The Andhra Agric. J 69 (1): 137-144, 2022

Statistical Modelling on India Coffee Exports - A Time Series Approach

K Gnana Prakash, D Ramesh, V Srinivasa Rao and G Raghunadha Reddy
Department of Statistics and Computer Applications, Agricultural College, Bapatla, A. P.

ABSTRACT

The research paper is an attempt to model and forecast the exports of coffee from India by using annual time series data from 1971-2020. A comparative study was made among linear and non-linear regression models, Auto Regressive Integrated Moving Average (ARIMA) and Artificial Neural Network models (ANN) as to find out an appropriate model to capture the trend of coffee exports. The model selection criterion for this study were Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The Ljung-Box test was also tried to verify the adequacy of the selected model. Finally, NNAR (5-10-1) model was found as the most appropriate to capture the trend of coffee exports from India, and forecasts were estimated as 5502.49 (Cr.) for the year 2024.

Keywords: *Coffee, Forecasting and Regression.*

Coffee is the second most important commodity in the world trade after petroleum. Since the establishment of the first commercial coffee plantations in India in the 18th century, the country has made significant progress in terms of export orientation and commercial cultivation. Coffee contains antioxidants and nutrients that are beneficial to one's health. Coffee contains caffeine, which has been scientifically proven to improve mental performance. Caffeine provides a significant energy and mental boost, allowing us to face the day's challenges and also helps to burn fat and improve physical performance. In India, area under coffee cultivation was 459.98 thousand hectares during 2019-20 and it is the world's seventh largest coffee producer (after Brazil, Vietnam, Columbia, Indonesia, Ethiopia, and Honduras). Production and yield of Coffee were recorded as 2,98,000 tonnes, 647 kg per hectare respectively, and the exports earned an amount of Rs.5,199.41 crores in foreign exchange, in 2019. The present study was undertaken with the objective of analyse the model and forecast the future Coffee exports.

MATERIAL AND METHODS

Data with respect to exports of coffee from India for period of 1971-2020 was collected from Food and Agricultural Organisation (FAO STAT) and Indiastat website (MOA & FM, 2019). The analysis was conducted using R software version 4.2.1 to model and forecast coffee exports from India upto 2024. Before analysis, as the study is dealing with time series, present data set had been verified initially for existence of outlier.

Test for Outlier

For detecting the outlier in the time series, Grubbs' test was used in the current scenario as the test is particularly useful in case of large sample and easy to follow. Graph pad software was employed in the present study to identify the existence of outliers (Sahu, 2010).

Linear and Non-linear regression models.

In this study, various selected linear and nonlinear regression models namely Linear, Logarithmic, Inverse, Quadratic, Cubic, Power, S-curve and Exponential models were applied to the Coffee export (value) of India. The models were given as shown below.

S. No.	Name of the model	Model form
1	Linear model	$Y_t = a + bt + e_t$
2	Quadratic model	$Y_t = a + bt + ct^2 + e_t$
3	Cubic Model	$Y_t = a + bt + ct^2 + dt^3 + e_t$
4	Logarithmic Model	$Y_t = a + b \ln(t) + e_t$
5	Power Model	$Y_t = a\, t^b + e_t$
6	Inverse Model	$Y_t = a + b/t + e_t$
7	S Curve	$Y_t = \text{Exp}(a + b/t) + e_t$
8	Exponential	$Y_t = a*\text{Exp}(bt) + e_t$

Where,

'Y_t' is the dependent variable i.e., Export

'a' is the intercept,

't' is the independent variable, time in years

e_t is the error term,

'b', 'c' & 'd' are the regression coefficients

'Method of least squares' is used to estimate the parameters of the linear model, where Levenberg-Marquardt (LM) algorithms, or Does not Use Derivatives (DUD) procedures are to be employed for nonlinear models (Seber and Wild, 1989). In the present study SPSS package was used to estimate parameters of the selected models.

ARIMA Model

According to Box and Jenkins (1976), a non seasonal ARIMA model is denoted by ARIMA (p,d,q) which is a combination of Auto Regressive (AR) and Moving Average (MA) with an order of integration or differencing (d), where p and q are the order of autocorrelation and moving average respectively (Gujarati *et al.* 2012).

The Auto-regressive model of order p denoted by AR (p) is as follows:

$$Z_t = c + \emptyset_1 z_{t-1} + \emptyset_2 z_{t-2} + \ldots + \emptyset_p z_{t-p} + e_t$$

where,

c is constant term,

$\emptyset_p$ is the p-th autoregressive parameter and

e_t is the error term at time t.

The general Moving Average (MA) model of order q or MA (q) can be written as:

$$Z_t = c - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} + e_t$$

where,

c is constant term,

θ_q is the q-th moving average parameter and

e_{t-k} is the error term at time t-k.

ARIMA in general form is as follows:

$$\Delta^d Z_t = c + (\emptyset_1 \Delta^d Z_{t-1} + \ldots + \emptyset_p \Delta^d Z_{t-p}) - (\theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}) + e_t$$

where Δ denotes difference operator like

$\Delta Z_t = Z_t - Z_{t-1}$ (data form of first order differentiation)

$\Delta^2 Z_{t-1} = \Delta Z_t - \Delta Z_{t-1}$ (data form of second order differentiation)

Here, Z_{t-1}, …, Z_{t-p} are values of past series with lag 1, …, p respectively.

Here, Z_{t-1}, …, Z_{t-p} are values of past series with lag 1, …, p respectively.

Modeling using ARIMA methodology consists of four steps viz. model identification, model estimation, diagnostic checking and forecasting (Sankar, 2011).

Artificial Neural Network model (ANN)

The ANNs are generally constructed by layers of units i.e., artificial neurons or nodes, hence termed as multilayer ANNs, such that each unit in a layer performs a similar task. The very first layer consists of the input units, which are statistically known as the independent variables. Similarly, the last layer contains the output units, statistically known as the response or dependent variables. The rest of units in the model are known as the hidden units and constitute the hidden layers.

In the present study, single hidden layer with multilayer feed forward network was employed in developing ANN model, which is considered as the most popular for time series modeling and forecasting (Rathod *et al.*, 2017).This model is characterized by a network of three layers of simple processing units. The first layer is input layer, the middle layer is the hidden layer and the last layer is output layer, as shown in the Figure 1.

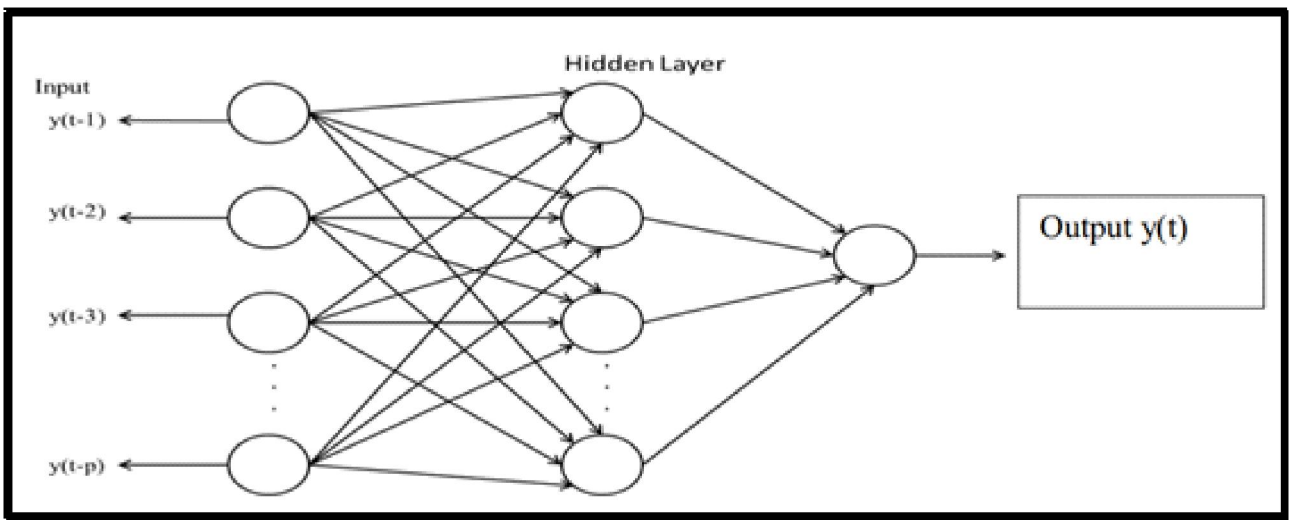

Figure 1: Architecture of ANN for time series forecasting

In this study, model performance among the selected ANN models was verified by diagnostic criterion namely MAPE, MAE and RMSE. In addition to this diagnostics, Ljung-Box test was also employed to verify the residual of series were independently distributed or not.

Comparison among the Selected Models

The appropriate model among the linear and non-linear regression, ANN and ARIMA model were further compared by the diagnostic criterion (RMSE & MAPE).

RESULTS AND DISCUSSION

The secondary data on, Coffee exports from India from 1971- 2020 was tested for any outliers by Grubbs method. It was observed that the number of extreme observations in the present data was zero, as presented in Table 1.

Table 1. Grubbs test for detecting Outliers

Mean:	1680.91
SD:	1911.54
No of observations:	50
Outlier detected?	NO

Linear and Non-linear regression models were applied to the dataset under consideration. From Table 2, it was concluded that the Exponential model was found to be superior to other selected regression models based on diagnostic criteria.

Table 2. Linear and Non-linear regression models for estimation of Coffee exports from India

Model	RMSE	MAPE	Fitted Equation
Linear	892.51	380.57	$Z_t = -1267.65 + 115.62t + e_t$
Logarithmic	1386.41	606.86	$Z_t = -2664.12 + 1463.19\ln(t) + e_t$
Inverse	1772.62	569.67	$Z_t = 2062.43 - 4239.92/t + e_t$
Quadratic	799.08	111.70	$Z_t = 489.22 - 87.08t + 3.97t^2 + e_t$
Cubic	872.88	39.23	$Z_t = 13.05 + 19.70t - 1.20t^2 + 0.06t^3 + e_t$
Power	1108.71	47.94	$Z_t = 5.13t^{1.64} + e_t$
S-curve	1925.80	131.86	$Z_t = \mathrm{Exp}(7.12 - 6.54/t) + e_t$
Exponential	746.46	29.64	$Z_t = 48.53\ \mathrm{Exp}(0.10^t) + e_t$

Box Jenkins (ARIMA) model

For employing the selected linear time series modelling, stationary of data series had examined first. For this, Augmented Dickey Fuller (ADF) test was applied to the data series and from Table 3, it was also found that the data series was non stationary and became stationary at first difference as the null hypothesis was not accepted at 5% LOS, as the p-value was 0.03 (<0.05).

After fixing the d=1, as per autocorrelation and partial autocorrelation considerations, possible ARIMA (p,d,q) models were formulated and compared to each other as depicted in Table 4.

Based on selected diagnostic criterion i.e., RMSE (322.04) and MAPE (34.14) values and significance of all estimated parameters at 5% LOS, the model ARIMA (1,1,1) was identified as one of the appropriate models to forecast the Indian coffee export (value). Further, residual analysis was also carried out to check the adequacy of the selected ARIMA (1,1,1) model and it was revealed that none of the lags of residual ACF and PACF charts were found to be significant as per Figure 2 and The Ljung-box test was also confirmed the same as its p-value was 0.802 (> 0.05), hence null hypothesis was accepted as residuals were independently distributed, which also indicated good fit of the selected model i.e., ARIMA (1,1,1). Similar model was reported by Amry (2021) as to forecast the Coffee Exports in Indonesian, and then the ARIMA model was

Coffee export (Y_t) = 104.89-0.24Y_{t-1}-0.46 e_{t-1}+e_t

Table 3. Augmented Dickey-Fuller (ADF) Test for Indian Coffee export (value)

	Data type	ADF statistic	Critical value (P value)	Decision
Coffee export (value)	ADF at level	-1.852	0.63	Data Non-Stationary
	ADF at 1st difference	-3.66	0.03	Data became Stationary

Table 4. ARIMA Model fit statistics of Indian Coffee export (value)

Model		Parameter Estimates					Goodness of Fit	
	a	Autoregressive Coefficient			Moving Average Coefficient		RMSE	MAPE
		AR1	AR2	AR3	MA1	MA2		
(1,1,0)	104.23	0.20*					325.34	35.33
(1,1,1)	104.89	-0.24*			0.46*		322.04	34.14
(1,1,2)	104.72	-0.35*			0.57*	0.03*	324.06	36.35
(2,1,2)	102.93	0.27*	0.22*		-0.04*	-0.30	323.74	34.71
(2,1,1)	104.60	-0.30*	0.03*		0.52		324.11	36.61
(3,1,1)	102.87	-0.18*	0.01	0.05	0.40		324.58	34.8
(3,1,2)	102.07	0.26*	0.14	0.03	-0.04	-0.22	324.58	36.28
** Significant at 1%, * Significant at 5%								

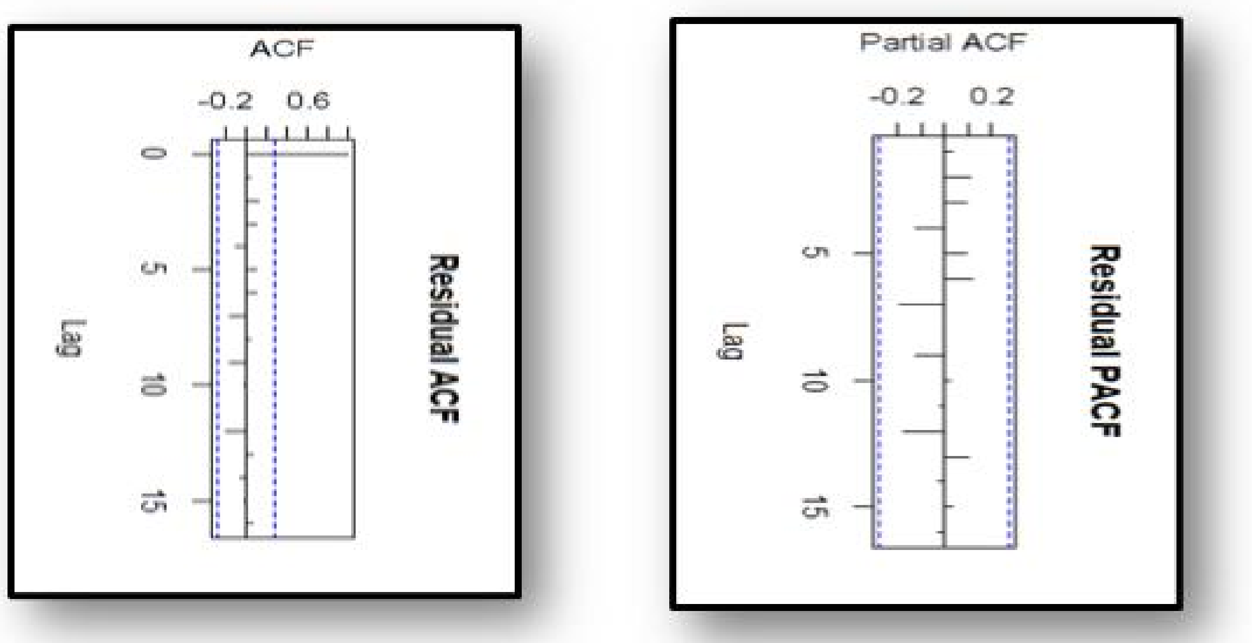

Figure 2. Residual ACF and PACF for Indian Coffee export (value)

Artificial Neural Network (ANN) Model

A multilayer feed forward neural network architect with back propagation was considered for the Coffee export (value) from India. As a result, 5 lags were identified as optimal for network as input nodes and there by various network topologies were trained by increasing the number of hidden nodes from 1 to 15, with using the sigmoid function as an activation function in the hidden layer. Among several tentative models, the top performing models were listed in Table 5, based on RMSE, MAE and MAPE. A neural network model with 5-10-1 was outperformed among all other neural networks with lower RMSE (5.37),) and MAPE (10.51) values.

Table 5. Performance of different numbers of neural network models

Network structure Coffee	RMSE	MAPE
05-01-2001	25.41	21.55
05-02-2001	14.15	16.11
05-03-2001	9.62	14.13
05-04-2001	7.56	12.78
05-05-2001	7.18	12.51
05-06-2001	6.69	12.42
05-07-2001	5.43	10.82
05-08-2001	5.55	11.23
05-09-2001	5.41	10.88
05-10-2001	5.37	10.51
05-11-2001	5.49	10.69
05-12-2001	5.53	11.06

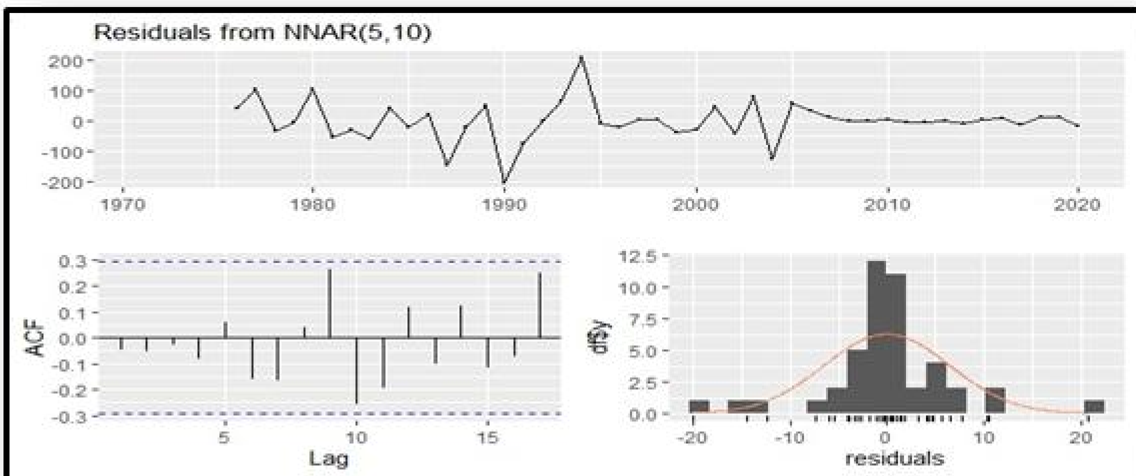

Figure 3. Residual plots for Coffee export (value) from India

Further, residual analysis was also carried out to check the adequacy of the selected Neural network model and it was discovered that none of the lags of residual ACF chart were found to be significant as per Figure 3 and The Ljung-box test was also confirmed the same as its p-value was 0.288, hence null hypothesis was accepted at 5% LOS as residuals were distributed independently, which also indicated good fit of the selected model i.e., NNAR (5-10-1).

Forecasting the exports (Value) of Coffee from India through the best fitted model

Among the selected models from nonlinear regression (Exponential model), linear time series (ARIMA (1,1,1)) and non-linear time series (NNAR (5-10-1)) models, the most plausible model was recognized as NNAR (5-10-1), due to the better diagnostic criterion after comparing from the Table 2., Table 4 and Table 5. Similar kind of report was obtained by Singh and Mishra (2015), forecasting the prices of Groundnut oil in Mumbai. Now the best fitted model was further compared with its actual and predictions for last three years, as illustrated in the following Table 6.

Table 6. Sample of predicted values using ANN and ARIMA for comparison

Year	Actual	Forecasted Values	Forecast Error (%)
		ANN (5-10-1)	ANN
2018	5661.37	5684.52	0.40
2019	5549.27	5648.20	1.78
2020	5255.35	5420.54	3.14
2021		5387.77	
2022		5400.66	
2023		5464.50	
2024		5502.49	

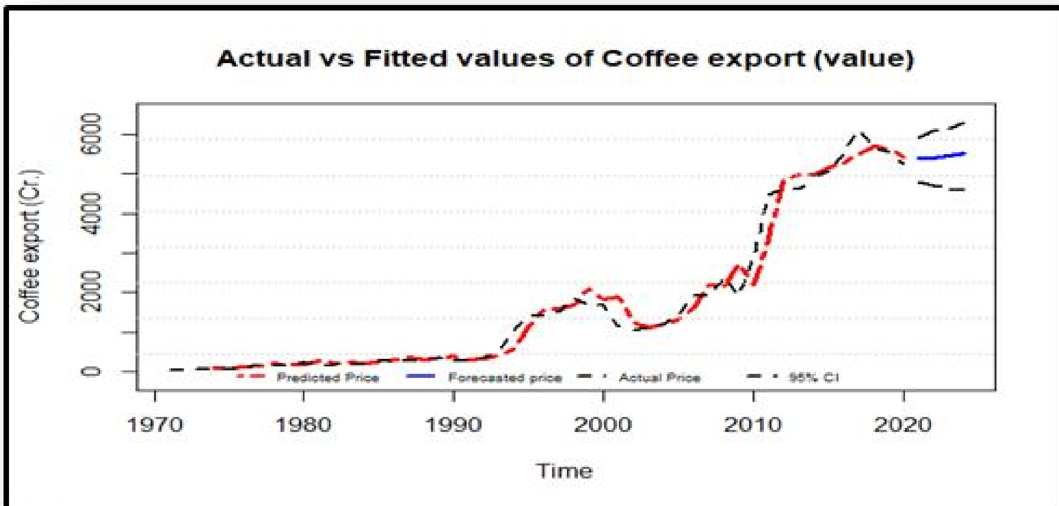

Figure 4. Actual vs fitted graph of ANN for exports of Coffee (value) from India

In this study, the actual and fitted graph of Coffee exports (value) was made by NNAR (5-10-1) as depicted in Figure 4. It was revealed that there were narrow variations in between the actual and predicted values. It was found that the exports of coffee were forecasted as 5387.77 (Cr.) and 5502.49 (Cr.) for the year 2021 and 2024 respectively.

CONCLUSION

In this study, an attempt was made to capture the trend of coffee exports different linear, nonlinear time series models, and Neural Network models were used. Among all, the best fitted model was recognized as NNAR (5-10-1), due to the better model selection criterion. By using this model, Coffee exports (value) were forecasted as 5502.49 (in Cr.) for the year 2024. It was concluded as there would be steady increasing trend of exports in future. This study on forecasting the coffee exports from India would help the coffee exporters and policymakers to make appropriate decisions.

LITERATURE CITED

Amry Z 2021 Time Series Modeling and Forecasting for Indonesian Coffee Export. *International Journal of Engineering Research & Technology*. 10(3): 211-214.

Fao.org/faostat *Food and Agricultural Organization of the United Nations*, 2022. https://www.fao.org/faostat/en/#data.

Govindaraju Ramesh D Srinivasa Rao V and Sitarambabu V 2021 Price Forecasting of

Ragi in Hassan Market of Karnataka – A Timeseries Approach. *The Andhra Agricultural Journal*. 68 (1): 126-132.

Gujarati D Porter D and Gunasekar S 2012 Basic Econometrics. *Mc Graw Hill*, New Delhi.

MOA & FM 2019 *Ministry of Agriculture and Formers Welfare*, 2019. https://www.indiastat.com/.

Ramesh D Pal S Rao VS and Bhattacharyya B 2015 Time series modeling for trend analysis and forecasting wheat production of India. *International Journal of Agriculture, Environment and Biotechnology*. 8(2): 303-308.

Rathod S Singh K N Paul R K Meher S K Mishra G C Gurung B Ray M and Sinha K 2017 An improved ARFIMA Model using Maximum Overlap Discrete Wavelet Transform (MODWT) and ANN for forecasting agricultural commodity price. *Journal of the Indian Society of Agricultural Statistics*. 71(2): 103-111.

Sankar J 2011 Forecasting fish product export in Tamilnadu - A stochastic model approach. Recent Research in Science and Technology 3:104-108.

Sahu P K 2010 Forecasting production of major food crops in four major SAARC countries. *International Journal of Agricultural and Statistics Science* 10:71-92.

Seber G A F and Wild C 1989 Nonlinear regression. John Willy, New York.

Singh A and Mishra G C 2015 Application of Box-Jenkins method and Artificial Neural Network procedure for Time Series Forecasting of Prices. *Spring*. 16 (1): 83–96.

Sirisha K J V K Subba Rao D V Srinivasa Rao V and P Rambabu 2020 A Time Series Analysis on Shrimp Production from India and Andhra Pradesh Using ARIMA Model. *The Andhra Agricultural Journal*. 67 (2): 98-104

Received on 17.10.2022 and Accepted on 19.03.2022