Visit us - www.researchjournal.co.in ■ DOI : 10.15740/HAS/IRJAES/8.2/281-286
International Research Journal of Agricultural Economics and Statistics
Volume 8 | Issue 2 | September, 2017 | 281-286 ■ e ISSN-2231-6434

Research Paper

Study on modeling and forecasting of tobacco production in India

■ PREMA BORKAR AND V. M. BODADE

See end of the paper for authors' affiliations

Correspondence to :
PREMA BORKAR
Gokhale Institute of Politics and Economics, PUNE (M.S.) INDIA

ABSTRACT : The paper describes an empirical study of modeling and forecasting time series data of tobacco production in India. Yearly tobacco production data for the period of 1950-1951 to 2014-2015 of India were analyzed by time-series methods. Autocorrelation and partial autocorrelation functions were calculated for the data. The Box Jenkins ARIMA methodology has been used for forecasting. The diagnostic checking has shown that ARIMA (1, 1, 1) is appropriate. The forecasts from 2015-2016 to 2019-2020 are calculated based on the selected model. The forecasting power of autoregressive integrated moving average model was used to forecast tobacco production for five leading years. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of tobacco production, import and/or export and adopt appropriate measures in this regard.

KEY WORDS : ACF - Autocorrelation function, ARIMA - Autoregressive integrated moving average, Forecast, PACF - Partial autocorrelation function, Tobacco

***Paper History* :**
Received : *16.01.2017;*
Revised : *17.07.2017;*
Accepted : *29.07.2017*

HOW TO CITE THIS PAPER : Borkar, Prema and Bodade, V.M. (2017). Study on modeling and forecasting of tobacco production in India. *Internat. Res. J. Agric. Eco. & Stat.,* **8** (2) : 281-286, **DOI : 10.15740/HAS/IRJAES/8.2/281-286.**

INTRODUCTION :

India is the world's second largest producer of tobacco, endowed with rich agro-climatic attributes such as fertile soils, rainfall and ample sunshine. India produces various types of tobacco. Currently, Indian tobacco is exported to more than 100 countries spread over all the continents. A few of the top multinational companies such as British American Tobacco (BAT), Philip Morris, RJ Reynolds, Seita, Imperials, Reemtsma etc. and many companies with government monopoly all over the world import Indian tobacco either directly or indirectly. Over the years, a combination of strong prices, domestic consumption, good export demand for tobacco and low prices of other crops helped the growth of tobacco from a cash crop to a manufacturing industry linked with commercial considerations. The tobacco industry in India includes the production, distribution and consumption of (i) leaf tobacco, (ii) smoking products such as cigarettes and beedis and (iii) various chewing tobacco products.

It is a robust and largely irrigation-independent crop, provides substantial employment, has significant export potential and most importantly, is a source of ever-growing tax revenues on one hand. On the other, there are public health concerns about the effects of smoking and consumer-led lobbies asking for more controls on cigarette sales, smoking and advertising. In spite of its proven adverse implications for public health, the industry continues to be supported in many quarters on the grounds of its contribution to employment and national revenue.

The organized sector of the industry, dominated by multinational corporations, is at the forefront of canvassing support for the sector.

The total area and production of tobacco in India for the year 2013-14 were 0.46 million hectares and 0.74 million tonnes, respectively. It occupies a meagre 0.24 per cent of the country's total arable land area. India ranks 4th in the total tobacco consumption in the world. But India's cigarette consumption ranks 11th in the world. Out of the total production, only 19 per cent of the total consumption of tobacco is in the form of cigarette whereas 81 per cent is in other forms like, chewing, bidi, snuff, Gutka paste, Jarda, hookah paste etc. The per capita consumption of cigarette in India is one of the lowest in the world in comparison to major tobacco consuming countries like Zimbabwe, UK, Brazil, U.S.A. and Pakistan. The annual level for demand of cigarette in India remains the same as it was 15 years ago, despite the cumulative growth in population during the same period. However the consumption of tobacco has been a matter of national debate in view of the emerging anti tobacco drive in the country. India is one of the leading tobacco exporting countries in the world. The principal markets for Indian tobacco are U.S.S.R, U.K., Japan and the Middle East countries.

Forecasts have traditionally been made using structural econometric models. Alteration has been given to the univariate time series models known as auto regressing integrated moving average (ARIMA) models, which are primarily due to the work of Box and Jenkins (1970). These models have been extensively used in practice for forecasting economic time series, inventory and sales modeling (Brown, 1959 and 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 and 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 and 1927); Bartlett (1964); Quenouille (1949); Ljunge and Box (1978) and Pindycke and Rubinfeld (1981) have also emphasized the use of ARIMA models.

In this study, these models were applied to forecast the production of tobacco crop in India. This would enable to predict expected tobacco production for the years from 2016 onward. Such an exercise would enable the policy makers to foresee ahead of time the future requirements for tobacco production, import and/or export of tobacco 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 :

The time series data of tobacco production in India has been collected from the website of Directorate of Economics and Statsitics, Department of Agriculture and Co-operation, Ministry of Agriculture from 1950-51 to 2014-15. Box and Jenkins (1976) linear time series model was applied. Auto regressive integrated moving average (ARIMA) is the most general class of model for forecasting a time series. Different series appearing in the forecasting equations are called "Auto-regressive" process. Appearance of lags of the forecast errors in the model is called "moving average" process. The ARIMA model is denoted by ARIMA (p,d,q),

where,

"p" stands for the order of the auto regressive process,

"d" is the order of the data stationary and

"q" is the order of the moving average process.

The general form of the ARIMA (p,d,q) can be written as described by Judge *et al*. (1988).

$$a^d y_t = ¢ + {}^{-}_{1} a^d y_{t-1} + {}^{-}_{2} a^d y_{t-2} + \text{----} + {}^{-}_{p} y_{t-p} + e_{t-1} \sim e_{t-1} - {}^{\sim}_{2} e_{t-2} \sim_q e_{t-2} \quad \text{...(1)}$$

where,

Δ^d denotes differencing of order d, *i.e.*, $\Delta y_t = y_t - y_{t-1}$, $\Delta_2 y_t = \Delta y_t - \Delta_{t-1}$ and so forth,

Y_{t-1} ------ y_{t-p} are past observations(lags),

δ, θ_1 ------ θ_p are parameters (constant and co-efficient) to be estimated similar to regression co-efficients of the auto regressive process (AR) of order "p" denoted by AR (p) and is written as :

$$Y = ¢ + {}^{-}_{1} y_{t-1} + {}^{-}_{2} y_{t-2} + \text{-----} + {}^{-}_{p} y_{t-p} + e_t \quad \text{....(2)}$$

where,

e_t is forecast error, assumed to be independently distributed across time with mean θ and variance

${}^{-}_{2} e, e_{t-1}, e_{t-2}$ ------- e_{t-q} are past forecast errors,

$\sim_1, \text{--------} \sim_q$ are moving average (MA) co-efficient that needs to be estimated.

While MA model of order q (*i.e.*) MA (q) can be written as:

$$Y_t = e_t - \sim_1 \sim_{t-1} - \sim_2 e_{t-2} \text{ ------- } \sim_q e_{t-q} \quad \text{....(3)}$$

The major problem in ARIMA modeling technique is to choose the most appropriate values for the p, d and

q. This problem can be partially resolved by looking at the Auto correlation function (ACF) and partial auto correlation functions (PACF) for the series (Pindycke and Rubinfeld, 1981). The degree of the homogeneity, (d) *i.e.* the number of time series to be differenced to yield a stationary series was determined on the basis where the ACF approached zero.

After determining "d" a stationary series $\Delta d\ y_t$ its auto correlation function and partial autocorrelation were examined to determined values of p and q, next step was to "estimate" the model. The model was estimated using computer package "SPSS".

Diagnostic checks were applied to the so obtained results. The first diagnostic check was to draw a time series plot of residuals. When the plot made a rectangular scatter around a zero horizontal level with no trend, the applied model was declared as proper. Identification of normality served as the second diagnostic check. For this purpose, normal scores were plotted against residuals and it was declared in case of a straight line. Secondly, a histogram of the residuals was plotted. Finding out the fitness of good served as the third check. Residuals were plotted against corresponding fitted values: Model was declared a good fit when the plot showed no pattern.

Using the results of ARIMA (p,q,d), forecasts from 2016 upto 2020 were made. These projections were based on the following assumptions.

– Absence of random shocks in the economy, internal or external.

– Agricultural price structure and polices will remain unchanged.

– Consumer preferences will remain the same.

RESULTS AND DATA ANALYSIS :

The results obtained from the present investigation as well as relevant discussion have been summarized under following heads :

Building ARIMA model for tobacco production data in India :

To fit an ARIMA model requires a sufficiently large data set. In this study, we used the data for tobacco production for the period 1950-1951 to 2014-2015. As we have earlier stated that development of ARIMA model for any variable involves four steps: identification, estimation, diagnostic checking and forecasting. Each of these four steps is now explained for tobacco production. The time plot of the tobacco production data is presented in Fig. 1.

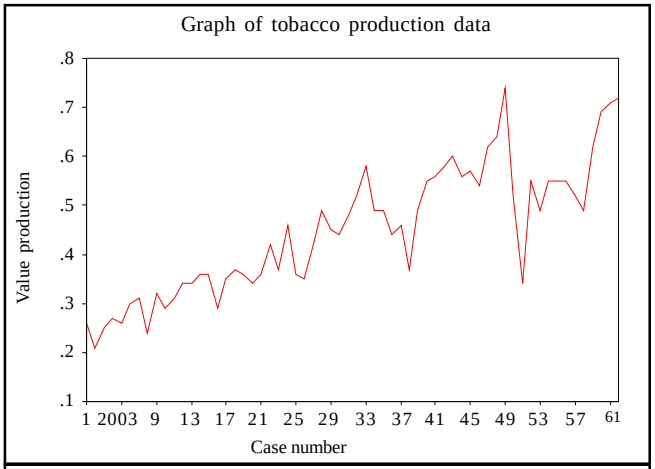

Fig. 1 : Time plot of tobacco production data

The above time plot indicated that the given series is non-stationary. Non-stationarity in mean is corrected through appropriate differencing of the data. In this case difference of order 1 was sufficient to achieve stationarity in mean.

The newly constructed variable X_t can now be examined for stationarity. The graph of X_t was stationary in mean. The next step is to identify the values of p and q. For this, the autocorrelation and partial autocorrelation co-efficients of various orders of X_t are computed (Table 1). The ACF and PACF (Fig. 2 and 3) shows that the order of p and q can at most be 1.We entertained three tentative ARIMA models and chose that model which has minimum AIC (Akaike Information Criterion) and

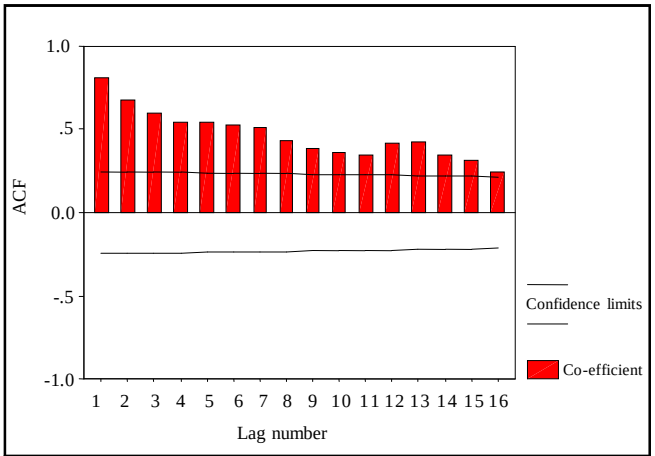

Fig. 2 : ACF of differenced data

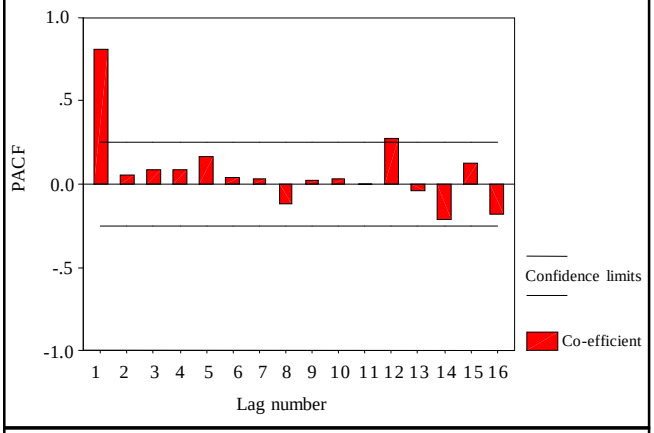

Fig. 3 : PACF of differenced tobacco data

BIC (Bayesian Information Criterion). The models and corresponding AIC and BIC values are :

ARIMA (p, d, q)	AIC	BIC
1 0 0	174.05	162.24
1 1 1	165.05	158.72
1 0 1	168.69	159.43

So the most suitable model is ARIMA (1,1,1) this model has the lowest AIC and BIC values.

Model parameters were estimated using SPSS package. Results of estimation are reported in Table 2. The model verification is concerned with checking the residuals of the model to see if they contain any

Table 1: Autocorrelations and partial autocorrelations

Lag	Autocorrelation	Std.error	Partial autocorrelation	Std.error
1	0.814	0.124	0.814	0.127
2	0.679	0.123	0.051	0.127
3	0.596	0.122	0.088	0.127
4	0.547	0.121	0.087	0.127
5	0.546	0.120	0.164	0.127
6	0.531	0.119	0.039	0.127
7	0.508	0.118	0.034	0.127
8	0.435	0.117	-0.121	0.127
9	0.387	0.116	0.025	0.127
10	0.366	0.114	0.033	0.127
11	0.346	0.113	-0.003	0.127
12	0.419	0.112	0.274	0.127
13	0.427	0.111	-0.042	0.127
14	0.343	0.110	-0.214	0.127
15	0.318	0.109	0.127	0.127
16	0.244	0.108	-0.183	0.127

Table 2 : Estimates of the fitted ARIMA model

		Estimates	Std. error	t	Approx sig.
Non- seasonal lag	AR1	0.99623	0.00707	140.9135	0.0000
	AR2	0.34106	0.12038	2.83312	0.0062
	MA1				
Constant		14820.85	12395.97	1.19562	0.23654
Number of residuals		65			
Number of parameters		2			
Residual df		63			
Adjusted residual sum of squares		190370719.3			
Residual sum of squares		1465609344.3			
Residual variance		2969019.2			
Model std. error		1723.0842			
Log-likelihood		-559.42			
Akaike's information criteria (AIC)		1124.84			
Schwarz's bayesian criterion (BIC)		431.1630			
		1131.27			

systematic pattern which still can be removed to improve on the chosen ARIMA. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations upto 16 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 3. As the results indicate, none of these correlations is significantly different from zero at a reasonable level. This proves that the selected ARIMA model is an appropriate model. The ACF and PACF of the residuals (Fig. 4 and 5) also indicate 'good fit' of the model.

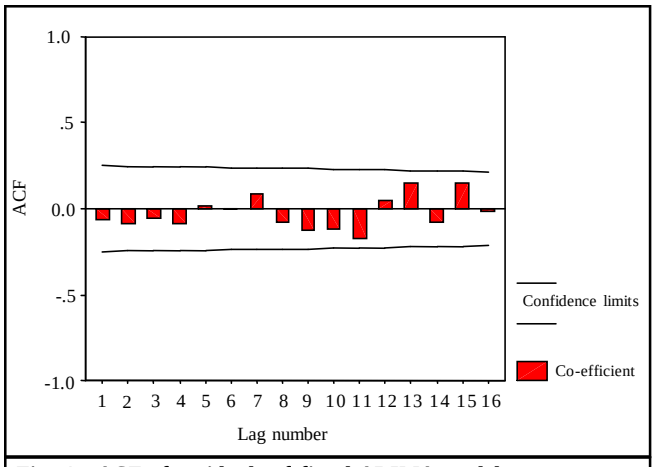

Fig. 4 : ACF of residuals of fitted ARIMA model

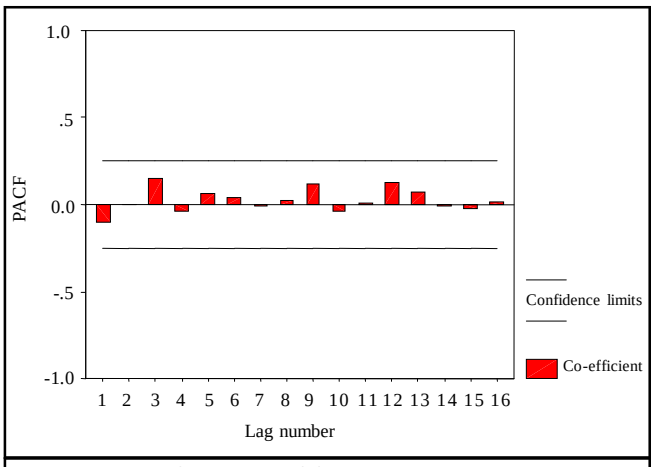

Fig. 5 : PACF of residuals of fitted ARIMA model

Table 3 : Autocorrelations and partial autocorrelations of residuals

Lag	Autocorrelation	Std.error	Box- ljung	Df	Sig.	Partial autocorrelation	Std.error
1	-0.100	0.123	0.658	1.000	0.417	-0.100	0.126
2	0.006	0.122	0.660	2.000	0.719	-0.004	0.126
3	0.150	0.121	2.192	3.000	0.534	0.152	0.126
4	-0.067	0.120	2.503	4.000	0.644	-0.038	0.126
5	0.074	0.119	2.893	5.000	0.716	0.064	0.126
6	0.047	0.118	3.050	6.000	0.803	0.040	0.126
7	-0.026	0.117	3.099	7.000	0.876	-0.004	0.126
8	0.050	0.116	3.282	8.000	0.915	0.024	0.126
9	0.113	0.115	4.254	9.000	0.894	0.119	0.126
10	-0.065	0.114	4.581	10.000	0.917	-0.043	0.126
11	0.037	0.113	4.687	11.000	0.945	0.008	0.126
12	0.140	0.112	6.270	12.000	0.902	0.124	0.126
13	0.021	0.110	6.307	13.000	0.934	0.071	0.126
14	0.015	0.109	6.325	14.000	0.958	-0.010	0.126
15	0.009	0.108	6.332	15.000	0.974	-0.024	0.126
16	0.015	0.107	6.351	16.000	0.984	0.019	0.126

Table 4 : Forecasts for tobacco production (2015-16 to 2019-2020) **(t/hec)**

Years	Forecasted production	Lower limit	Upper limit
2015-2016	28.5	23.3	33.7
2016-2017	28.4	22.8	34.1
2017-2018	28.4	22.3	34.5
2018-2019	28.3	21.8	34.8
2019-2020	28.3	21.4	35.2

The last stage in the modeling process is 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. The residuals calculated during the estimation process, are considered as the one step ahead forecast errors. The forecasts are obtained for the subsequent agriculture years from 2015-16 to 2019-2020.

In our study, the suitable model for tobacco production was found to be ARIMA (1,0,1). The forecasts of tobacco production, lower control limits (LCL) and upper control limits (UCL) are presented in Table 4. The validity of the forecasted values can be checked when the data for the lead periods become available. The model can be used by researchers for forecasting of tobacco production in India. However, it should be updated from time to time with incorporation of current data.

This paper forecast future tobacco production based on the data from 1950-51 to 2014-15, using ARIMA model. The forecast will help policy makers to design future tobacco production strategies.

Authors' affiliations:
V. M. BODADE, Department of Agricultural Economics and Statistics, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, AKOLA (M.S.) INDIA

LITERATURE CITED :

Bartlett, M.S. (1964). On the theoretical specification of sampling properties of autocorrelated time series. *J. Roy. Stat. Soc.*, **B 8** : 27–41.

Box, G.E.P. and Jenkins, G. M. (1970). *Time series analysis: forecasting and control*, Holden Day, San Francisco, CA.

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control.* Rev. Ed. San Francisco. Holden-Day.

Brockwell, P.J. and Davis, R. A. (1996). *Introduction to time series and forecasting*, Springer.

Brown, R.G. (1959). *Statistical forecasting for inventory control*. McGraw-Hill, NEW YORK, U.S.A.

Holt, C.C., Modigliani, F. , Muth, J.F. and Simon, H.A. (1960). Planning, production, inventores and work force. *Prentice Hall, Englewood Cliffs,* NJ, U.S.A.

Iqbal, N., Bakhsh, K., Maqbool, A. and Ahmad, A.S. (2005). Use of the ARIMA Model for forecasting wheat area and production in Pakistan. *J. Agric. & Soc. Sci.*, **2** : 120-122.

Jenkins, G. M. and Watts, D.G. (1968). *Spectral analysis and its application*, *day,* San Francisco, California, USA.

Kendall, M. G. and Stuart, A. (1966). The advanced theory of statistics. Vol. **3**. Design and Analysis and Time-Series. Charles Griffin & Co. Ltd.,LONDON, UNITED KINGDOM.

Ljunge, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65** : 67–72.

Makridakis, S., Anderson, A., Filds, R., Hibon, M., lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *J. Forecasting Competition. J. Forecasting*, **1**: 111–153.

Meese, R. and Geweke, J. (1982). A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. Manuscript, University of California, Berkeley, CA, U.S.A.

Muhammad, F., Javed, M. S. and Bashir, M. (1992). Forecasting sugarcane production in Pakistan using ARIMA Models, *Pak. J. Agric. Sci.*, **9**(1): 31-36.

Prindycke, R.S. and Rubinfeld, D.L. (1981). *Econometric models and economic forecasts*, 2nd Ed. McGraw-Hill, NEW YORK, U.S.A.

Quenouille, M.H. (1949). Approximate tests of correlation in time- series. *J. Roy. Stat. Soc.*, **B11**: 68–84.

Saeed, N., Saeed, A., Zakria, M. and Bajwa, T. M. (2000). Forecasting of wheat production in Pakistan using ARIMA models, *Internat. J. Agric. & Biol.*, **4**: 352-353.

Yule, G.U. (1926). Why do we sometimes get nonsence-corrleations between times deries. A study in sampling and the nature of series. *J. Roy. Stat. Soc*., **89**: 1–69.

Yule, G.U. (1927). On a method of investigation periodicities in disturbed series, with specia; Reference To Wolfer's Sunspot Number. *Phill. Trans*., **A 226** : 267–98.