Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.5409

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **5 (5):** 341-347 (2017)

International Journal of Pure & Applied Bioscience

Research Article

OPEN ACCESS

Forecasting of Area and Production of Cotton in India: An Application of ARIMA Model

Vijaya B. Wali[1*], Devendra Beeraladinni[2] and Lokesh H[3].

[1, 2]Assistant Professor, [3] Professor

Department of Agricultural Economics, UAS, Raichur-584104

*Corresponding Author E-mail: vbw06@rediffmail.com

Received: 8.08.2017 | Revised: 12.09.2017 | Accepted: 16.09.2017

ABSTRACT

The paper attempts forecasting of the area and production of cotton in India using the univariate autoregressive integrated moving average (ARIMA) model. The time series data on area and production of cotton in India for the period of 65 years from 1950-51 to 2015-16 was analyzed for the study. The best models were selected by comparing Akaike Information Criterion (AIC), Schwartz's Bayesian Criterion (SBC), Normalized BIC; Mean Absolute Percentage Error (MAPE) and maximum values of R^2*. The study revealed that ARIMA (0, 1, 0) and ARIMA (1, 1, 1) were the best fitted models for forecasting area and production of cotton in India respectively. Selected models were used to forecast area and production of cotton for four years from 2017-18 to 2020-21. The analysis showed an increasing trend in area and production of cotton. If the present trend continues, the cotton area in India in the year 2020-21 would be 123.83 lakh ha with upper and lower limits of 146.51 and 101.16 lakh ha, respectively and the production of cotton would be 316.16 lakh bales (170 kg each) with upper and lower limits of 412.70 and 219.63 lakh bales, respectively.*

***Key words:** Area, production, Autoregressive integrated moving average (ARIMA) model, Box and Jenkins, Forecasting.*

INTRODUCTION

Cotton is one of the major commercial crops grown in India. Since 2015-16 India stands first in both area and production of cotton in the world. The major cotton growing states in India are Punjab, Haryana, Rajasthan, Madhya Pradesh, Gujarat, Maharashtra, Tamil Nadu and Karnataka. India has constituted around 25 per cent of the total area in the world. During 2014-15 area under cotton was varying between 110-121 lakh ha with a production of 34.81 million bales. During 2015-16 the cotton production in India declined to 30.80 million bales, which was the lowest during the last five years. This drastic reduction in production of cotton was due to the sucking pest and white fly attack especially in the northern region.

Cite this article: Wali, V.B., Beeraladinni, D. and Lokesh, H., Forecasting of Area and Production of Cotton in India: An Application of ARIMA Model, *Int. J. Pure App. Biosci.* **5(5):** 341-347 (2017). doi: http://dx.doi.org/10.18782/2320-7051.5409

The country was expected to produce 34.15 million bales during 2016-17, on account of the better weather conditions across all cotton growing regions of the country (Cotton Advisory Board, 2016).

The Government of India has launched " Technology Mission on Cotton" in February 2000 with an objective of improving the production and productivity of cotton through development of high yielding varieties; enhance the income of the cotton growers by reducing cost of cultivation, appropriate transfer of technology and better farm management practices, cultivation of Bt-cotton hybrids etc. Area estimation and forecasting of production are essential procedures supporting in policy decisions with respect to production, price structures as well as consumption of cotton in the country. Increased global demand for cotton should induce higher production in the next decade. With these backgrounds, it is necessary to know the extent of cotton production in future with available resources. Various approaches have been used for forecasting such agricultural systems. Borkar Prema *et al*.[1] in their empirical study showed that ARIMA (2, 1, 1) is the appropriate model for forecasting the production of cotton in India. The study of Debnath *et al*.[2] revealed that area, production and yield of cotton in India would increase from 2016-17 to 2020-21. Similar studies have been conducted by Payyamozhil, S. *et al*.[4] and Sudar Rajan *et al*.[5] for forecasting cotton production in India, the analysis revealed that ARIMA (0, 1, 0) is the best model for forecasting cotton production. The present study has been undertaken with an objective to forecast the area and production of cotton in India in future using Box- Jenkins ARIMA model.

MATERIAL AND METHODS

The time series forecasting is an important area of forecasting in which past observations of the variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to interpolate the time series into the future values of the variable. In the present study, the data on cotton area and production in India was collected for 65 years from 1950-51 to 2015-16 from www.Indiastat.com (Statistical database). By using SAS 9.3 and SPSS 16 software, the data was analyzed to fit the best model using autoregressive integrated moving average (ARIMA) model. The selected best models were used to forecast the area and production of cotton in India up to 2020-21.

Description of ARIMA Model

One of the most important and widely used time series models is the Auto Regressive Integrated Moving Average (ARIMA) Model. The popularity of the model is due to its statistical properties as well as the well known Box- Jenkins methodology in the model building process. In an autoregressive integrated moving average (ARIMA) model, the future value of a variable is assumed to be a linear function of several past observations and random errors. That is, the underlying process that generates the time series has the form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (1)$$

where y_t and ε_t are the actual value and random error at time t. ϕi $(i=0,1,2,\ldots.p)$ and θj $(j=0,1,2,\ldots.q)$ are model parameters, p and q are integers, often referred to as order of the model. ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,...n. A first order auto-regressive process is denoted by ARIMA (1, 0, 0) or simply AR (1) and is given by, $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$ and a first order moving average process is denoted by ARIMA (0, 0, 1) or simply MA (1) and is given by $y_t = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t$. Equation (1) entails several important special cases of ARIMA family of models. One central task of the ARIMA model building is to determine the appropriate model order (p,q). The Box- Jenkins methodology includes three iterative steps model identification, parameter estimation and diagnostic checking.The basic idea of model identification is that if a time series is generated from ARIMA process, it should have some theoretical autocorrelation properties. By matching the empirical autocorrelation pattern with the theoretical one, it is often possible to identify the potential models for the given time series. Box- Jenkins proposed to use the autocorrelation function and partial autocorrelation function of the sample data as basic tools to identify the order of the ARIMA model.

In the identification steps, data transformation is often needed to make the time series stationary. Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting. A stationary time series has the property that its statistical characteristics such as the mean and the autocorrelation structure are constant over time. When the observed time series presents trend and heteroscedasticity, differencing and power transformation are often applied to the data to remove the trend and stabilize the variance before an ARIMA model can be fitted. Once the tentative model is specified, the parameters are estimated such that an overall measure of errors is minimized. This can be done with a non linear optimization procedure. The last step of model building is the diagnostic checking of model adequacy. This is basically to check if the model assumptions about the errors are satisfied. Several diagnostic statistics viz, , Akaike Information Criterion(AIC), Schwartz's Bayesian Criterion(SBC), Normalized BIC, Mean Absolute Percentage Error (MAPE) and plots of residuals can be used to examine the goodness of fit of tentative models. This three step model building process is repeated several times until a satisfactory model is finally selected. The finally selected model can then be used for prediction process.

RESULTS AND DISCUSSION

The ARIMA model was applied according to three steps namely model identification, model parameter estimation, diagnostic checking. Sixty-five years data on cotton area and production was used for modeling purpose.

Model Identification

The model specification involved the plots of the auto correlation function (ACF), partial auto correlation function (PACF) and the plot of the differenced series. The autocorrelation functions of 1st differenced time series shows stationarity for cotton area and production, as the autocorrelation declines faster than the auto correlation of un-differenced series. The auto correlation function and partial auto correlation function for cotton area and production were obtained and presented in Table 1 & 2 and the plots of the differenced series were depicted in Fig. 1& 2.

Table 1: ACF and PACF of cotton area in India

Lags	ACF	SE	Box-Ljung Statistic			PACF	SE
			Values	DF	Sig.		
1	-.046	.121	.143	1	.705	-.046	.124
2	-.159	.120	1.882	2	.390	-.161	.124
3	.280	.119	7.392	3	.060	.272	.124
4	-.122	.118	8.452	4	.076	-.144	.124
5	-.160	.117	10.309	5	.067	-.083	.124
6	.166	.116	12.336	6	.055	.060	.124
7	-.030	.115	12.403	7	.088	.001	.124
8	-.206	.114	15.630	8	.048	-.145	.124
9	-.082	.113	16.159	9	.064	-.196	.124
10	.193	.112	19.095	10	.039	.210	.124
11	-.146	.111	20.809	11	.035	-.108	.124
12	-.093	.110	21.516	12	.043	-.052	.124
13	.133	.109	22.999	13	.042	-.057	.124
14	-.088	.108	23.657	14	.050	.020	.124
15	.051	.107	23.888	15	.067	.158	.124
16	.279	.106	30.803	16	.034	.139	.124

ACF - Autocorrelation Function PACF- Partial Autocorrelation Function
DF- Degrees of Freedom SE- Std. Error

Table 2: ACF and PACF of cotton Production in India

Lags	ACF	SE	Box-Ljung Statistic			PACF	SE
			Values	DF	Sig.		
1	.068	.121	.318	1	.573	.068	.124
2	-.202	.120	3.133	2	.209	-.207	.124
3	.210	.119	6.243	3	.100	.253	.124
4	.097	.118	6.920	4	.140	.006	.124
5	-.245	.117	11.285	5	.046	-.179	.124
6	.005	.116	11.287	6	.080	.035	.124
7	.285	.115	17.402	7	.015	.200	.124
8	-.118	.114	18.470	8	.018	-.107	.124
9	-.187	.113	21.198	9	.012	-.079	.124
10	-.112	.112	22.188	10	.014	-.294	.124
11	.042	.111	22.328	11	.022	.101	.124
12	-.117	.110	23.460	12	.024	-.051	.124
13	-.160	.109	25.608	13	.019	-.118	.124
14	.077	.108	26.109	14	.025	-.029	.124
15	.196	.107	29.459	15	.014	.228	.124
16	-.055	.106	29.730	16	.019	.031	.124

ACF - Autocorrelation Function PACF- Partial Autocorrelation Function
DF- Degrees of Freedom SE- Std. Error

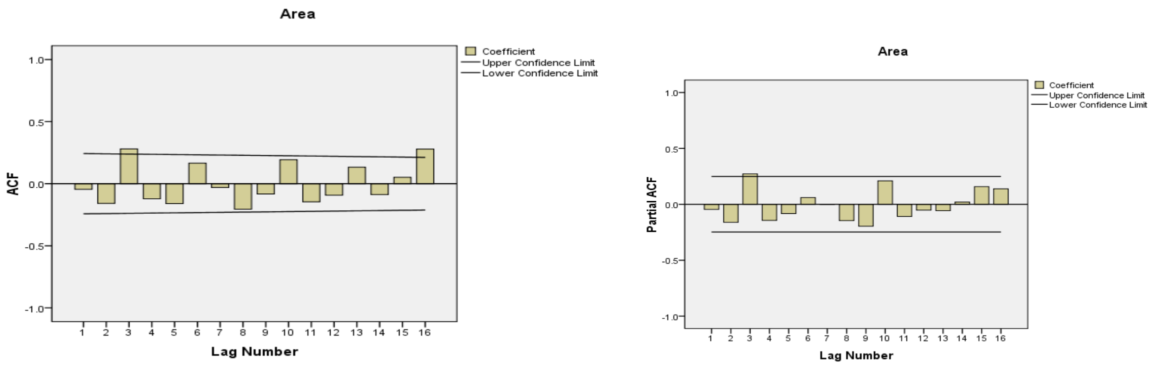

Fig. 1: ACF and PACF of differenced data of Cotton Area

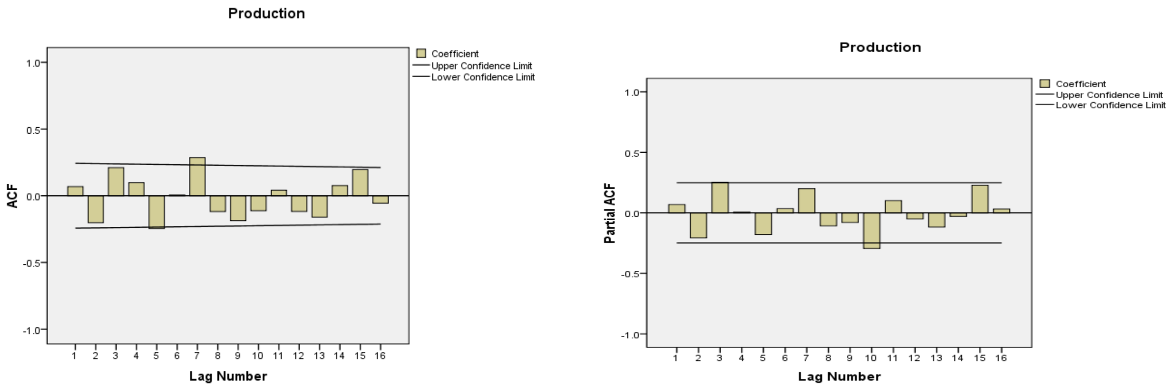

Fig. 2: ACF and PACF of differenced data of Cotton Production

The various models were fitted. The models which had minimum AIC, SBC value were selected as a best fit model for forecasting the future values of area and production of cotton. The various models and their AIC, SBC values are presented in Table 2. It was observed from the table that ARIMA (0, 1, 0) and ARIMA (1, 1, 1) had the lowest AIC and SBC values for area and production of cotton, respectively.

Table 2: AIC and SBC values of ARIMA Models

	ARIMA(p, d, q)	AIC	SBC
Area	ARIMA(1,1,1)	997.39	1003.91
	ARIMA(0,1,1)	997.09	1001.44
	ARIMA(1,1,0)	997.15	1001.50
	ARIMA(0,1,0)	**995.30**	**997.47**
	ARIMA(0,1,2)	996.38	1002.91
	ARIMA(2,1,0)	997.32	1003.84
	ARIMA(1,1,2)	996.94	1006.64
	ARIMA(2,1,1)	995.07	1003.76
Production	**ARIMA(1,1,1)**	**1171.48**	**1176.01**
	ARIMA(0,1,1)	1175.51	1179.86
	ARIMA(1,1,0)	1175.89	1180.23
	ARIMA(0,1,0)	1174.29	1176.39
	ARIMA(0,1,2)	1173.72	1180.24
	ARIMA(2,1,0)	1174.70	1181.22
	ARIMA(1,1,2)	1173.06	1181.76
	ARIMA(2,1,1)	1173.15	1181.85

AIC- Akaike information Criterion SBC- Schwartz's Bayesian Criterion

Model estimation

By using SPSS 16 package the model parameters were estimated and presented in Table 3 and 4. It was observed from Table 3 that for cotton area R^2 was 0.87. The value of AIC and SBC were lowest and worked out to be 995.30 and 997.47, respectively for ARIMA(0, 1 ,0) model and the MAPE value was recorded as 4.80, indicated that ARIMA (0, 1, 0) was the best model for forecasting cotton area in India. Similarly from Table 4, for production, R^2 value was 0.95. The value of AIC and SBC were lowest and worked out 1171.48 and 1176.01, respectively for ARIMA(1, 1, 1) model and the MAPE value was recorded as 14.10 indicated that ARIMA (1, 1, 1) was the best model for forecasting cotton production in India.

Table 3: Diagnostic tools and model selection criteria for area of cotton in India

Model Fit statistics								Ljung-Box Q(18)		
Stationary R^2	**R^2**	**RMSE**	**MAPE**	**MAE**	**AIC**	**SBC**	**Normalized BIC**	**Statistics**	**DF**	**Sig.**
-1.70E-16	0.87	507.54	4.80	397.56	995.30	997.47	12.52	30.97	18	0.03

R^2 – Coefficient of Determination **RMSE**- Root Mean Square Error **MAPE**- Minimum Absolute Percentage Error **MAE**- Minimum Absolute Error AIC- Akaike information Criterion **Normalized BIC, SBC**- Schwartz's Bayesian Criterion **DF**-Degrees of Freedom

Model parameters for Cotton Area

Type	Estimate	Std Error	t-value	Sig.
Constant	92.88	62.958	1.48	0.15

Table 4: Diagnostic tools and model selection criteria for production of cotton in India

Model Fit statistics								Ljung-Box Q(18)		
Stationary R^2	R^2	RMSE	MAPE	MAE	AIC	SBC	Normalized BIC	Statistics	DF	Sig.
0.12	0.95	1.91E3	14.10	1.39E3	1171.48	1176.01	15.31	26.70	16	0.05

R^2 – Coefficient of Determination **RMSE**- Root Mean Square Error **MAPE**- Minimum Absolute Percentage Error **MAE**- Minimum Absolute Error AIC- Akaike information Criterion **Normalized BIC, SBC**- Schwartz's Bayesian Criterion **DF**-Degrees of Freedom

Model parameters for Cotton Production

Type	Estimate	Std Error	t-value	Sig.
Constant	402.814	271.604	1.483	0.143
AR(1)	-0.728	0.114	-6.390	0.000
MA(1)	-0.999	1.261	-0.792	0.431

Diagnostic checking

The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen ARIMA model, which has been done through examining the autocorrelation and partial autocorrelation of the residuals of various order. The ACF and PACF of residuals (Fig 3) indicated the goodness of fit of the models ie, none of these autocorrelation was significantly different from zero. This proved was that the selected models were the appropriate models for forecasting area and production of cotton.

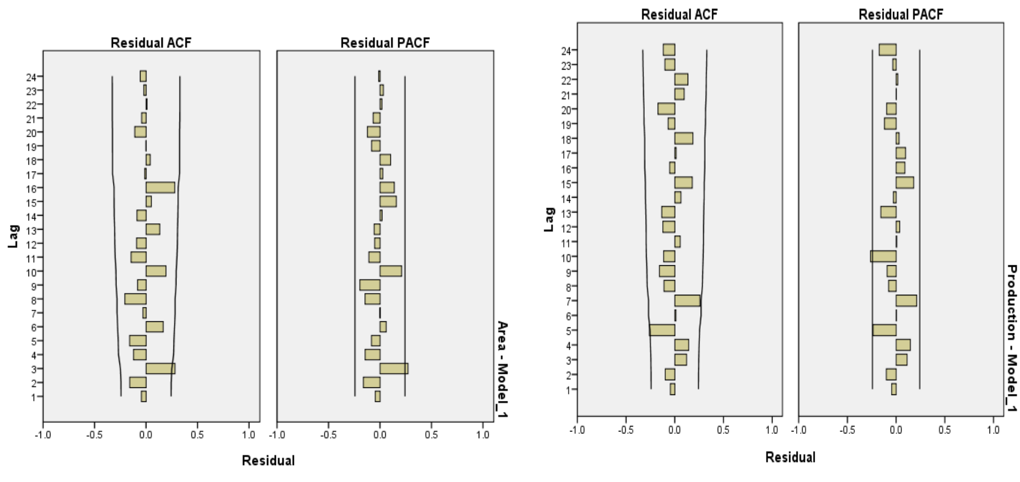

Fig. 3: ACF and PACF of residuals of fitted ARIMA Model for Area and Production

Forecasting of area and production of cotton by fitted ARIMA Models.

ARIMA models are basically developed to forecast the corresponding variables. To judge the forecasting ability of the ARIMA model, important measures of the sample period accuracies were calculated. The Mean Absolute Percentage Error (MAPE) for cotton area and production worked out to be 4.80 and 14.10 respectively. These measures indicate that the inaccuracy is low. The best fitted models were used to forecast the area and production of cotton in India for the period from 2017-18 to 2020-21 and presented in Table 5. The results (depicted in Fig 4) indicated an increasing trend in area and production of cotton.

Table 5: Forecast of cotton area and production in India For the period 2017-18 to 2020-21

Year	**Area** (lakh Ha)			**Production** (lakh Bales)		
	Forecast	**Lower Limit**	**Upper Limit**	**Forecast**	**Lower Limit**	**Upper Limit**
2017-18	121.05	106.98	135.12	306.54	246.30	366.76
2018-19	121.98	104.75	139.21	307.60	235.33	380.38
2019-20	122.91	103.01	142.80	316.65	228.30	398.99
2020-21	123.83	101.16	146.51	316.16	219.63	412.70

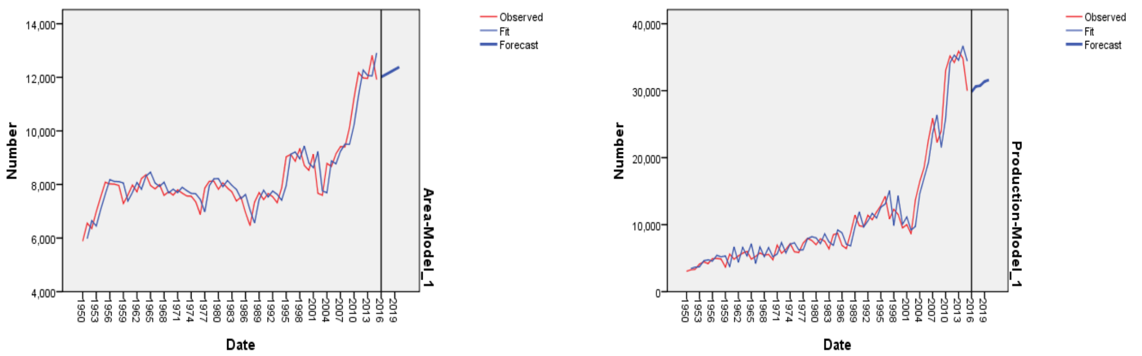

Fig. 4. Forecasting of area and production of cotton in India

CONCLUSION

In the study, ARIMA (0, 1, 0) and ARIMA (1, 1, 1) models were developed for forecasting area and production of cotton, respectively. From the forecasts available by using the best fitted models it can be seen that the cotton area and production will increase in the next four years. The area under cotton will increase from 121.05 to 123.83 lakh ha with the upper and lower limits of 135.12 - 106.98 lakh ha in 2017-18 and 146.51 - 101.16 lakh ha in 2020-21, respectively. Similarly, the production of cotton in India will increase from 306.54 to 316.16 lakh bales with the upper and lower limits of 366.76 - 246.30 lakh bales in 2017-18 and 412.70 - 219.63 lakh bales in 2020-21, respectively.

REFERENCES

1. Borkar, Prema and Tayade, P.M. Forecasting of Cotton Production in India Using ARIMA Model, *International Journal of Research in Economics and Social Sciences*, **6(5):** 1-7 (2016).
2. Debnath, M.K., Kartic Bera and Mishra, P. Forecasting Area, Production and Yield of Cotton in India Using ARIMA Model, *Research and Reviews: Journal of Space & Technology*, **2(1):** 16-20 (2013).
3. Prabakaran, K. and Sivapragasam, C. Forecasting Areas and Production of Rice in India Using ARIMA Model, *International Journal of Farm Sciences*, **4(1):** 99-106 (2014).
4. Payyamozhil, S., Kachi Mohideen,A. Forecasting of Cotton Production in India Using ARIMA Model. *Asia Pacific Journal of Research- A peer reviewed International Journal*, **XLVIII (I):** 70-74 (2017).
5. Sudar Rajan and Palanivel. Time Series Model to Forecast Production of Cotton in India: An Application of ARIMA Model. *AE International Journal of Multi Disciplinary Research*, **2(1):** 1-9 (2014).