending with the coffee of other origins. About 60 to 75 percent of the Indian production goes towards export over 88 countries of the world. Due to restriction from ICA and incessant increase in coffee production has given rise to great fluctuation with respect to export. So producers of Indian coffee were greatly affected and Political, economic disruption has taken place. Hence forecasting of the export of Indian coffee, appears to be a solution to this problem. So, in this study both linear, nonlinear parametric and nonparametric time series models like ARIMA and ANN are used to analyse the past behaviour of a time series data, in order to make inferences about its future behaviour for Indian coffee export.
Keywords: *Coffee, Artificial Neural Network, ARIMA, Export.*

Received 02.07.2017 Revised 19.08.2017 Accepted 29.08.2017

INTRODUCTION

Coffee is one of the world's most popular beverages. Some claim it is the most widely consumed liquid in the world aside from water. This one having great significance in world economy as the largest single commodity entering the international trade after petroleum and petroleum products. It's farmed in 80 countries and exported by over 50 countries in central and South America, Africa and Asia. More than 100 million people are engaged in produce and processing of coffee. It is estimated that more than 20 million people throughout the world earn their living from coffee. A majority of them are involved in its production and 40 percent of them consume coffee on a regular basis. Thus, coffee is extremely important in the economies of many countries and in the world trade in general.

Indian coffee is known for its quality and is exported to almost all countries in the world. It is being used for the purpose of blending with the coffee of other origins. On the export front, India cannot remain isolated from the global stream, although India's production accounts for 3.65 per cent of the global production and 4.62 percent of the total world exports in 2012-13. About 60 to 75 percent of the Indian production goes towards export over 88 countries of the world (coffee board). Export of coffee was regulated by a system of international coffee agreement (ICA) quotas. Range of price and fixed quantities were allotted to member countries, based on exportable production and previous performance in order to achieve a reasonable balance between supply and demand on a basis which would ensure adequate supplies of coffee to consumers.

Due to this restriction from ICA and incessant increase in coffee production has given rise to great fluctuation with respect to export. During the period 1961-72, India's share remained broadly constant in the range of 0.8 per cent to 1.2 per cent. During 1973-1979, the country's share marginally improved but fluctuated in the range of 1.2 per cent to 1.6 per cent. A Further increase to about 2 per cent can be noticed during the early 1980, but then the share was declined and remained mostly below 2 per cent until 1992. The first half of the 1990s witnessed a significant increase in India's market share. The share has been declining since 1995 but remains above 2 per cent for most of the years (Veeramani, 2012). Because of this fluctuation in export, producers of Indian coffee were greatly affected. Political and economic disruption has taken place. Thus forecasting of the export of Indian coffee, appears to be a solution to this problem. Generally agricultural data often contain both linear and nonlinear patterns, so no single model is capable to identify all the characteristics of time series data. So, in this study both linear, nonlinear parametric and nonparametric time series models like ARIMA and ANN are used to analyse the past behaviour of a time series data, in order to make inferences about its future behaviour for Indian coffee export.

The data

The data for this study are annual export quantity (thousand tonnes) of coffee seeds in India (Source; Bangalore market). The Secondary data for the study were recorded from various published sources including coffee board, Bangalore.

MATERIAL AND METHODS

Auto Regressive Integrated Moving Average (ARIMA) Model:

The ARIMA model has been one of the most popular approaches for forecasting, introduced by Box and Jenkins (1970). The main objective in fitting this ARIMA model is to identify the stochastic process of the time series and predict the future values accurately. These methods have also been useful in many types of situation which involve the building of models for discrete time series and dynamic systems. In general, an ARIMA model is characterized by the notation ARIMA (p, d, q), where *p*, *d* and *q* denote orders of Auto-Regression (AR), Integration (differencing) and Moving Average (MA), respectively. Which can be denoted as ARIMA (p,d,q) which is expressed in the following form:

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

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.. Random errors e_t are assumed to be independently and identically distributed with mean zero and the constant variance, σ_e^2.

The first step in the process of ARIMA modelling is to check for the stationarity of the series as the estimation procedure is available only for a stationary series. A series is regarded stationary if its statistical characteristics such as the mean and the variance are constant over time. It can be achieve by differencing, while Logarithmic transformation is employed to stabilize the variance. After appropriate transformation and differencing, multiple ARMA models are chosen on the basis of Auto-Correlation Function (ACF) and Partial Auto- Correlation Function (PACF) that closely fit the data. This is not a hard and fast rule, as sample autocorrelation coefficients are poor estimates of population autocorrelation coefficients. Still they can be used as initial values while the final models are achieved after going through the stages repeatedly. Then, the parameters of the tentative models are estimated by method of least squares is used to estimate the parameters which utilizes the concept of minimizing the sum of squares due to residuals. Lastly, diagnostic checking for model adequacy is performed for all the estimated models through the plot of residual ACF and using Ljung-Box text. The most suitable ARIMA model is selected based on following characteristics like It is parsimonious (uses the smallest number of coefficients needed to explain the given data), It is stationary (has *AR* coefficients which satisfy some mathematical inequalities), It is invertible (has MA coefficients which satisfy some mathematical inequalities), It has uncorrelated residuals*.* The lowest root mean square error (RMSE) and Mean absolute percent error (MAPE). In this study, all estimations and forecasting of ARIMA model have been done using SPSS 16 statistical package.

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). Neural networks are good at input and output relationship modeling even for noisy data. The greatest advantage of a neural network is its ability to model complex nonlinear relationship without a priori assumptions of the nature of the relationship.

A neural network can be made active by setting either long-term or short-term memory, depending on the retention time, into the structure of a static network. One simple way of building short–term memory into the structure of a neural network is through the use of time delay, which can be implemented at the input layer of the neural network.

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. It has been proved that neural networks with one hidden layer can approximate any non-linear function given a sufficient number of nodes at the hidden layer and adequate data points for training. In this study, we have used neural network with one hidden layer. In time series analysis, the determination of number of input nodes which are lagged observations of the same variable plays a crucial role as it helps in modelling the autocorrelation structure of the data. The determination of number of output nodes is relatively easy. In this study, one output node has been used.

The most popular form of neural network architecture is the multilayer perceptron (MLP). It has any number of inputs, one or more hidden layers with any number of units, any number of outputs with any activation function connections between the input layers. The important aspect of artificial neural Network is its learning and training algorithm which is a procedure for updating and determining weights of the connections. Learning algorithm is a procedure of modifying weights and biases of network *i.e.* method of driving next changes that might be made in ANN, while Training algorithm is a procedure whereby network is actually adjusted to do a particular job. In the present study method used to estimate the synaptic weights is Scaled conjugate gradient. The assumptions that justify the use of conjugate gradient methods apply only to batch training types, so this method is not available for online or mini-batch training.

The most popular forecasting evaluation methods like root mean squared error (RMSE), and mean absolute percentage error (MAPE) are evaluate above models.

RESULT AND DISCUSSION

The Export series on Indian coffee covered yearly data from 1995-96 to 2012-13(3 year dataset used for validation of result).These series vary from 25.74 to 248.99 thousand tones, so it illustrate the variation of Indian coffee Export. The basic characteristics of the Export series used in the study are presented in Table 1

Table 1: Discriptive statistics of Indian Export series used in this study

	N	Range	Min.	Max.	Mean	Std. Deviation	Skewness	Kurtosis
Export	45	223.26	25.74	248.99	117.83	73.191	0.444	-1.277

ARIMA Model

An ARIMA model was endeavored using the SPSS 16.0 statistical package. The model was then used to forecast three years out-of-sample set. The first step in time series analysis is to plot the data. A perusal of Figure 1 reveals an increasing trend in the total export of Indian coffee over years, which indicates the non-stationary nature of series. It is confirmed, through the Auto correlation function (ACF) and partial autocorrelation function (PACF). The autocorrelations are significant for a large number of lags**,** showing a slow linear decay over a years (Figure 2) and PACF plot (Figure 3) has a significant spike only at lag 1, meaning that all the higher-order autocorrelations are effectively explained by the lag1, so it indicates non stationarity of time series. To make the series stationary it was first differenced.

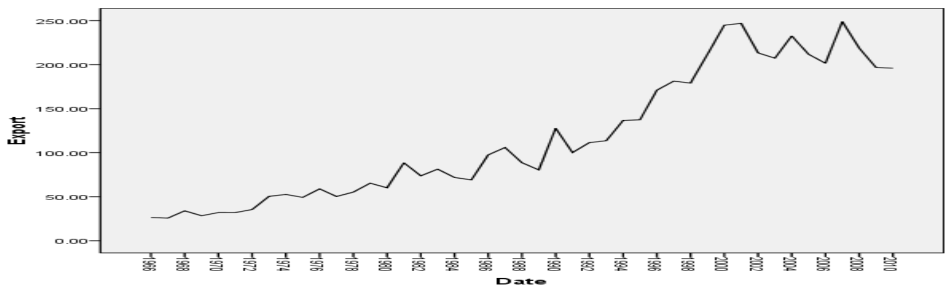

Fig 1: The time plot of Export of Indian coffee

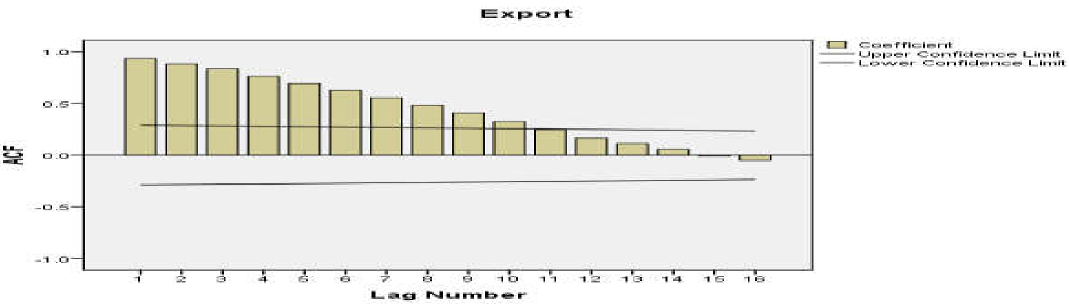

Fig 2: Autocorrelations at different lags ofIndian Export of Indian coffee

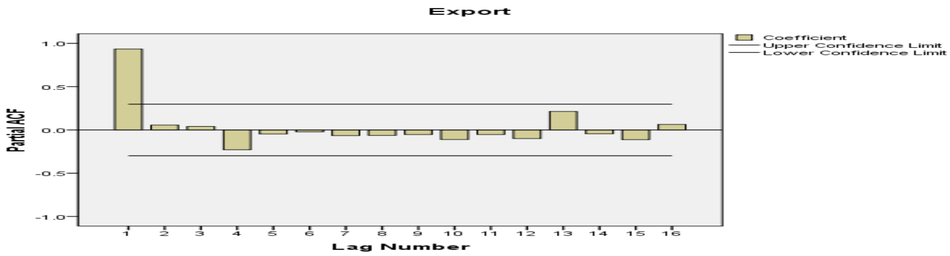

Fig 3: Partial Autocorrelations at different lags of Indian Export of Indian coffee

Fig 4 shows the time plot of the differenced series and it clearly indicates that the series has now become stationary. However, the judgment about stationarity was withheld until plotting differenced ACF and PACF plot.

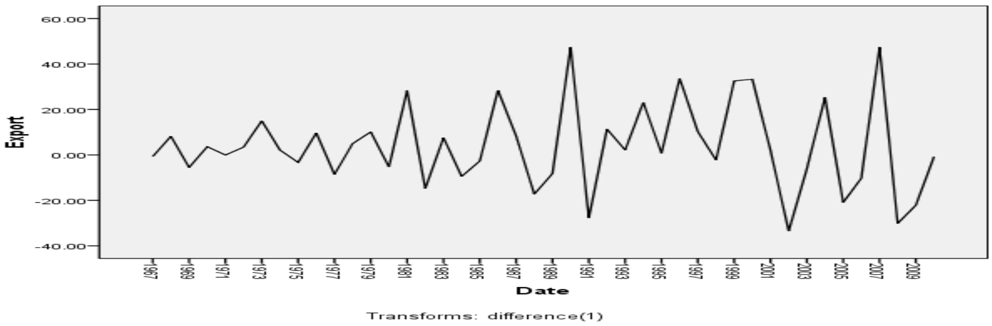

Fig 4: The time plot of the differenced series of Indian Export of Indian coffee

In fig 5 and fig 6 Autocorrelation function and Partial autocorrelation function of the differenced series are shown. Here in both the ACF and PACF plot most of all the lags are non-significant, so it was concluded that mean of the series is stationary.

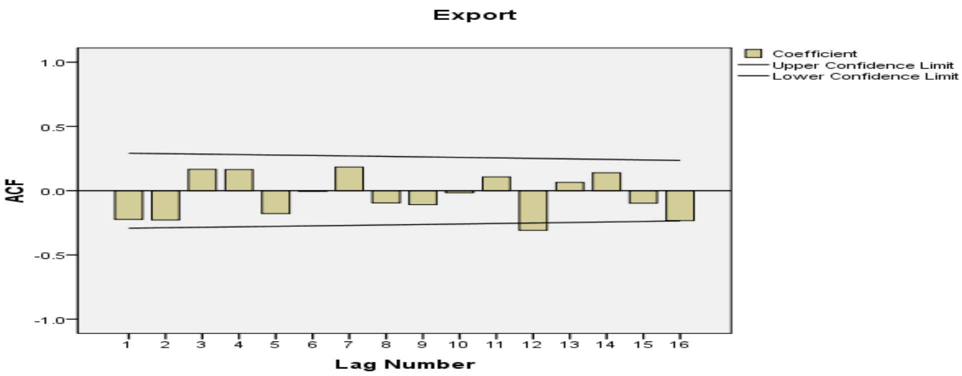

Fig 5: Autocorrelations at different lags of 1st differenced time series of Export of Indian coffee

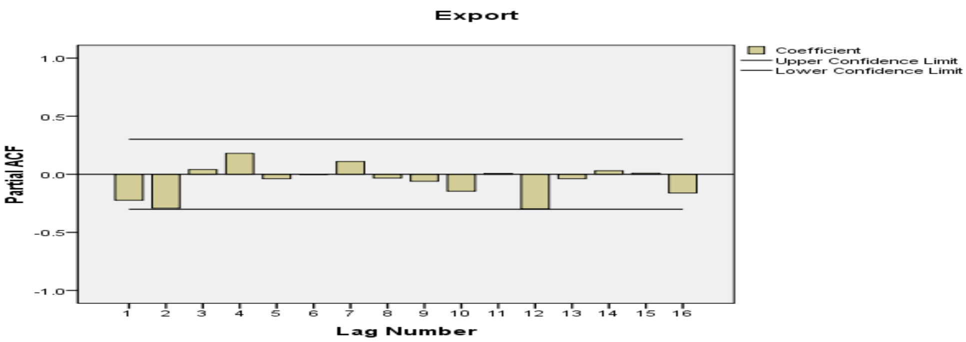

Fig 6: Partial Autocorrelations at different lags of 1st differenced time series of Export of Indian coffee

Once the time series has become stationary, Using Expert Modeler option in SPSS, the ARIMA model was estimated. In particular, the ACF has one significant spikes, while the PACF has no significant spike. Thus, the differenced series displays ARIMA (0,1, 1) model as best among the family of ARIMA models . Model parameters are given in Table-3

Table-2: Estimate of the ARIMA Model parameter of Indian coffee Export

	Estimate	**SE**	**Test stat.**	**Sig.**
Constant	0.047	0.012	3.953	0.000
Difference	1			
MA LAG1	0.487	0.144	3.391	0.002

In table 2. ARIMA (0,1,1) was found to be the best model. This model satisfies the invertibility requirement $|\theta_1| \leq 1$ and the moving average (MA) was found to be statistically significant at 1% and 5% level of significance. Model fit statistics in table 3.

Table -3: Model fit statistics and Ljung-Box Q statistics for Export of Indian coffee

Model Fit statistics						Ljung-Box Q(18)		
Stationary R-squared	R-squared	RMSE	MAPE	MAE	Normalized BIC	Statistics	DF	Sig.
0.195	0.919	20.987	12.530	15.282	6.260	13.847	17	0.678

Residual analysis was carried out to check the adequacy of the models. The residuals of ACF and PACF were obtained from the tentatively identified model, they were found most of all lags are to be non-significant (fig 7). So we can state that model is adequate.

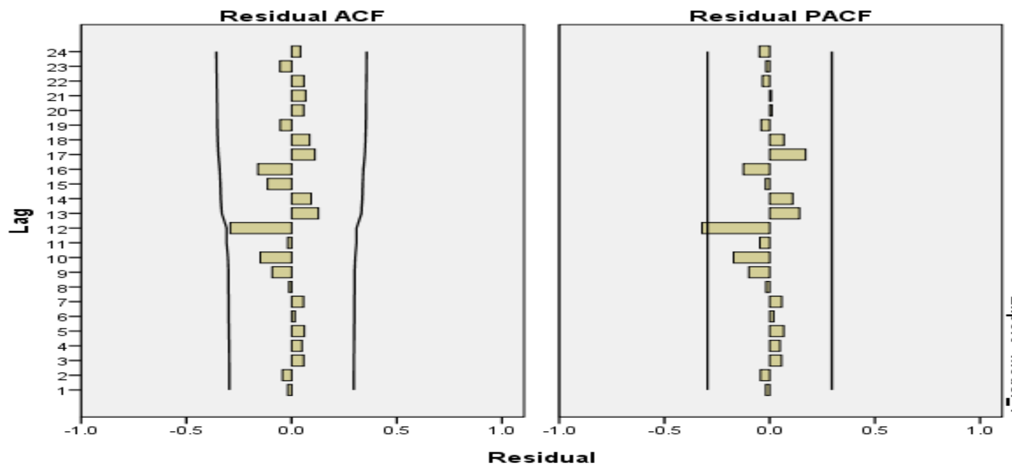

Fig 7: Residual auto and partial autocorrelations for Indian Export of Indian coffee

The adequacy of the model was also judged based on the values of Box-Pierce Q statistics in table 3, it found to be non-significant. Thus we can reviewed that ARIMA (0, 1, 1) model showing satisfactory, among different ARIMA models for Forecasting of export of Indian coffee.

Neural network Model

A Feed forward neural network was fitted to the data with the help of SPSS 16.0 statistical package, where we varied the number of input nodes from 1 to 3, the number of hidden nodes from 2 to 10. Thus, different numbers of neural network models are tried before arriving at the final structure of the model (Jha *et al*, 2009). Out of all neural network structures a neural network model with two input nodes and six hidden nodes (3:2s:1*i*) performed better than other competing models in respect of out-of sample forecasting for export series. Out of 46 cases, 33 cases were assigned to the training sample, 7 to the testing sample and 6 to hold out set. This is in conformity with the rules that 70%, 15%, and 15%, dividing the dataset into equal thirds among training, testing, and holdout (Prity Kumari *et al*, 2013).
The information about the Neural network architecture shows that network has a input layer with three input nodes, a single hidden layer with two hidden node and a output layer with one output node. The activation function used is the hyperbolic tangent. The error is the sum-of-squares error because identity, activation function is applied to the output layer. The architecture of the network has been shown in the fig 8 light color lines show weights greater than zero and the dark color lines show weight less than zero.

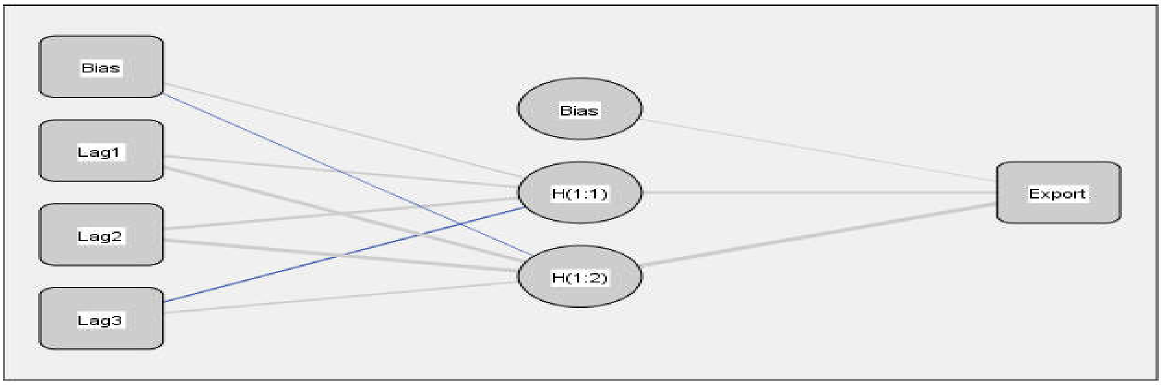

Fig 8: The architecture of the network fitted to time series of Export of Indian coffee

Table 5 displays information on the result of training and applying the network to the testing and holdout sample. Sum of squares error is the error function that the network tries to minimize during training. The relative error for export is the ratio of the sum of squares error for export variable to the sum of squares error for the "null" model in which the mean value of the export variable is used as the predicted values. There appeared to be 0.153 for training, 0.019 for testing and 0.636 for holdout set. Here there is one step to allow before checking for a decrease in error. With training time 00:00:00.012.

Table 5: The training summary and the fit statistics of ANN Indian coffee export

Training	Sum of Squares Error	2.442
	Relative Error	0.153
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	00:00:00.012
Testing	Sum of Squares Error	0.074
	Relative Error	0.019
Holdout	Relative Error	0.636

Dependent Variable: Export

a. Error computations are based on the testing sample.

Table 6: The estimates of the weights and Bias of ANN fitted Export of Indian coffee

		Predicted		
		Hidden Layer 1		Output Layer
Predictor		H(1:1)	H(1:2)	Export
Input Layer	(Bias)	0.162	-0.049	
	Lag1	0.284	0.519	
	Lag2	0.495	0.601	
	Lag3	-0.149	0.265	
Hidden Layer 1	(Bias)			0.084
	H(1:1)			0.410
	H(1:2)			1.058

The estimates of the weights and Bias are given in table 6. This table shows the value of weights from input to the hidden layer and from the hidden layer to the output layer. There are 6 weights from input nodes to hidden layer nodes and 2 weights from hidden layer nodes to output node.

Fig 9 displays scatter plot of predicted values on the y-axis by observed values on the x-axis for the combined training and testing samples. Which depicts that except for few outliers it is a straight line. It indicates almost one to one correspondence between the observed and predicted values. Hence it can be inferred that the performance of the ANN is satisfactory.

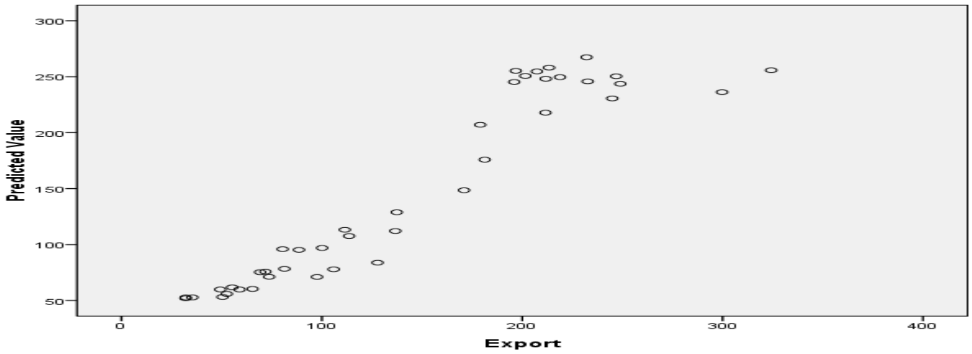

Fig 9: The observed vs the predicted graph of ANN of Indian Export of Indian coffee

The residual vs predicted chart (fig 10) also shows that the residual do not follow a definite pattern and therefore are not correlated.

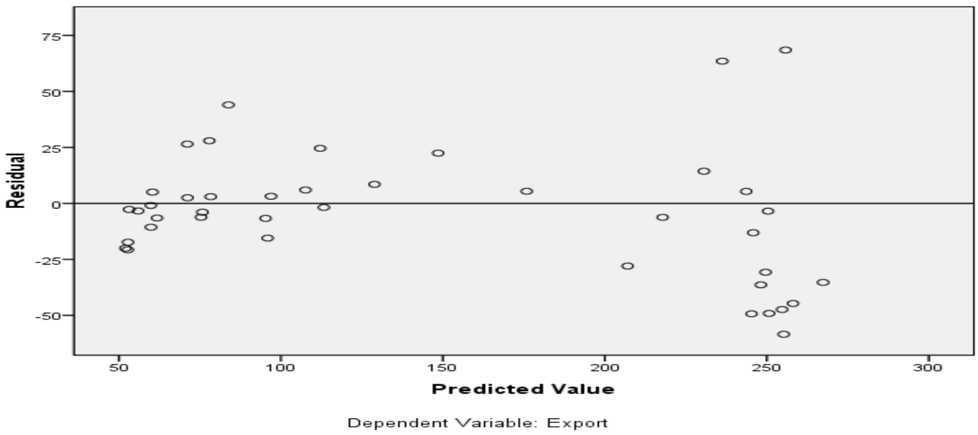

Fig 10: The residual vs the predicted graph of ANN of Indian Export of Indian coffee

Comparison

Table 7: Observed and predicted Exports of Indian Export of Indian coffee

Months	Observed (thousand ton)	Predicted (thousand ton)	
		ARIMA	ANN
2010-11	299.737	224.53	236.23
2011-12	324.253	236.12	255.77
2012-13	232.207	248.31	267.37
RMSE		67.534	57.640
MAPE		25.768	22.283

The forecasting ability of ARIMA and ANN models is assessed with respect to two common performance measures, viz. Root mean squared error (RMSE) and Mean absolute percentage error (MAPE). In this study, we have included the results for one year, two year, and three year ahead forecast. The comparative results for the best ARIMA and ANN models are given Table 7 As assessed by RMSE and MAPE statistic gives the indication of the superiority of ANN model than in ARIMA model respectively for forecasting of overall Indian coffee Export. Also for one step and two step head forecasting ANN gave best result. In case of three step ahead forecasting ARIMA is gave better results.

Higher Root Mean squared error and Mean absolute percentage error because, forecasted three years we can see very large fluctuation in export as compare to previous year export. From 2010-11 and 2011-12 there is great export because upsurge in coffee production specially Robusta coffee production (coffee statistics, 2013), whereas due to Sluggish demand in the European market on account of economic slowdown, incidence of pest attacks and an erratic monsoon have taken a toll on coffee exports as well as production in 2012-13th year (Nagesh Prabhu, Hindu paper 2013).

CONCLUSION

This paper has compared the ANN and ARIMA models in terms of both modelling and forecasting of annually export of coffee in India. The ANN model in has provided a better forecast accuracy in terms of conventional RMSE and MAPE values as compared to the ARIMA model. The study has suggested that best model for coffee export projection plays a vital role for government to make policies and also to establish relations with other countries of the world by making proper export plan.

REFERENCE

1. Box, G.E.P. and Jenkins G. 1970. Time series analysis, Forecasting and control, Holden-Day, San Francisco.
2. Coffee Board, Indian Coffee (various issues), Bangalore.
3. Girish K. Jha, Parimala Thulasiraman and Ruppa K. Thulasiram. 2009. PSO based neural network for time series forecasting. *In proceeding of the IEEE International Joint Conference on Neural Networks, US*.1422-1427.
4. McCulloch, W. S., and Pitts, W. A .1943. Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, **5:**115-133.
5. Nagesh Prabhu . Jan, 2013. Coffee exports, production decline. *The Hindu paper.*
6. Prity Kumari, G.C.Mishra and C.P. Srivastava. 2013. Forecasting of Productivity and Pod Damage by Helicoverpa armigera using Artificial Neural Network Model in Pigeonpea (Cajanus cajan). *International Journal of Agriculture, Environment & Biotechnology*, **6(2)**: 335-340.
7. Veeramani,c., 2012. Competitive structure of plantation commodity exports: India's tea and coffee exports in a comparative perspective. *Nrppd Discussion Paper:* 1-47.

CITATION OF THIS ARTICLE

Naveena K, Subedar singh and Abhishek Singh . Evaluation of univariate time series models for forecasting of coffee export in India . Bull. Env. Pharmacol. Life Sci., Vol 6 Special issue 2, 2017: 433-440