FORECASTING OF AREA AND PRODUCTION OF CASHEW NUT IN DAKSHINA KANNADA USING ARIMA AND EXPONENTIAL SMOOTHING MODELS

[1]Chaithra M.,[1*]Pramit Pandit and [1]Bishvajit Bakshi
[1]Department of Agricultural Statistics, Applied Mathematics and Computer
Science, University of Agricultural Sciences, Bengaluru, India
E Mail:chstatuasb@gmail.com,*pramitpandit@gmail.com and
bbishvajit@gmail.com

*Corresponding Author

Abstract
        The cultivation and marketing of cashew nut involve a considerable amount of work force. Hence, it plays a vital role in the Indian economic scenario. In this context, an attempt has been made to forecast the area and production of cashew nut with a view to help the planners in recommending policies regarding cashew nut. Due to autocorrelation in the data, time series forecasting models such as ARIMA and exponential smoothing models were adopted. Detection and removal of 3 significant outliers, i.e. 1 for area under cashew nut and 2 in case of cashew nut production, were done before fitting the models. Holt's model was found to have better forecasting ability with lowest RMSE value (1386.13) among the different models fitted for forecasting the area under cashew nut. From this model, area (ha) under cashew nut was forecasted to be 34492.10, 34974.81 and 35474.87 for the year 2018, 2019 and 2020, respectively. In case of cashew nut production, Brown's linear trend model, with RMSE value (10020.19), was observed to have better forecasting ability among the tried models. Production of cashew nut (in tonnes) was forecasted to be 10230.20, 10996.81 and 11833.00 for the year 2018, 2019 and 2020, respectively.

Key Words: ARIMA, Cashew Nut, Exponential Smoothing, Forecasting, Time Series Analysis.

1. Introduction
        Cashew nut had been introduced to India in 16th century with a view to prevent soil erosion (Elakkiya *et al.*, 2017). As it can be cultivated in diverse agro-climatic conditions, it has become a crop of high commercial value. A substantial amount of work force is required in production and marketing of cashew nut. Hence, it plays a vital role in the Indian economic scenario. However, in recent years, the cashew nut growers are not being able to obtain optimum yield and return (Senthil and Mahesh, 2013). In this context, forecasting of area and production can aid the planners in recommending policies regarding cashew nut (Chand *et al.*, 2007).

Pal *et al.* (2007) forecasted the milk production in India using ARIMA and double exponential smoothing models for the period from 1980-81 to 2004-05. ARIMA (1, 1, 1) model performed better in their study with lowest AIC, MAPE, MAE and MSE value. Debnath *et al.* (2013) forecasted the area, production and yield of cotton in India using ARIMA model for the period from 1950-51 to 2010-11. Krishnarani (2013) performed outlier analysis of tea price data. Four outliers were detected among which one was additive outlier and rest three were innovational outliers. Masuda and Goldsmith (2008) conducted a study on world Soybean production, yield and harvested area. Using damped trend exponential model, they finalized that world production of soybeans was predicted to be increased by 2.1 per cent annually to 359.7 million tonnes by 2030.

Hence, in this paper, an attempt has been made to forecast area and production of cashew nut of Dakshina Kannada in order to aid the planners substantially in formulating policies regarding cashew nut.

2. Materials and Methods

2.1 Description of the Study Area

Dakshina Kannada district of Karnataka has a total geographical area of 4,559 square km with shelter of the Western Ghats on the east. The district has 2 agro-climatic divisions namely, Coastal region and Malnad region.

2.2 Source of the Data

The secondary data on area and production of cashew nut for Dakshina Kannada district of Karnataka were collected for period of 1987 to 2017 from Directorate of Economics and Statistics (DES), Government of Karnataka, Bengaluru. Data from 1987 to 2013 and from 2014 to 2017 were used for the model building and model validation purpose, respectively.

2.3 Analytical Tools Used

For the purpose of forecasting, both Auto Regressive Integrated Moving Average (ARIMA) models and different exponential smoothing models were fitted and compared on the basis of RMSE (root mean square error), MAPE (mean absolute percentage error) value and significance of the parameter estimates. Durbin-Watson test was also employed to check autocorrelation in the time series data. R and SPSS 20.0 have been used for the purpose of statistical computations.

2.3.1 Auto Regressive Integrated Moving Average (ARIMA) Model

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

model is usually represented as ARIMA (p, d, q), where p, d, q denotes the order of AR process, differencing and MA process, respectively.

$$\left(1 - \sum_{s=1}^{p} \phi_s L^s\right) \Delta y_t = \mu + \left(1 + \sum_{s=1}^{q} \theta_s L^s\right) \varepsilon_t$$

where, y_t and ε_t are the actual observation and random error at time period t, respectively; ϕ_s (s = 1, 2, …, p) and θ_s (s = 1, 2, …, q) are model parameters. L is the usual lag operator, i.e. $L^s y_t = y_{t-s}$ and $\Delta y_t = y_t - y_{t-1}$.

2.3.2 Exponential Smoothing Model (ES)
One of the most popular parametric forecasting technique for producing a smoothed time series is exponential smoothing. Exponential smoothing process allots exponentially decreasing weights to the older observations, implying allocation of more weights to the recent observations.

2.3.2.1 Single Exponential Smoothing (SES) Model
This method is also known as method of estimation of future value by single weight. Let S_{t+1} denote the estimator of the level at time t+1. Given S_t, if X_t, which is the observation at time t and also the previous observed value, becomes available, then SES due to Brown (1963) updates the level estimator via the recurrence equation.

$$S_{t+1} = \alpha X_t + (1 - \alpha)S_t$$

Where, α is a smoothing parameter taking values in the interval (0, 1) and S_{t+1} is the estimator at time t.

2.3.2.2 Holt's Model
In presence of local linear trend, SES does not perform satisfactorily (Brown, 1963). In order to overcome this lacuna, Holt (1957) proposed an extension of SES by adding one more updating equation for the slope (trend).

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + b_{t-1})$$
$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

Where,

S_t = smoothened value at time period t
S_{t-1} = smoothened value at time period t-1
   α   = level smoothing constant
X_t = actual value at time period t
b_t = trend estimate of the time period t
b_{t-1} = trend estimate of the time period t-1
   γ   = trend smoothing constant

2.3.2.3 Brown's linear trend model
        A local linear trend forecasting procedure, called double exponential smoothing, was proposed by Brown (1983), which utilizes a single smoothing parameter to smooth both the level and the trend.

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$$
$$b_t = \alpha (S_t - S_{t-1}) + (1 - \alpha)S_{t-1}$$

Where,

S_t = smoothened value at time period t

S_{t-1} = smoothened value at time period t-1

α = level smoothing constant

X_t = actual value at time period t
b_t = trend estimate of the time period t
b_{t-1} = trend estimate of the time period t-1

2.3.2.4 Damped trend model
        In conjunction with the smoothing parameters α (level) and γ (trend), this method includes a damping parameter ϕ; $0 < \phi < 1$.

$$\widehat{y_{t+h \mid t}} = l_t + (\phi + \phi^2 + \cdots + \phi^h)b_t$$
$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + \phi b_{t-1})$$
$$b_t = \gamma (S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

Where,

S_t = smoothened value at time period t

S_{t-1} = smoothened value at time period t-1

α = level smoothing constant

X_t = actual value at time period t
b_t = trend estimate of the time period t
b_{t-1} = trend estimate of the time period t-1
Φ =   damping parameter

2.3.3 Durbin–Watson statistic
        This test statistic is employed to test for autocorrelation.

$$d = \frac{\sum_{t=2}^{n}(\widehat{e_t} - \widehat{e_{t-1}})^2}{\sum_{t=1}^{n} \widehat{e_t}^2}$$

Where,
n = total number of observation

2.3.4 R^2

It is known as co-efficient of determination. It indicates the proportion of variation present in response explained by the model.

$$R^2 = 1 - \frac{SS_{res}}{SS_T}$$

Where, SS_{res} and SS_T are the residual and total sum of squares, respectively.

2.3.5 RMSE

Root Mean Square Error is nothing but the standard deviation of the residuals. It measures the spread of these residuals.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

Where, y_i and $\hat{y_i}$ are the actual and predicted values of the response variable, respectively.

2.3.6 MAPE

It is another measure of prediction accuracy of forecasting methods in statistics.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \times 100\%$$

Where, y_i and $\hat{y_i}$ are the actual and predicted values of the response variable, respectively.

3. Results and Discussions

3.1 Forecasting of Area of Cashew Nut in Dakshina Kannada

The data of area under cashew nut were tested for autocorrelation using Durbin Watson test since it was time series data. The Durbin Watson statistic value was found to be 0.41. As the value did not lie between1.5 to 2.5, ARIMA and Exponential smoothing models were used to forecast the area.

One significant outlier was detected and removed, which was shown in Table 1. An overview of the data of area under cashew nut revealed a positive trend over time, indicating non-stationary nature of the series. It was confirmed through the Auto Correlation Function (ACF) and Partial Autocorrelation Function (PACF). ACF of the time series in figure 1 showed a slow linear decay of the autocorrelation coefficients with a few significant spikes. Figure 2 represents the PACF plot, which showed one significant spike at lag 1. These indicated non-stationarity of the time series. To make

the series stationary, it was first differenced. After differencing, the data attained stationarity as showed in figure 3.

Year	Type of outlier	Estimate	SE	t	p-value
1983	Additive	0.04**	0.009	4.158	<0.001

**: Significant at 1 % level
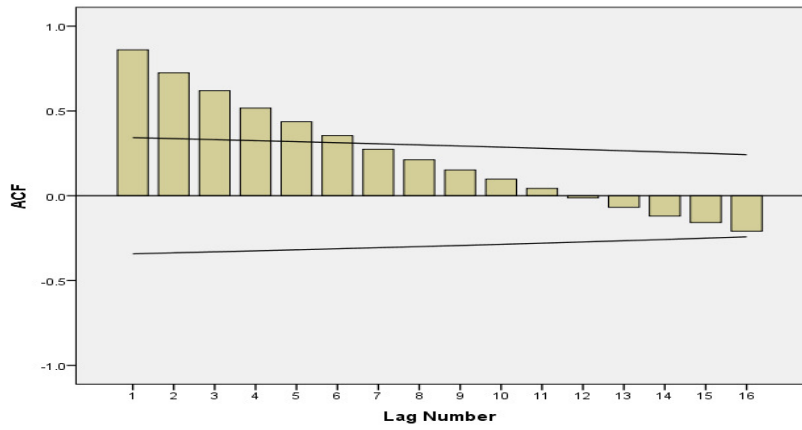
Table 1: Detected outlier of the area under cashew nut

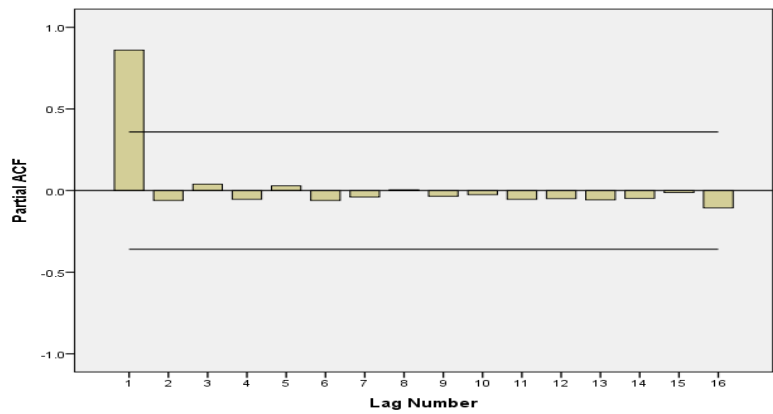**Figure 1: Autocorrelations at different lags of area under cashew nut**

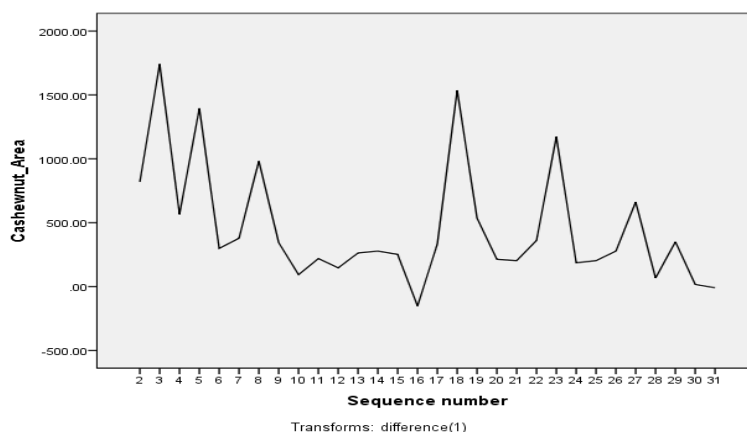**Figure 2: Partial autocorrelations at different lags of area under cashew nut**

Figure 3: Time plot of the differenced series for area under cashew nut

Among the ARIMA models, ARIMA (0, 1, 1) models was found to be the best fit and the estimates of the parameters were given in the Table 2.

Transformation	Parameters		Estimate	SE	T	p-value
	Constant		2.532*	1	2.533	0.01
Natural Log	Difference		1			
	MA	Lag 1	-0.44*	0.179	-2.486	0.02

*: Significant at 5 % level

Table 2: Estimates of the ARIMA (0, 1, 1) model parameters for area under cashew nut

Models were fitted based on the ACF and PACF plot of the differenced series as shown in figure 4 and 5, respectively. One spike was outside the limit in ACF plot and none of the spikes were outside the limit in PACF plot, indicating ARIMA (0, 1, 1) model as the most appropriate. Log transformation was done as there was skewness (-0.29) in the data. The adequacy of the model was also appraised based on the value of Ljung-Box Q statistic, which was found to be non-significant. Value of model fit statistics namely R^2, RMSE and MAPE value for ARIMA (0,1,1) model were also given in Table 3.
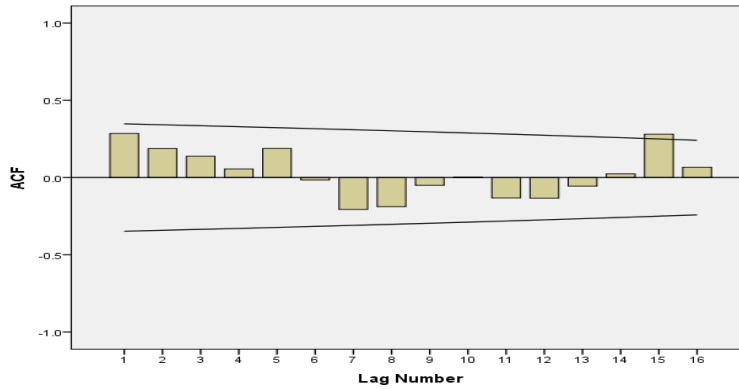
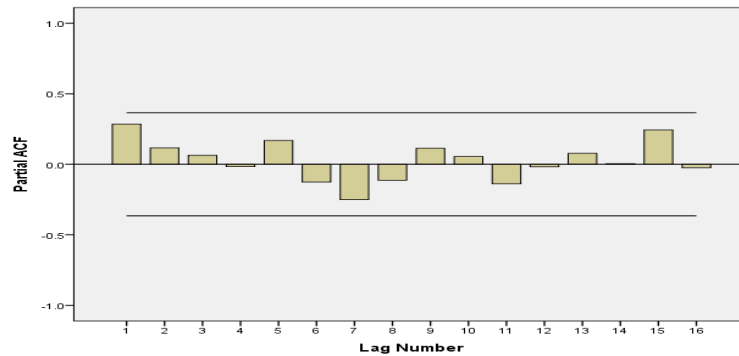**Figure 4: Autocorrelations at different lags of 1st differenced time series for area under cashew nut**

Figure 5: Partial Autocorrelations at different lags of 1st differenced time series for area under cashew nut

Model Fit statistics			Ljung-Box Q			Number of Outliers
R^2%	RMSE	MAPE	Statistic	DF	p-value	
98.88	428.606	1.283	14.06^{NS}	17	0.66	1

NS: Non significant

Table 3: Model fit statistics and Ljung-Box Q statistic of ARIMA for area under cashew nut

In order to check the adequacy of ARIMA (0, 1, 1) model, residual analysis was carried out. ACF and PACF plot of the residuals were obtained, in which all the spikes were found within limits as shown in figure 6, indicating that the model was adequate.
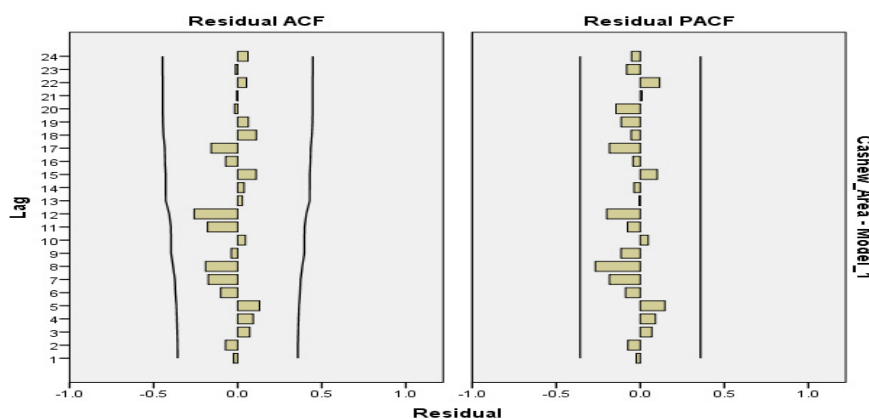
Figure 6: Residual autocorrelation and partial autocorrelations for area under cashew nut

Different exponential smoothing models were also fitted. Based on the Ljung-Box Q statistic given in Table 4, it could be said that the Holt's model fitted well with the data as the statistic was non-significant. Model fit statistics like R^2 (98.43%), RMSE (498.86) and MAPE (1.47) value for Holt's model were also more satisfactory than the other exponential smoothing models, as given in the same Table. The estimates of the parameters were given in Table 5.

Model	Model Fit statistics			Ljung-Box Q		
	R (%)	RMSE	MAPE	Statistic	DF	p-value
SES	97.30	642.65	1.86	17.13[NS]	17	0.45
Holt	98.43	498.86	1.47	20.89[NS]	16	0.18
Brown	98.26	529.37	1.67	27.70[*]	17	0.05
Damped Trend	98.64	477.60	1.33	28.56[*]	15	0.02

*: Significance at 5 % level, NS: Non significant

Table 4: Model fit statistics and Ljung-Box Q statistic of exponential smoothing models for area under cashew nut

Transformation	Parameters	Estimate	SE	T	p-value
Natural Log	Alpha (Level)	0.78[**]	0.16	4.65	<0.001
	Gamma (Trend)	0.52[*]	0.23	2.30	0.03

**: Significance at 1% level, *: Significance at 5 % level

Table 5: Estimates of the Holt's model parameters for area under cashew nut

Values predicted by ARIMA (0, 1, 1) and exponential smoothing model (Holt's model) were given in Table 6 and compared for forecasting ability. Among these models, Holt's model was observed to have better forecasting accuracy with lower RMSE (1386.13) and MAPE (3.75) value. Forecasting was done for the next three years using the Holt's model, as presented in Table 7. Outcomes emanated from the investigation were in line with the findings of Suresh and Priya (2011) and Prabakaran *et al.* (2013).

Year	Observed values (ha)	Predicted values using ARIMA (0, 1, 1) (ha)	Predicted values using Holt's model (ha)
2014	30967	30978.08	31097.81
2015	32756	30951.66	31187.71
2016	32863	30886.35	31284.84
2017	33040	30782.41	31391.32
	RMSE	1750.69	1386.13
	MAPE	4.60	3.75

Table 6: Predicted values and model fit statistics for area under cashew nut crop

Year	Forecasted values (ha)
2018	34492.10
2019	34974.81
2020	35474.87

Table 7: Forecasted values for area under cashew nut in Dakshina Kannada

3.2 Forecasting of Cashew Nut Production in Dakshina Kannada

The data of cashew nut production were tested for autocorrelation using Durbin Watson test since it was time series data and the test statistic value was found to be 0.52. As the value did not lie between1.5 to 2.5, ARIMA and exponential smoothing models were used to forecast the production.

Two significant outliers were detected and removed, which were shown in Table 8. An overview of the data revealed a positive trend over time, indicating non-stationary nature of the series. It was also confirmed through the ACF and PACF. ACF of the time series in figure 7 showed a slow linear decay of the autocorrelation coefficients with a few significant spikes. PACF plot, represented in figure 8, showed a significant spike at lag 1. These indicated non-stationarity of the time series. To make the series stationary, it was first differenced. After differencing, the data attained stationarity, as shown in figure 9.

Year	Type of outlier		Estimate	SE	t	p-value
1983	Transient	Magnitude	0.07**	0.021	3.547	<0.001
		Decay factor	0.93*	0.407	2.294	0.03
2010	Innovational		0.19**	0.021	9.084	<0.001

**: Significant at 1 % level, *: Significance at 5 % level

Table 8: Detected outliers of cashew nut production
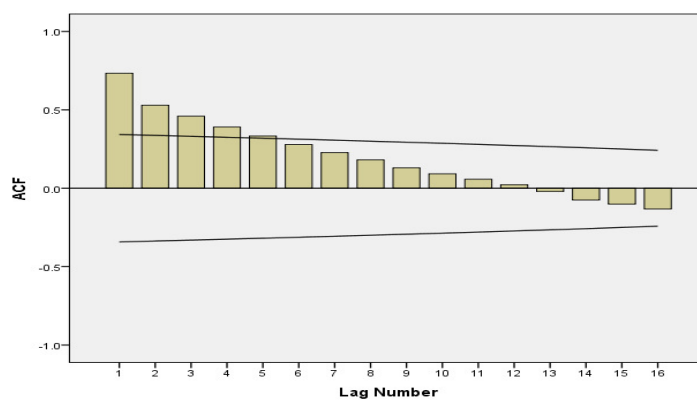

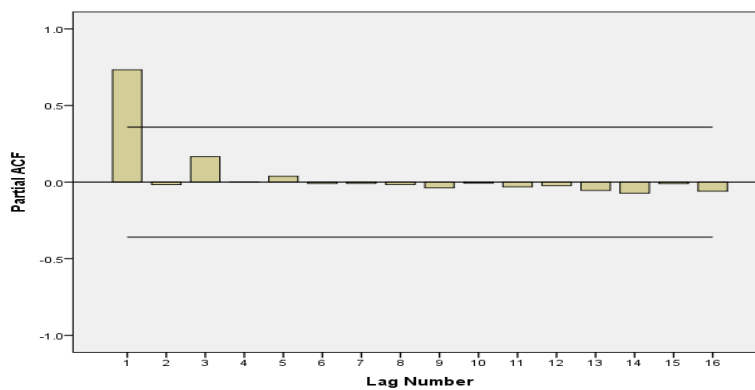**Figure 7: Autocorrelations at different lags of cashew nut production**

Figure 8: Partial Autocorrelations at different lags of cashew nut production
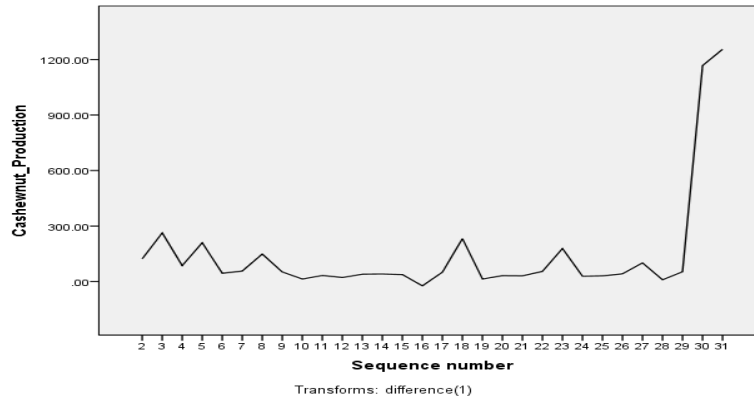
Transforms: difference(1)

Figure 9: Time plot of the differenced series for cashew nut production

Among the ARIMA models, ARIMA (1, 1, 1) model was found to be the best fit and the estimates of the parameters were given in the Table 9.

Transformation	Parameters		Estimate	SE	t	p-value
Natural Log	Constant		1.03^{NS}	2.365	0.436	0.67
	AR	Lag 1	-0.155^{NS}	0.206	-0.75	0.46
	Difference		1			
	MA	Lag 1	-1^{**}	0.233	-4.294	<0.001

**: Significant at 1 % level, NS: Non significant

Table 9: Estimates of the ARIMA (1, 1, 1)model parameters for cashew nut production

Models were fitted on the basis of the ACF and PACF plot of the differenced series as shown in figure 10 and 11, respectively. One spike had crossed the limit in both ACF and PACF plot at lag 1, indicating ARIMA (1, 1, 1) model as the most appropriate. Log transformation was done as there was skewness in the data (1.99). The adequacy of the model was also appraised in Table 10 on the basis of Ljung-Box Q statistic value, which was found to be non-significant. Value of model fit statistics namely R^2, RMSE and MAPE were also presented.
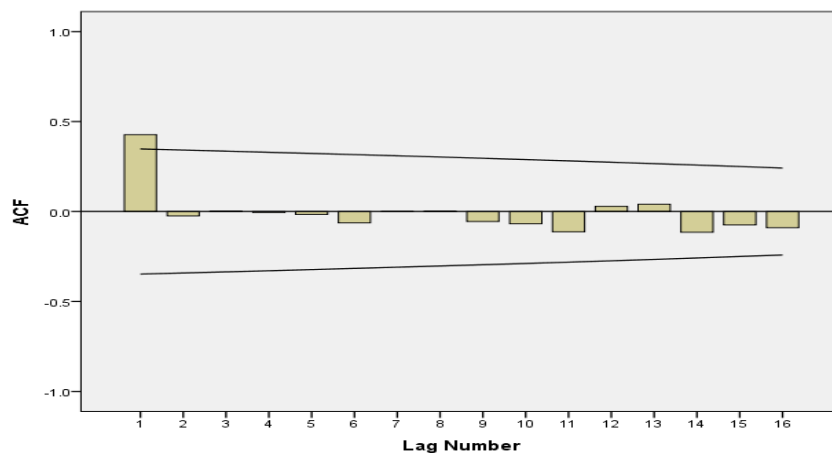
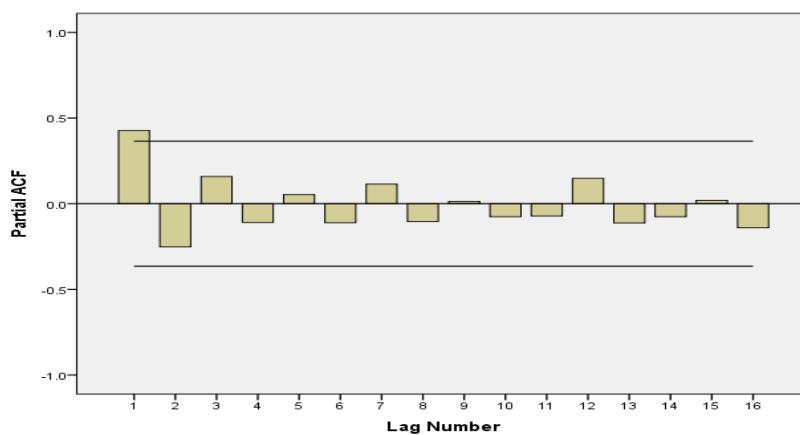**Figure 10: Autocorrelations at different lags of 1st differenced time series for cashew nut production**

Figure 11: Partial Autocorrelations at different lags of 1st differenced time series cashew nut production

Model Fit statistics			Ljung-Box Q			Number of Outliers
R^2 (%)	RMSE	MAPE	Statistic	DF	p-value	
99.24	81.547	1.448	25.58[NS]	16	0.06	2

NS: Non significant

Table 10: Model fit statistics and Ljung-Box Q statistic of ARIMA for cashew nut production

To check model adequacy, residual analysis was carried out. ACF and PACF plot of the residuals were obtained, in which most of the spikes were found within limits except at lag 1 in both ACF and PACF plot as seen in figure 12. So, because of the presence of significant spikes in residual ACF and PACF plot, it could be inferred that even ARIMA (1, 1, 1) was not found appropriate.
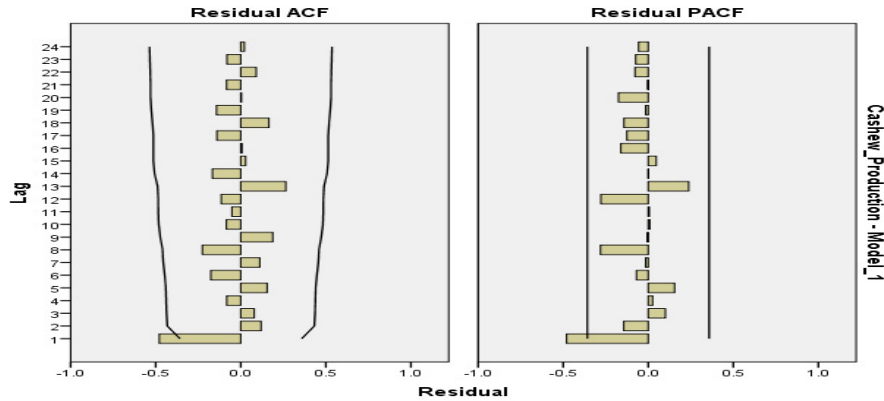

Figure 12: Residual autocorrelation and partial autocorrelations for cashew nut production

Different exponential smoothing models were also fitted. Brown's linear trend model was said to be the best fit based on high R^2 (93.23%) and minimum RMSE (222.86), MAPE (2.26) value, which were given in Table 11. The estimated values of the parameters were given in Table 12.

	Model Fit statistics			Ljung-Box Q		
Model	R^2 (%)	RMSE	MAPE	Statistic	DF	p-value
SES	85.64	323.94	2.85	9.88[NS]	17	0.91
Holt	93.12	228.17	2.27	5.18[NS]	16	0.99
Brown	93.23	222.86	2.26	5.98[NS]	17	0.99
Damped Trend	93.10	232.09	2.27	5.20[NS]	15	0.99

NS: Non significant

Table 11: Model fit statistics and Ljung-Box Q statistic of exponential smoothing models for cashew nut production

Transformation		Estimate	SE	t	p-value
Natural Log	Alpha (Level and Trend)	0.88[**]	0.09	9.93	<0.001

**: Significant at 1 % level

Table 12: Estimate of the Brown's linear trend model parameter for cashew nut production

As none of the ARIMA models was found appropriate, based on the predicted values of only Brown's linear trend model, value of model fit statistics like RMSE and MAPE were given in Table 13. Forecasting was done for the next three years using Brown's linear trend model, as presented in Table 14.

Year	Observed values (tonnes)	Predicted values using Brown's linear trend model (tonnes)
2014	7879	8650.44
2015	5266	10507.55
2016	6885	12806.64
2017	9576	15680.86
	RMSE	10020.19
	MAPE	64.77

Table 13: Predicted values and model fit statistics for cashew nut production

Year	Forecasted values (tonnes)
2018	10230.20
2019	10996.81
2020	11833.00

Table 14: Forecasted values for cashew nut production in Dakshina Kannada

4. Conclusion

Due to presence of autocorrelation in the data, time series forecasting models such as ARIMA and exponential smoothing models were adopted. Detection and removal of several significant outliers were done before fitting the models. Holt's model was found to have better forecasting ability with the lowest RMSE value (1386.13) among the different models fitted for forecasting the area under cashew nut. From this model, area (ha) under cashew nut was forecasted to be 34492.10, 34974.81 and 35474.87 for the year 2018, 2019 and 2020, respectively. In case of cashew nut production, Brown's linear trend model, with RMSE value (10020.19), was observed to have better forecasting ability among the tried models. Production of cashew nut (in tonnes) was forecasted to be 10230.20, 10996.81 and 11833.00 for the year 2018, 2019 and 2020, respectively. Outcomes emanated from this investigation are expected to help the planners in recommending policies regarding cashew nut with a view to strengthen the economic backbone of India.

References

1. Bera, K., Debnath, M.K. and Mishra, P. (2013). Forecasting area, production and yield of cotton in India using ARIMA model, Research & Reviews: Journal of Space Science & Technology, 2 (1), p. 16-20.

2. Box, G.E.P. and Jenkins, G. (1994). Time series analysis, Forecasting and control, San Francisco, CA, Holden-Day.

3. Brockwell, P.J. and Davis, R.A. (1996). Introduction to time series and forecasting, 2nd Edn., New York, Springer.

4. Brown, R.G. (1963). Smoothing forecasting and prediction of discrete time series, USA, Prentice- Hal.

5. Chand, R., Raju, S.S. and Pandey, L.M. (2007). Growth crisis in agriculture: severity and options at national and state levels, Economic and Political Weekly, 42(26), p. 2528-2533.

6. Elakkiya, E., Sivaraj, P. and Prabhakar, V.A. (2017). Growth and Performance of Cashew Nut in India – an analysis, International Journal of Current Microbiology and Applied Sciences, 6(6), p. 1817-1823.

7. Holt, C.C. (1957). Forecasting seasonal and trends by exponentially weighted moving averages, ONR Research Memorandum, Carnigie Institute.

8. Krishnarani, S.D. (2013). Time series outlier analysis of tea price data, American Journal of Theoretical and Applied Statistics, 2(1), p. 1-6.

9. Masuda, T. and Goldsmith, P.D. (2008). World soybean production: Area harvested yield, and long- term projections, International Food and Agribusiness Management Review, 12(4), p. 143-161.

10. Pal, S., Ramasubramanian, V. and Mehta, S.C. (2007). Statistical models for forecasting milk production in India, Journal of Indian Society of Agricultural Statistics, 61(2), p. 81-83.

11. Prabakaran, K., Sivapragasam, C., Jeevapriya, C. and Narmatha, A. (2013). Forecasting cultivated area and production of wheat in India using ARIMA model, Golden Research Thoughts, 3 (3), p. 1-7.

12. Senthil, A. and Mahesh, M.P. 2013. Analysis of cashew nut production in India, Asia Pacific Journal of Marketing and Management Review, 2(3), p. 106-110.

13. Suresh, K.K. and Priya, K.S.R. (2011). Forecasting sugarcane yields of Tamil Nadu using ARIMA models, Sugar Tech., 13 (1), p. 23-26.