THINK INDIA JOURNAL

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

Forecasting of Cotton Production In India- An ARIMA Modelling Approach

Banti Kumar[1] and Manish Kumar Sharma[2]

[1]Assistant Professor, Department of Mathematics, School of Chemical Engineering & Physical Sciences, Lovely Professional University, Phagwara-144411, Punjab

[2]Professor and Head, Division of Statistics and CS, SKUAST Jammu-180009, J&K

Corresponding Author's Email: - bantikumar1573@gmail.com

Abstract

An attempt has been made in this study to propose an autoregressive integrated moving average (ARIMA) model for forecasting the cotton production for next 10 years. The stationarity of the data have been testedthroughAugmented Dickey Fuller (ADF) test which showed the data is not stationary. For making it stationary, first order differencing has been done. Correlogram have been used to find the suitable order of ARIMA (p,d,q) values. ARIMA (1, 1,1) was found suitable for cotton production on the basis of minimum values of AIC, SBIC and Hannan Quinn values. Using data for 1950-51 to 2017-18, production of cotton in India were forecasted for 10 years starting from 2019 to 2028 and have shown a rising trend in cotton production in India.

Keywords: *Autoregressive Integrated Moving Average, Stationarity, Correlogram, Autocorrelation function.*

Introduction

India is one of the leading country in terms of cotton production all over the world because it is economically produced among all agricultural crops.Cotton in a raw form is used by the textile industry. Cotton seeds are used for making Vanaspati in industry and also used as fodder for milch animals for getting good quality milk.India constitute 25 per cent of the total global area under cotton cultivation. It is grown in mostly tropical and subtropical regions where the temperature ranges from 21^0C to 30^0C. There has been a consistent decrease in terms of area under cotton cultivation from 9.1 million hectares to 7.6 million hectares from 2001-02 to 2003-04 respectively. However, the production has been increased manifolds. In year 2000, Indian Government has started a "Technology Mission on Cotton" in order to improve the production and productivity by developing high yielding varieties of cotton and hence increasing the income of cotton growers. For this purpose, estimation of area and

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

forecasting of production becomes essential procedures for policy making in the country. Keeping above in view, ARIMA methodology have been applied in this study using time series data of cotton production. Two most widely used time series models viz., AR and MA models are available in the literature. Combination of these models yields ARIMA model. In this paper, forecasting of cotton production has been done for future. Since, this model was first proposed by Box and Jenkins (1976), hence it is also known as Box-Jenkins model [1]. Wali et al. (2017) have used ARIMA modelling for forecasting of area and production of cotton in India [2]. They have takendata of 65 years i.e., from1950-51 to 2015-16 for prediction of the area and production of cotton. They have selected ARIMA (0, 1, 0) for forecasting of area and ARIMA (1, 1, 1) models for forecasting production of cotton in India respectively on the basis of various model selection criterionfrom 2017-18 to 2020-21. Ghosh (2017) utilized ARIMA for forecasting of export of cotton in India makes short run forecasting of cotton exports in India. He selected ARIMA (1, 1, 0) model using lowest values of AIC and BIC[3]. Using the above selected model, forecasting of next years have been made for cotton exports which shown a rising trend.

The present work has been taken with the objectives:

(i) to investigate the past behavior of cotton production in India
(ii) to test the stationarity of cotton production
(iii) to develop an ARIMA model for cotton production in India
(iv) to forecast the cotton production for next 10 years.

Materials and Methods

Time series data on cotton production of 68 years i.e., from 1950-51 to 2017-18 has been procured from Agricultural Statistics at a Glance 2018 published by Ministry of Agriculture and Farmers welfare, Department of Agriculture, cooperation and farmers welfare, Directorate of Economics and Statistics Govt. of India [4,5].

Autoregressive Integrated Moving Average Model(ARIMA)

For better understanding of the ARIMA model, one should know the following terms:Autoregression (AR): is a model obtained by regressing the variable of interest on its own lagged values. Integration (I): It depicts the number of times the differencing has been done to make series stationary. Moving average (MA): represents the output variable

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

depends linearly on the present and various past values of error term. The general equation of the ARIMA model is given as

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \theta_q e_{t-q} + e_t$$

Where, μ denotes the general mean,$\phi_1, \dots, \phi_p$ represents the estimates of the AR process and$\theta_1, \dots \theta_q$ denotes the estimates of the MA process.

Correlogram: It is a plot of the autocorrelations and partial correlation versus time lag of a time series. A typical ARIMA models has (p, d, q) order where p denotes the order of AR process, d denotes the no of times differencing has been done to make the series stationary and q denotes order of MA process. The patterns in a correlogram are used to analyze underlying behavior of a time series. Autocorrelation represents simple correlation between X_t and, say, X_{t+h}, it is a correlation between a series but with a lag. Mathematically, autocorrelation function is given by

$$\rho(p_{or}q) = \frac{\mathrm{cov}(y_t, y_{t-p_{or}q})}{\mathrm{var}(y_t)}$$

Partial autocorrelation is a relationship between two observations after removing the linear relationship of all observations that fall between those two observations. Mathematically,

$$\phi(p) = \frac{\rho_p - \sum_{j=1}^{p-1}\left(\phi_{p-2,j} - \phi_{pp}\phi_{p-2,p-j}\right)\rho_{p-j}}{1 - \sum_{j=1}^{p-1}\left(\phi_{p-2,j} - \phi_{pp}\phi_{p-2,p-j}\right)\rho_j}, p \geq 3$$

Ljung-Box Q statistic: This test is a used to check whether the observation are independent and random over time. The null hypothesis for this test is that the first k autocorrelations are jointly zero, H_0: $\rho_1 = \rho_2 = \dots = \rho_k = 0$. Alternatively, at least one of $\rho_i$is not zero. Note that it is a test related to autocorrelation. Ljung-Box Q statistic is having asymptotic chi-squaredistribution.

$$Q(p) = T(T+2)\sum_{i=1}^{p}\frac{\rho_i^2}{T-1} \sim \chi_p^2$$

In a time series, time series plot and correlogram can help us in the identification of a suitable model for our time series data.

Stationarity: It means that the statistical properties of a time series generating process do not change over time. More specifically, stationarity means mean, variance, autocorrelation, etc.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

remains constant over time.A time series $Y_t$is said to be stationary if for all t the following holds true:

$$E(Y_t) = m < \infty, V(Y_t) = E(Y_t - m)^2 = g_0 < \infty \text{ and } Cov(Y_t, Y_{t-k})$$
$$= E\{(Y_t - m)(Y_{t-k} - m)\} = g_k$$

Unit root test: This test is used to test the stationarity of a time series and check whether the given time series possesses a unit root. Augmented Dickey fuller (1979) test is most widely used which presumes that the error term in the Model are correlated [6]. ARIMA is a mathematical model designed to forecast data within a time series. Time series is of two types:

- Stationary: A stationary process is one whose statistical properties remains same over time.
- Non- stationary: This process has properties which changes over time. The Box-Jenkins model makes the non- stationary time series stationary by taking the difference between data points.

Box and Jenkins proposed a methods for identification, estimation and diagnostic checking of a time series model. This methodology is now referred to as the Box-Jenkins Methodology.

Identification. It is used to checkthe stationary of a time series. Once a stationary model is found, it can be used for forecasting. If the series is not stationary, it is to be made stationary by differencing the series. The original series is replaced by a series of differencing and an ARIMA model is then fitted for the differenced series.

Estimation and Model Checking: This steps in ARIMA modelling includes diagnostic checking which involves the examination of autocorrelation and partial autocorrelation functions for estimation of residual behaviour. This helps us to identify the areas where the model is inadequate. In this stage, it is tested whether the estimated models has the specifications of univariate time series.

Forecasting: In this step, the estimated model is used to forecast the future values of the given time series.

Results and Discussions

The given dataset have been used for developing the forecasting model. A time series plot has been plotted in order to study the pattern or behaviour of the given time series data which showed that cotton production has the increasing trend from year 1950 to 2018 (Fig.1).

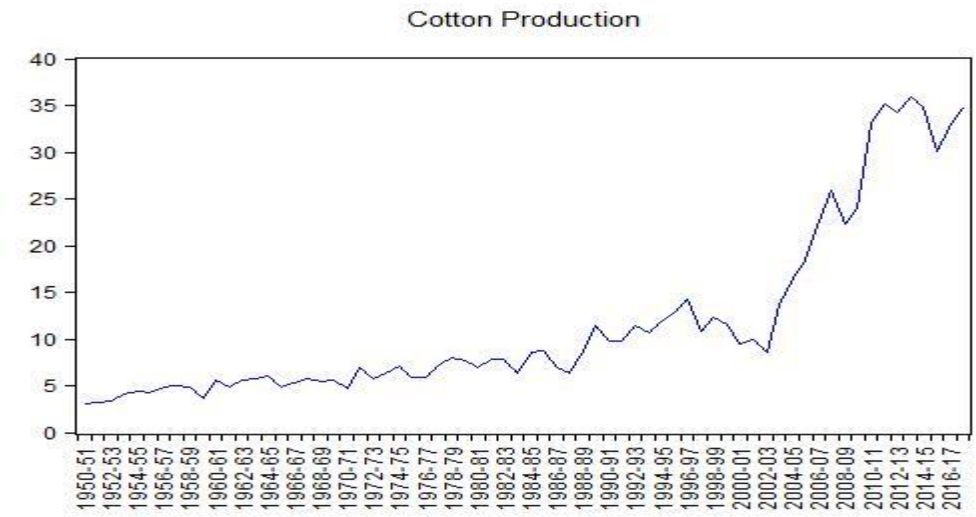


Fig.1: Cotton Production (Million Tonnes) in India from 1950-51 to 2017-18.

Table 1: Unit Root test for checking the stationary of data

			t-Statistic	Sig.
ADF test statistic			-1.091	0.923
Critical values:	1% los		-4.101	
	5% los		-3.478	
	10% los		-3.167	

los denotes level of significance

ADF test is most widely used test for determining stationary of data. The ADF test has the hypothesis that H0: the non-stationary of the data against H1: the data is stationary. The critical value for the rejection of the null hypothesis of unit root test was non-significant with p-value (>.05). The results showed that the data regarding cotton production was non-stationary. Therefore, differencing of the series was required to make the series stationarity.

Table2: ADF test after first differencing

			t-Statistic	Sig.
ADF test statistic			-7.667	0.000
Critical values:	1% los		-3.533	
	5% los		-2.906	
	10% los		-2.591	

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

Table 2 reveals that that ADF test statistics value is 7.667 and its p value is less than .01 meaning that the cotton production sequence after the first-order difference and logarithmic change has turned to be stationary and hence the significance test of the stationarity is passed.

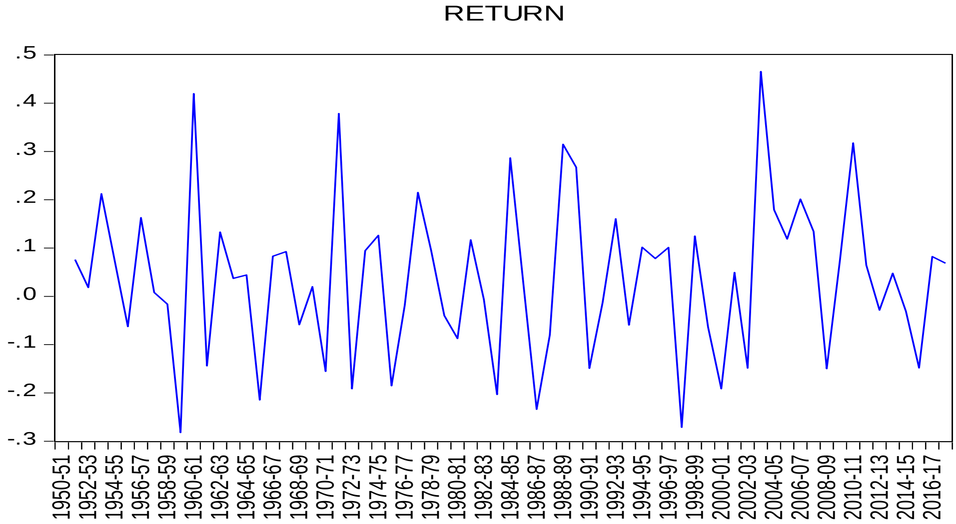
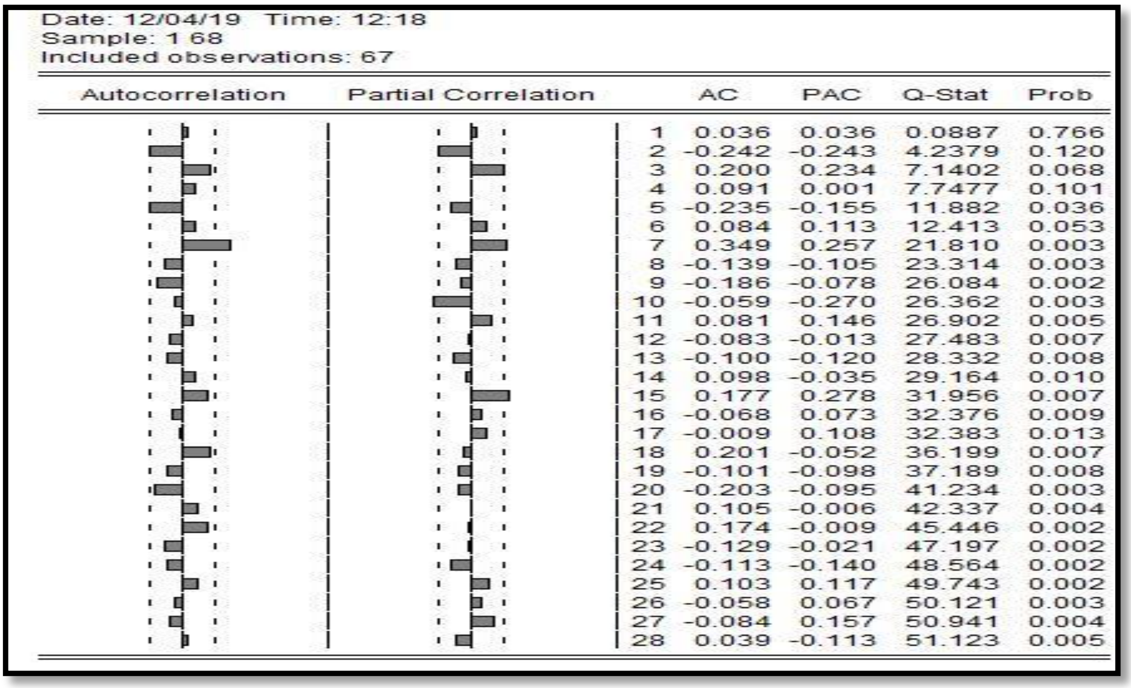


Fig.2: Graph of differenced log series of cotton production



Date: 12/04/19 Time: 12:18
Sample: 1 68
Included observations: 67

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.036	0.036	0.0887	0.766
		2	-0.242	-0.243	4.2379	0.120
		3	0.200	0.234	7.1402	0.068
		4	0.091	0.001	7.7477	0.101
		5	-0.235	-0.155	11.882	0.036
		6	0.084	0.113	12.413	0.053
		7	0.349	0.257	21.810	0.003
		8	-0.139	-0.105	23.314	0.003
		9	-0.186	-0.078	26.084	0.002
		10	-0.059	-0.270	26.362	0.003
		11	0.081	0.146	26.902	0.005
		12	-0.083	-0.013	27.483	0.007
		13	-0.100	-0.120	28.332	0.008
		14	0.098	-0.035	29.164	0.010
		15	0.177	0.278	31.956	0.007
		16	-0.068	0.073	32.376	0.009
		17	-0.009	0.108	32.383	0.013
		18	0.201	-0.052	36.199	0.007
		19	-0.101	-0.098	37.189	0.008
		20	-0.203	-0.095	41.234	0.003
		21	0.105	-0.006	42.337	0.004
		22	0.174	-0.009	45.446	0.002
		23	-0.129	-0.021	47.197	0.002
		24	-0.113	-0.140	48.564	0.002
		25	0.103	0.117	49.743	0.002
		26	-0.058	0.067	50.121	0.003
		27	-0.084	0.157	50.941	0.004
		28	0.039	-0.113	51.123	0.005

Fig.3: Correlogram of cotton production for the period of 1951 to 2017

From fig.3, it can be observed from partial autocorrelations that only one spike is going outside the confidence interval i.e., one significant autocorrelation is present which gives order of AR as one. Likewise if we look at autocorrelation graph, we can find that only one spike is going outside the confidence interval give order of moving average as one. Thus, ARIMA model has possible order as (0,1,1), (1,1,1) and (1,1,0). Among all possible models,

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

ARIMA (1, 1, 1) was found to be best on the basis of minimum values of AIC, SBIC and also on significance of model parameters.

Table 3: AIC and SBC values of ARIMA Models

ARIMA(p,d,q) model	AIC	SBIC	Hannan Quinn
ARIMA(1,1,0)	288.793	295.406	291.410
ARIMA(0,1,1)	288.704	295.318	291.322
ARIMA(1,1,1)	**284.725**	**293.544**	**288.215**

From table 3, it has been found that among all possible models, ARIMA (1, 1, 1) was found to be best on the basis of minimum values of AIC, SBIC , Hannan-Quinn value and also on significance of model parameters for forecasting of cotton production in India .

Table 4: Estimates of the model parameters for production of cotton in India

	Coefficient	Std. Error	z	p-value
constant	0.465	0.261	1.781	0.075
AR(1)	-0.743	0.102	-7.309	0.000
MA(1)	0.968	0.058	16.824	0.000

From the table 4, it has been observed that the estimates of the AR (1) and MA (1) were found to be highlysignificant and hence can be used for forecasting of cotton production in India.

Table 5: Forecasting of cotton production India using ARIMA (1, 1, 1) model.

Year	Prediction	Std. Error	95% interval (lower confidence limit, upper confidence limit)
2019	34.812	1.894	(31.099, 38.524)
2020	35.680	2.994	(29.811, 41.548)
2021	35.845	3.602	(28.784 42.905)
2022	36.533	4.241	(28.221, 44.845)
2023	36.832	4.716	(27.588, 46.075)
2024	37.420	5.201	(27.226, 47.614)
2025	37.793	5.608	(26.809, 48.784)
2026	38.326	6.012	(26.542, 50.110)
2027	38.740	6.374	(26.248, 51.232)
2028	39.243	6.728	(26.056, 52.429)

Table 5 shows the forecasted values of cotton production in India using ARIMA (1, 1, 1) model for next 10 years which reveals that the production of cotton will increase in next 10 years.

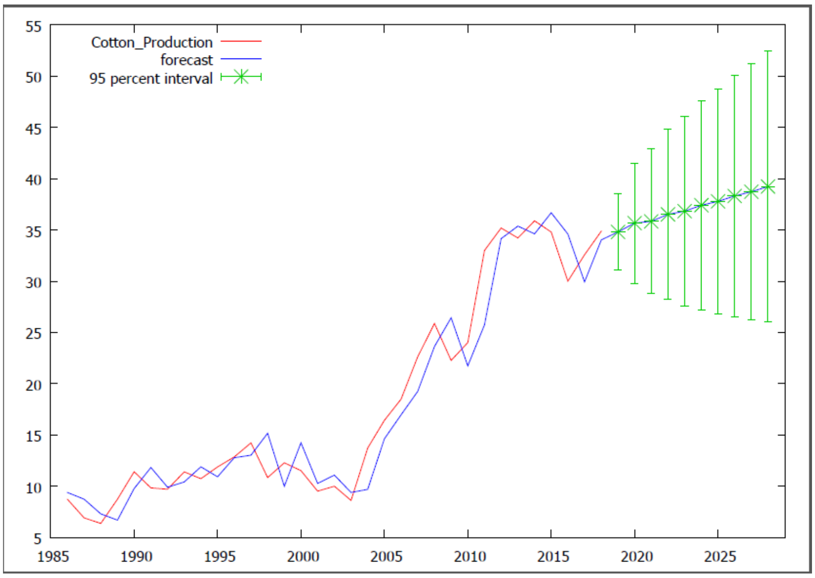


Fig.4: Forecasting of cotton production in India (2019-2028) using ARIMA (1, 1, 1) model.

The graph of forecasting of cotton production in India shows displays that the production will increase in coming years. The ends of the green represents the upper and lower confidence intervals.

Conclusion

In the present work, past behaviour of the time series data of cotton production have been studied which have shown an overall increasing trend. The ADF test of the given time series was found to be non-significant which indicates that the series is not stationary. For making it stationary, first differencing has been done. This stationary time series have been then used for developing ARIMA model in order to obtain the forecasted values of Indian cotton production of next ten years. The results shows that ARIMA(1,1,1) was better than other ARIMA models like ARIMA(1,1,0) and ARIMA(0,1,1) on the basis of lowest values of some model selection criteria's*viz.*, AIC, SBIC and Hannan Quinn values. The forecasted values showed that there will be increase in cotton production in coming 10 years.s

References

1. G.E. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, 1976.
2. V.B. Wali, D. Beeraladinni, and H. Lokesh, "Forecasting of Area and Production of Cotton in India: An Application of ARIMA Model".*Int. J. Pure App. Biosci.* Vol.5, pp. 341-347, 2017.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

3. S. Ghosh, "Forecasting Cotton Exports in India using the ARIMA model". Amity Journal of Economics, Vol. 2 pp. 36-52, 2017.
4. www.agricoop.nic.in
5. http://eands.dacnet.nic.in
6. D.A. Dickey and W.A. Fuller, "Distribution of estimators for Autoregressive Time Series with a Unit root". Journal of the American Statistical Association Vol. 74, pp. 427-431, 1979