J. Cotton Res. Dev. **28** (2) 342-345 (July, 2014)

Forecasting cotton (*Gossypium* spp) production in India

D.J.CHAUDHARI*, C.J.CHAUDHARI AND A.S.TINGRE
Department of Agricultural Economics and Statistics, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola- 444 104
**E-mail : datta1616@rediffmail.com*

ABSTRACT : The present study aimed to forecast the cotton production in India by using the time series cotton production data for the period from 1950-1951 to 2010-2011. To forecast the cotton production ARIMA models, introduced by Box and Jenkins were used. To test the reliability of model R^2, Mean Absolute Percentage Error (MAPE), and Bayesian Information Criterion (BIC) were used. Among the different class of ARIMA models, lowest BIC value worked out to be 254 for ARIMA (0,1,0), which was the best fitted model. Based on model results the estimated cotton production in India would increase from 33.93 million bales during the year 2011-2012 to 37.98 million bales during the year 2019-2020.

Key words : ACF, ARIMA, auto regression, cotton, forecasting, moving average, non stationary, PACF

Cotton is one of the major commercial crops grown in India. India ranks first in the world for the total area planted under cotton and second in cotton production next to China. The major cotton growing states are Punjab, Haryana, Rajasthan, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Tamil Nadu and Karnataka. With a objective to improve the quality of cotton, enhance/ha productivity, enhance the income of cotton growers by reducing the cost of cultivation, to improve the processing facilities etc. The Government of India has launched "Technology Mission On Cotton" in February, 2000. Increased global demand for cotton should induce higher production in the next decade. Considering these points, it is necessary to know the extent of cotton production in future with available resources. The present study has been under taken with a objective to forecast the cotton production in near future of India.

The present study was conducted during 2012. ARIMA model, introduced by Box and Jenkins (1970), was frequently used for discovering the pattern and predicting the future values of the time series data. While Najeeb (2005) used ARIMA model for forecasting wheat area and production in Pakistan. He reported that ARIMA (1,1,1) and (2,1,2) was the best fitted model for area and production of wheat. Nasiru and Solomon (2012) forecasted milled rice production in Ghana by using Box Jenkins approach. The results showed the increasing trend in production of rice in next ten years. Padhan (2012) applied ARIMA model for forecasting agricultural productivity in India. Seydou and Ying (2012) forecasted niger grains production in China by using ARIMA model. They found ARIMA (1,1,0) was best fitted model for the niger grain production.

Stochastic time series ARIMA models were widely used in time series data having the characteristics of parsimonious, stationary, invertible, significant estimated coefficients and statistically independent and normally distributed residuals. When a time series is non stationary, it can often be made stationary by taking first differences of the series *i.e.*, creating a new time series of successive differences (Y_t-Y_{t-1}). If first differences do not convert the series to stationary form, then second differences can be created. This is called second order differencing. A distinction is made between a second order differences (Y_t-Y_{t-2}).

ARIMA process are mathematical models used for forecasting. The ARIMA approach is based on the two ideas *i.e.* the forecast are based on linear functions of the sample observations and the aim is to find out the simplest models that provide an adequate description of the observed data. This also called principle of parsimony. The time series when differenced follows both AR and MA models and is known as autoregressive integrated moving averages (ARIMA) model. The model are often written in short hand as ARIMA (p,d,q) where 'p' describes the AR part, 'd' describes the integrated part and 'q' describe the MA part. ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated using the Statistical Package for Social Sciences (SPSS) and to fit the ARIMA models.

Autoregressive process of order (p) is,

$$Y_t = \text{ì} + \text{Ø}_1 Y_{t-1} + \text{Ø}_2 Y_{t-2} + \ldots\ldots\ldots\ldots + \text{Ø}_p Y_{t-p} + e_t$$

Moving Average process of order (q) is,

$$Y_t = \text{ì} - \text{ʿ}_1 e_{t-1} - \text{ʿ}_2 e_{t-2} - \ldots\ldots\ldots\ldots - \text{ʿ}_q e_{t-q} + e_t$$

and the general form of ARIMA model of order (p, d, q) is

$$Y_t = \text{Ø}_1 Y_{t-1} + \text{Ø}_2 Y_{t-2} + \ldots\ldots\ldots + \text{Ø}_p Y_{t-p} + \text{ì} - \text{ʿ}_1 e_{t-1} - \text{ʿ}_2 e_{t-2} - \ldots\ldots\ldots - \text{ʿ}_q e_{t-q} + e_t$$

where Y_t is cotton production, e_t 's are independently and normally distributed with zero mean and constant variance for t = 1,2,..., n; d is the fraction differenced while interpreting AR and MA and Ø_p and ʿ_q are coefficients to be estimated.

The best model is obtained with the following diagnostics, by lowest values of Akaike's Information Criteria (AIC) and Schwartz Bayesian Criteria (SBC or BIC). To check the adequacy for the residuals using Q statistic. A modified Q statistic is the Box-Ljung Q statistic given as

$$Q = \frac{N(n+2)\text{£}rk^2}{(n-k)}$$

Where rk : the residual autocorrelation at lag k

n : the number of residuals

The Q statistic is compared to critical value of Chi squre distribution. If the p value associated with Q statistic is small, the model is consider inadequate. Forecasting the future periods using the parameters for the tentative model has been selected.

Trend fitting: For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Mean Absolute Percentage Error (MAPE), and Bayesian Information Criterion (BIC) were used. Lesser the various reliability statistics better will be the efficiency of the model in predicting the future production. The time series data related to cotton production in India is collected from the official website of Department of Cooperation, Ministry of Agriculture, Government of India for the period from 1950-1951 to 2010-2011. Using the data forecasting of cotton production was done upto 2019-2020.

Model identification: ARIMA model is estimated only after transforming the variable under forecasting into a stationary series. The stationary series is the one whose values vary over time only around a constant mean and constant variance. There are several ways to ascertain this. The most common method is to check stationarity through examining the graph or time plot of the data. Non stationarity in mean is corrected through appropriate differencing of the data. The newly constructed variable Y_t was stationary in mean, the next step is to identify the values of p and q. For this Autocorrelation (ACF) and Partial Autocorrelation (PACF) of various orders of Y_t were computed and presented in Table 1. The various ARIMA models were fitted. The model which had minimum normalized BIC value was chosen as a best fit model for forecasting the future values of cotton production. The various ARIMA models and their AIC and normalized BIC values are presented in Table 2. showed that ARIMA (0,1,0) had the lowest normalized BIC value.

Model estimation: By using SPSS package the model parameter were estimated and presented in Table 3. From this table it was observed that the R^2 was 0.66. The value of the normalized BIC was lowest and worked out to 254 for the ARIMA (0,1,0) and the MAPE value recorded to 13.8, indicated that ARIMA (0,1,0) was the most suitable model for forecasting cotton production in India.

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, which has been done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, various autocorrelations upto 16 lags were computed and the same along with their significance tested by Box Ljung statistic are provided in Table 4. As the results indicated, none of these autocorrelations was significantly different from zero at any reasonable level. This proved that the selected ARIMA model was an appropriate model for forecasting cotton

Table 1. ACF and PACF of cotton production in India

Lag	Auto correlation function (ACF)		Box Ljung Stat	Partial auto correlation function (PACF)	
	Value	Stand error		Value	Stand error
1	0.801	0.125	41.13	0.801	0.128
2	0.702	0.124	73.23	0.167	0.128
3	0.636	0.123	100.04	0.037	0.128
4	0.505	0.122	117.25	-0.170	0.128
5	0.390	0.121	127.67	-0.100	0.128
6	0.321	0.120	134.86	0.028	0.128
7	0.272	0.119	140.14	0.084	0.128
8	0.227	0.117	143.86	0.034	0.128
9	0.243	0.116	148.24	0.147	0.128
10	0.247	0.115	152.86	0.023	0.128
11	0.272	0.114	158.53	0.086	0.128
12	0.252	0.113	163.51	-0.119	0.128
13	0.227	0.112	167.63	-0.067	0.128
14	0.239	0.111	172.31	0.082	0.128
15	0.199	0.109	175.63	-0.032	0.128
16	0.154	0.108	177.66	-0.26	0.128

ACF : Auto Correlation Function, PACF : Partial Auto Correlation Function

Table 2. AIC and SBC values of ARIMA

ARIMA(p,d,q)	AIC	SBC
ARIMA(1,0,0)	266	270
ARIMA(1,1,0)	254	259
ARIMA(1,1,1)	251	258
ARIMA(0,1,1)	254	259
ARIMA(0,0,1)	334	338
ARIMA(1,0,1)	268	274
ARIMA(0,1,0)	**252**	**254**

‘p’ – Auto regression (AR); ‘d’ – Integration (I); ‘q’ – Moving Average (MA)

Table 3. Estimates of the ARIMA model fitted for cotton production.

Parameters	Estimates	SE	t value	Sig.
Constant	0.5064	0.2547	1.9883	0.0514
Numbers of residuals	60			
Log likelihood	-125.40			
Df	59			
R-^{2}	0.66			
MAPE	13.8			
SBC (BIC value)	254			

Table 4. Residuals of ACF and PACF of cotton production in India

Lag	Auto correlation function (ACF)		Box Ljung Stat	Partial auto correlation function (PACF)	
	Value	Stand error		Value	Stand error
1	0.006	0.126	0.002	0.006	0.129
2	-0.179	0.125	2.057	-0.179	0.129
3	0.210	0.124	4.938	0.219	0.129
4	0.162	0.123	6.684	0.129	0.129
5	-0.056	0.122	6.897	0.013	0.129
6	0.012	0.120	6.907	0.018	0.129
7	0.233	0.119	10.722	0.181	0.129
8	-0.077	0.118	11.141	-0.102	0.129
9	-0.079	0.117	11.156	-0.013	0.129
10	-0.111	0.116	12.518	-0.254	0.129
11	0.159	0.115	14.436	0.167	0.129
12	-0.036	0.114	14.537	-0.090	0.129
13	-0.227	0.112	18.610	-0.109	0.129
14	0.115	0.111	19.682	0.066	0.129
15	0.126	0.110	21.003	0.125	0.129
16	-0.057	0.109	21.277	0.058	0.129

ACF : Auto Correlation Function, PACF : Partial Auto Correlation Function

Table 5. Forecast of cotton production in India. (Quantity in million bales)

Year	Actual cotton production	Forecasted cotton production
2011-2012	—	33.93
2012-2013	—	34.44
2013-2014	—	34.94
2014-2015	—	35.45
2015-2016	—	35.96
2016-2017	—	36.46
2017-2018	—	36.97
2018-2019	—	37.48
2019-2020	—	37.98

production in India. Hence, the fitted ARIMA model for the cotton production data was

$$Y_t = 0.5064 – Y_{t-1} - e_t \text{ ....}(1)$$

Forecasting : ARIMA models are developed basically to forecast the corresponding variable. There are two kinds of forecasts: sample period forecasts and post-sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both these kinds of forecasts.

Sample period forecasts: The sample period forecasts are obtained simply by plugging the actual values of the explanatory variables in the estimated equation (1).The explanatory variables here are the lagged values of Y_t and the estimated lagged errors. To judge the forecasting ability of the fitted ARIMA model, important measures of the sample period forecasts' accuracy were computed. The Mean Absolute Percentage Error (MAPE) for cotton production worked out to be 13.8. This measure indicates that the forecasting inaccuracy is low.

Post sample forecasts: The principal objective of developing an ARIMA model for a variable is to generate post sample period forecasts for that variable. This is done through using equation (1). Based on the fitted model forecasting of cotton production in India was done for the period from 2011-2012 to 2019-2020 and presented in Table 5. From the table, the forecasted values showed that the cotton production will increase from 33.93 million bales during 2011-2012 to 37.98 million bales during 2019-2020.

REFERENCES

Box, G. E. P. and Jenkins, J. M. 1970. Time Series Analysis -Forecasting and Control. Holden-Day Inc., San Francisco, CA.

Najeeb, Iqbal, Khuda, Bakshi, Asif, Maqbool and Abid Shohab, Ahmad 2005. Use of the ARIMA model for forecasting wheat area and production in Pakistan. *J. Agri. Soc. Sci.,* **1** : 120-22.

Nasiru, Suleman and Solomon, Sarpong 2012. Forecasting milled rice production in Ghana using Box Jenkins approach. *Int. J. Agric. Manag. Dev.* **2** : 79-84.

Padhan, Purna Chandra 2012. Application of ARIMA Model for Forecasting Agricultural Productivity in India. *J. Agric. Soc. Sci.*, **8 :** 50-56

Seydou, Zakari and Liu, Ying .2012. Forecasting of niger grain production and harvested area. *Asian J. Agric. Sci.* **4** : 308-13.

Received for publication : December 22, 2012
Accepted for publication : March 1, 2013