AU
DOI: 10.15740/HAS/AU/12.2/252-257

Agriculture Update

Volume **12** | Issue 2 | May, 2017 | 252-257

Visit us : *www.researchjournal.co.in*

e ISSN-0976-6847

RESEARCH ARTICLE :

Modelling and forecasting of tur production in India using ARIMA model

PREMA BORKAR AND V.M. BODADE

ARTICLE CHRONICLE :
Received : *07.03.2017;*
Revised : *22.03.2017;*
Accepted : *04.04.2017*

SUMMARY : The paper describes an empirical study of modeling and forecasting time series data of tur production in India. Yearly tur 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 2024-2025 were calculated based on the selected model. The forecasting power of autoregressive integrated moving average model was used to forecast tur production for ten leading years. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of tur seed, import and/or export and adopt appropriate measures in this regard.

How to cite this article : Borkar, Prema and Bodade, V.M. (2017). Modelling and forecasting of tur production in India using ARIMA model. *Agric. Update*, **12**(2): 252-257; **DOI : 10.15740/HAS/AU/12.2/252-257.**

KEY WORDS :
ACF - autocorrelation function, ARIMA - autoregressive integrated moving average, PACF - partial autocorrelation function, Tur

Author for correspondence :

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

See end of the article for authors' affiliations

BACKGROUND AND OBJECTIVES

Tur is an important pulse crop in India. It is also known as pigeonpea, red gram and Arhar. It is mainly cultivated and consumed in developing countries of the world. This crop is widely grown in India. India is the largest producer and consumer of tur in the world. Tur is the second largest pulse crop in India accounting for about 20 per cent of total pulse production. The production of tur during 2014-15 as per fourth advance estimates in the country was about 2.78 million tonnes. The area under tur in current year 2014-15 is estimated about 39.05 lakh ha which is about 6.4 per cent less than the previous year.

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 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- 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 tur crop in India. This would enable to predict expected tur production for the years from 2013 onward. Such an exercise would enable the policy makers to foresee ahead of time the future requirements for tur storage, import and/or export of tur 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.

RESOURCES AND METHODS

Respective time series data for this study were collected from various Government Publications of India. Box and Jenkins (1976) linear time series model was applied. Auto Regressive Integrated Moving Average (ARIMA) is the most general class of models 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).

$$a^d y_t = ¢ + ¯_1 a^d y_{t-1} + ¯_2 a^d y_{t-2} + \text{-----} + ¯_p y_{t-p} + e_t - \sim e_{t-1} - \sim_2 e_{t-2} \sim_q e_{t-2} \quad (1)$$

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

$$Y = ¢ + ¯_1 y_{t-1} + ¢_2 y_{t-2} + \text{--------} + ¯_p y_{t-p} + e_t \quad (2)$$

where,

e_t is forecast error, assumed to be independently distributed across time with mean θ and variance $\theta_2 e$,

e_{t-1}, e_{t-2} ——— e_{t-q} are past forecast errors,

α_1, ————— α_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

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

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 (Prindycke 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 2013 upto 2025 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.

OBSERVATIONS AND ANALYSIS

The results obtained from the present study as well as discussions have been summarized under following heads:

Building ARIMA model for tur production data in India :

To fit an ARIMA model requires a sufficiently large data set. In this study, we used the data for tur production for the period 1950-1951 to 2014-15. 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 tur production. The time plot of the tur production data is presented in Fig. 1.

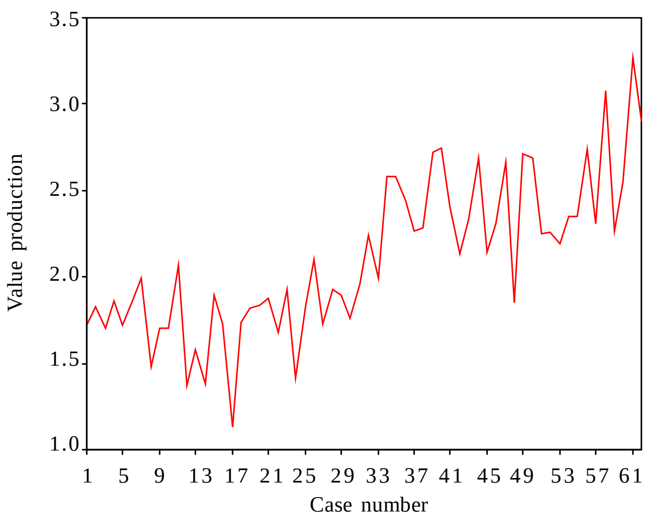

Fig. 1 : Time plot of tur production data

The above time plot indicated that the given series is nonstationary. 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 Xt 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 BIC (Bayesian Information Criterion). The models and corresponding AIC and BIC values (Table A).

Table A : AIC and BIC value

ARIMA (p, d, q)	AIC	BIC
1 0 0	47.03	51.28
1 1 0	38.77	42.99
1 1 1	25.02	31.35

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 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

Table 1 : Autocorrelations and partial autocorrelations

Lag	Autocorrelation	Std. error	Lag	Partial autocorrelation	Std. error
1	0.619	0.124	1	0.619	0.127
2	0.566	0.123	2	0.297	0.127
3	0.608	0.122	3	0.316	0.127
4	0.566	0.121	4	0.144	0.127
5	0.551	0.120	5	0.123	0.127
6	0.489	0.119	6	-0.023	0.127
7	0.441	0.118	7	-0.052	0.127
8	0.390	0.117	8	-0.098	0.127
9	0.399	0.116	9	0.032	0.127
10	0.324	0.114	10	-0.083	0.127
11	0.375	0.113	11	0.159	0.127
12	0.347	0.112	12	0.047	0.127
13	0.230	0.111	13	-0.127	0.127
14	0.237	0.110	14	-0.053	0.127
15	0.240	0.109	15	-0.003	0.127
16	0.166	0.108	16	-0.113	0.127

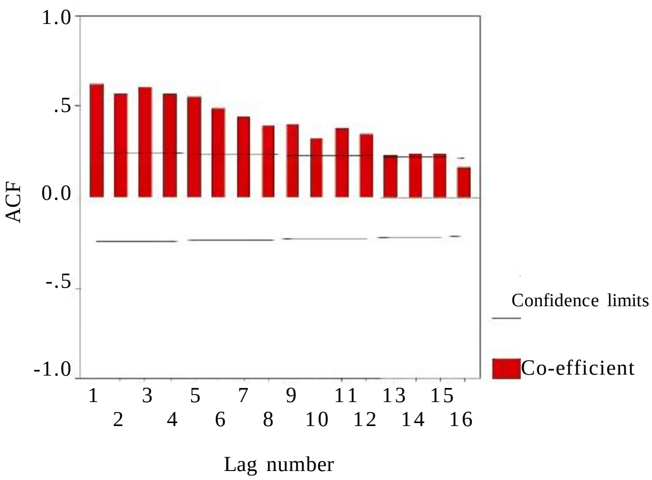

Fig. 2 : ACF of differenced data

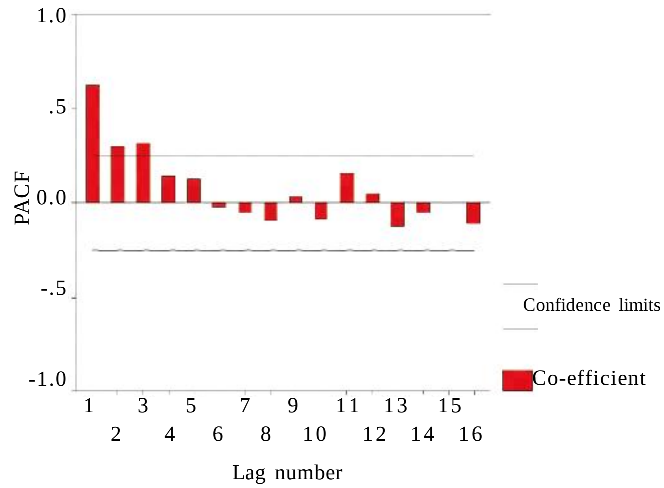

Fig. 3 : PACF of differenced data

Table 2 : Estimates of the fitted ARIMA model

		Estimates	Std Error	t	Approx sig
Non- Seasonal lag	AR1	-0.139022	0.170830	-0.8138	0.41908686
	MA1	0.715422	0.130183	5.49549	0.00000091
Constant		0.018101	0.009625	1.88055	0.06505709
Number of residuals		61			
Number of parameters		2			
Residual df		58			
Adjusted residual sum of squares		4.8710246			
Residual sum of squares		5.1143186			
Residual variance		0.08271759			
Model Std. error		0.28760667			
Log-Likelihood		-9.4831015			
Akaike's information criteria (AIC)		24.96			
Schwarz's bayesian criterion (BIC)		31.29			

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.

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 year from 2014-15 to 2024-2025.

Conclusion :

In our study, the developed model for tur production was found to be ARIMA (1,1,1). The forecasts of tur 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 tur production in India. However, it should be updated from time to time with incorporation of current data.

This paper discloses the production of tur from 1950-1951 to 2014-2015 and also shows the future movement.

Table 3 : Autocorrelations and partial autocorrelations of residuals

Lag	Autocorrelation	Std. error	Box- Ljung	df	Sig.	Lag	Partial autocorrelation	Std. error
1	-0.032	0.125	0.067	1.000	0.796	1	-0.032	0.128
2	-0.120	0.124	1.000	2.000	0.607	2	-0.121	0.128
3	0.108	0.123	1.766	3.000	0.622	3	0.101	0.128
4	0.038	0.122	1.861	4.000	0.761	4	0.030	0.128
5	0.168	0.121	3.794	5.000	0.579	5	0.199	0.128
6	0.026	0.120	3.841	6.000	0.698	6	0.036	0.128
7	-0.036	0.119	3.933	7.000	0.787	7	0.004	0.128
8	-0.086	0.117	4.469	8.000	0.813	8	-0.131	0.128
9	0.006	0.116	4.472	9.000	0.878	9	-0.031	0.128
10	-0.155	0.115	6.287	10.000	0.791	10	-0.236	0.128
11	0.021	0.114	6.319	11.000	0.851	11	0.022	0.128
12	0.023	0.113	6.362	12.000	0.897	12	-0.011	0.128
13	-0.178	0.112	8.895	13.000	0.781	13	-0.081	0.128
14	0.021	0.111	8.932	14.000	0.835	14	0.043	0.128
15	0.015	0.109	8.950	15.000	0.880	15	0.060	0.128
16	-0.088	0.108	9.607	16.000	0.886	16	-0.063	0.128

Table 4 : Forecasts for tur production (2015-16 to 2024-2025) **(Million Tonnes)**

Years	Forecasted production	Lower limit	Upper limit
2015-2016	2.89461	2.26349	3.52574
2016-2017	2.91268	2.26125	3.56411
2017-2018	2.93079	2.25916	3.60242
2018-2019	2.94889	2.25711	3.64067
2019-2020	2.96699	2.25511	3.67887
2020-2021	2.98509	2.25316	3.71703
2021-2022	3.00319	2.25124	3.75514
2022-2023	3.02130	2.24937	3.79322
2023-2024	3.03940	2.24753	3.83127
2024-2025	3.05750	2.24572	3.86928

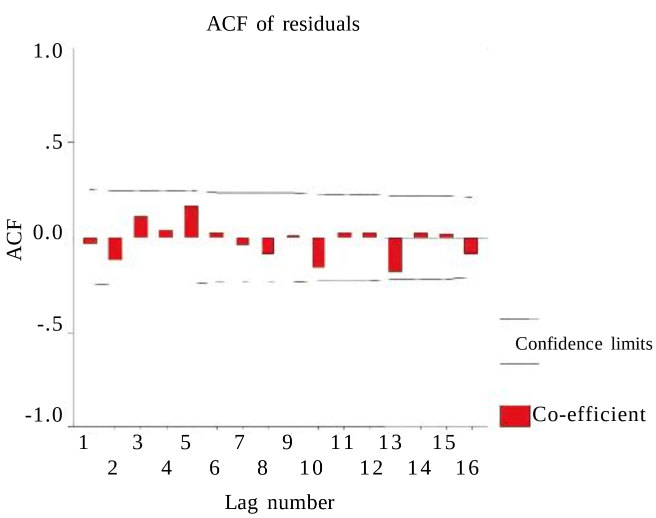

Fig. 4 : ACF of residuals of fitted ARIMA model

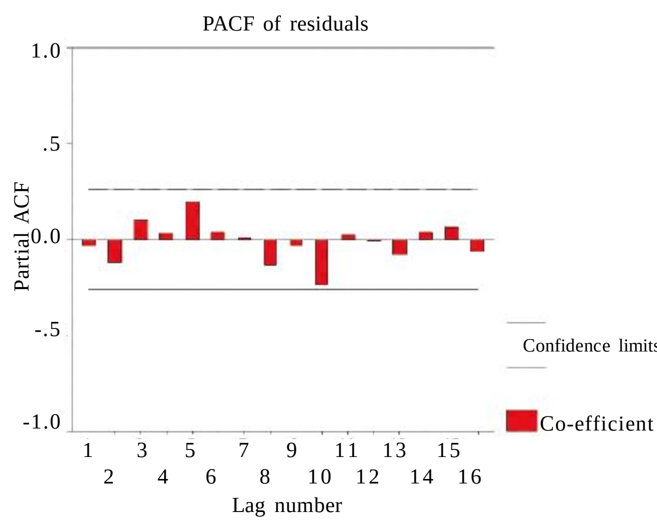

Fig. 5 : PACF of residuals of fitted ARIMA model

To formulate future development plan for tur production, it is essential to know the previous condition and also see the future trend. In this study, forecasting is done by using some sophisticated statistical tools so that the government and policy makers can easily realize about the future development of tur production and could take initiatives to improve the production.

Authors' affiliations :

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

REFERENCES

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*. New York, McGraw-Hill.

Holt, C.C., Modigliani, E., Muth, J.F. and Simon, H.A. (1960). *Planning, production, inventores and work force*. Prentice Hall, Englewood Cliffs, New Jersey, USA.

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.

Judge, G.G., Hill, R. Carter, William, E.G. and Helmut, I. (1988). *Introduction to the theory and practice of econometrics*. 2nd Ed., John Wiley and Son, INC. New York, Toronto, Singapore

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**: 111–53.

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

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. New York, McGraw-Hill.

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 series. 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.