FORECASTING OF COTTON PRODUCTION IN INDIA USING ARIMA MODELS

Prema Borkar[1],

Assistant Professor,
Gokhale Institute of Politics and Economics (Deemed to be a University),
Pune - 411004. Maharashtra.

P. M. Tayade[2]

Research Scholar,
Department of Agricultural Economics,
Vasantrao Naik Marathwada Krishi Vidyapeeth,
Parbhani – 431402. Maharashtra.

ABSTRACT

The paper describes an empirical study of modeling and forecasting time series data of Cotton Production in India. The Box Jenkins ARIMA methodology has been used for forecasting. The diagnostic checking has shown that ARIMA (2, 1, 1) is appropriate. The forecasts from 2015-16 to 2020-2021 are calculated based on the selected model. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of grain storage, import and/or export and adopt appropriate measures in this regard.

Key Words: Arima models; Cotton Production; Forecasting

INTRODUCTION

Cotton, the ***'white gold'*** or the ***"King of Fibres",*** enjoys a predominant position amongst all cash crops in India. In India, cotton occupies an area of nearly 117.27 million hectares, with a production of 398 lakh bales (2013-14), ranking third in the world after China and USA which accounts for about 18 per cent of the world cotton production. It has also the distinction of having the largest area under cotton cultivation in the world constituting about 27 per cent of the world area under cotton cultivation. The lint productivity of cotton is 577 kg/ ha, which is the lowest and far below that of the world average of 756 kg/ha. During last fifty years, production of cotton rose from 30 lakh bales (1 bale = 170 kg) in 1950-51 to 398 lakh bales in 2013-14. During the same period the area under cultivation increased from 56.48 lakh hectares to 112.42 lakh hectares. Significant increase in the area under cultivation of cotton was observed over a period of fifty years. Nearly 65 per cent cotton cultivation is rain dependent and subject to heavy vagaries of monsoon rains.

Forecasts have traditionally been made using structural econometric models. Concentration have been given on the univariate time series models known as auto regressing integrated moving average (ARIMA) models, which are primarily due to world of Box and Jenkins (1970). These models have been extensively used in practice for forecasting economic time series, inventory and sales modeling (Brown, 1959; Holt *et al.,* 1960) and are generalization of the exponentially weighted moving average process. Several methods for identifying special cases of ARIMA models have been suggested by Box-Jenkins and others. Makridakis *et al.* (1982), and Meese and Geweke (1982) have discussed the methods of identifying univariate models. Among others Jenkins and Watts (1968), Yule (1926, 1927), Bartlett (1964), Quenouille (1949), Ljune and Bos (1978) and Pindyck and Tubinfeld (1981) have also emphasized the use of ARIMA models.

In this study, these models were applied to forecast the production of cotton crop in India. This would enable to predict expected cotton production for the years from 2016 onward. Such an exercise would enable the policy makers to foresee ahead of time the future requirements for grain storage, import and/or export of cotton thereby enabling them to take appropriate measures in this regard. The forecasts would thus help save much of the precious resources of our country which otherwise would have been wasted.

MATERIALS AND METHODS

Our aim is to model the annual production of cotton for the years 1950-51 to 2014-15 and to forecast the production of future years. A choice is often to be made as to which type of the model should be developed. This choice may be difficult. A choice was made from a class of liner time series models introduced by Box and Jenkins (1970), which are now widely used and accepted. According to them, the ARIMA model is denoted by ARIMA (p, d, q)

Where,

'p' is the order of the autoregressive process;

'd' is the order of homogeneity i.e. the number of difference to make the series stationary; and

'q' the order of the moving average process.

The general from of the ARIMA (p, d, q) is

$$Z_t = C + (F_1 Z_{t-1} + ... + F_p Z_{t-p}) - (\theta_1 a_{t-1} + ... + \theta_q a_{t-q}) + a_t$$

Here C is a constant.

$Z_{t-1}, ..., Z_{t-p}$ are pasts series values (lags), the Φi's are the coefficients, similar to regression coefficients, to be estimated of the autoregressive model where autoregressive (AR) model of order p, denoted by AR (p) is

$$Z_t = C + F_1 Z_{t-1} + F_2 Z_{t-2} + \theta_p a_{t-p} + a_t$$

a_t is a random variable with zeros mean and constant variance.

Θ's are coefficients in the moving average (MA) model, where moving average model of order q or MA (q) is

$$Z_t = a_t + \theta_1 a_{t-1} - \theta_2 a_{t-2} \ldots - \theta_q a_{t-q}$$

RESULTS AND DISCUSSION

The maximum production of cotton in India was obtained in the year 2013-14 (398 lakh bales) and minimum in 1950-51 (30.4 lakh bales). The ARIMA model was applied according to four steps namely model specification, model estimation, diagnostic checking and forecast. These four steps are explained in the text while discussing forecasts of cotton production.

Forecasts of cotton production up to year 2020-21.

Last sixty four year's data of production of cotton was used for modeling purpose. The model specification involved the plots of the auto correlation function (ACF), partial auto correlation function (PACF) and the plot of the differenced series. Auto correlation function indicated the order of the autoregressive components 'q' of the model, while the partial correlation function gave an indication for the parameter p. First step was to check the stationarity of the data. The time series plot of production showed an increasing trend. Auto correlation function showed non-stationary as auto correlation function did not fall as quickly as the log (k) increased. To make the series stationary for production, differenced series was used and first difference series of cotton production showed stationarity. To check the further stationarity, second difference of the original series for production was taken. The autocorrelation function of second difference series and correlogram showed some more stationarity than that of the first difference. Time series data of original series of cotton production was not much beneficial to get a stationary series, therefore, the value of parameter 'd' was decided to be equal to '1'.

After this the values of autoregressive (AR) parameter "p" and moving average (MA) parameter 'q' was determined from correlograms of partial autocorrelation function and the auto correlation function, respectively. The correlogram of auto correlation function of first differenced series showed that the auto correlation function falls quickly after lag 1 for production, hence, the respective values of the parameter "q" was decided to be 1. Partial auto correlation function of the first differenced series of production was used to determine the parameter "p". It was observed that partial auto correlation function fell after lag 2 for production. Thus the value of "p" was decided equal to 2 for production which gave good results. Consequently, the respective values of p, d, q were determined for ARIMA i.e. ARIMA (2, 1, 1).

The model ARIMA (2, 1, 1) is estimated using the statistical software SPSS and the results are presented in Table 1.

Forecasts of cotton production

ARIMA (2, 1, 1) were taken for 6 year ahead and forecasts for cotton production which are given in Table 2 along with 95 per cent confidence interval values. For 2015-16 forecast of cotton production was about 29.73 lakh bales with lower and upper limits of 19.40 and 40.06 respectively. A cotton production forecast for the year 2020-21 was 28.99 lakh bales with lower and upper limits of 16.02 and 41.95 lakh bales, respectively.

SUGGESTIONS

Selection of high yielding varieties, massive education of farmers through a network of agriculture officers to make improvement in cultural practices, adequate supply and inputs and full scale use of latest technology are important to increase the production in future.

The supply projections of an agricultural commodity especially cotton plays an important role in the adjustment of supply and demand in the future. These projections help the government to make policies with regard to relative price structure, production and consumption and also to establish relations with other countries of the world.

REFERENCES

C.C. Holt, F. Modigliani, J.F. Muth and H.A. Simon. 1960. *Planning, Production, Inventores, and Work Force*. Prentice Hall, Englewood Cliffs, NJ, USA.

F. Muhammad, M. S. Javed and M. Bashir. 1992. Forecasting Sugarcane Production in Pakistan using ARIMA Models. Pak. J. Agri. Sci., 9(1): 31-36.

G.E.P. Box, and D.A. Pierce. 1970. Distribution of Residual Autocorrelations in Atuoreressive-Integrated Moving Average Models. *J. American Stat. Assoc.,* 65: 1509–26.

G. M. Jenkins and D.G. Watts. 1968. *Spectral Analysis and its Application,* Day, San Francisco, California, USA.

G.U. Yule. 1926. Why Do We Sometimes Get Nonsence-corrleations Between Times Series. A study in Sampling and the Nature of Series. *J. Roy. Stat. Soc.,* 89: 1–69.

Ljunge, G.M. and G.E.P. Box. 1978. On a Measure of Lack of Fit in Time Series Models. *Biometrika,* 65: 67–72.

M.S. Bartlett. 1964. On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series. *J. Roy. Stat. Soc.,* B 8: 27–41.

M.H. Quenouille. 1949. Approximate Tests of Correlation in Time- Series. *J. Roy. Stat. Soc.,* B11: 68–84.

N. Iqbal, K. Bakhsh, A. Maqbool and A. S. Ahmad. 2005. Use of the ARIMA Model for forecasting Wheat area and production in Pakistan". Journal of Agriculture & Social Sciences, 2: 120-122.

N. Saeed, A. Saeed, M. Zakria, T. M. Bajwa. 2000. Forecasting of Wheat Production in Pakistan using ARIMA models. International Journal of Agriculture and Biology., 4: 352-353.

R.G. Brown. 1959. *Statistical Forecasting For Inventory Control*. New York, McGraw-Hill.

R. Meese and J. Geweke. 1982. A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series. Unpublished Manuscript, University of California, Berkeley, CA, USA.

R.S. Prindycke and D.L. Rubinfeld. 1981. *Econometric Models and Economic Forecasts,* 2nd Ed. New York, McGraw-Hill.

S. Makridakis, A. Anderson, R. Filds, M. Hibon, R. lewandowski, J. Newton, E. Parzen and R. Winkler. 1982. The Accuracy of Extrapolation (time series) methods: Results of a Forecasting Competition, Journal of Forecasting Competition. *J. Forecasting*, 1: 111–53.

Table 1: Estimates of the fitted ARIMA model

		Estimates	Std Error	T	Approx sig
Non- Seasonal lag	AR1	0.9896	0.2152	45.9893	0.0000
Constant		15.1904	11.3447	1.3389	0.1859
Number of Residuals	61				
Number of Parameters	1				
Residual df	59				
Adjusted Residual Sum of Squares	207.1913				
Residual Sum of Squares	292.4865				
Residual Variance	3.4606				
Model Std. Error	1.8603				
Log-Likelihood	-119.236				
Akaike's Information Criteria (AIC)	242.472				
Schwarz's Bayesian Criterion (BIC)	246.5929				

Table 2: Forecasts for Cotton Production (2015-16 to 2020-2021)

(lakh bales)

Period	Forecast	Lower	Upper
2015-16	29.73	19.40	40.06
2016-17	29.58	18.65	40.50
2017-18	29.43	17.95	40.91
2018-19	29.28	17.28	41.28
2019-20	29.13	16.64	41.63
2020-21	28.99	16.02	41.95