MODELING OF GROUNDNUT PRODUCTION IN INDIA USING ARIMA MODEL

Prema Borkar

Assistant Professor,

Gokhale Institute of Politics and Economics (Deemed to be a University),

Pune - 411004. Maharashtra.

ABSTRACT

The paper describes an empirical study of modeling and forecasting time series data of groundnut production in India. Yearly groundnut production data for the period of 1950-1951 to 2013-2014 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 (0, 1, 1) is appropriate. The forecasts from 2015-2016 to 2024-2025 are calculated based on the selected model. The forecasting power of autoregressive integrated moving average model was used to forecast groundnut production for ten leading years. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of groundnut seed, import and/or export and adopt appropriate measures in this regard.

Key words: ACF - autocorrelation function, ARIMA - autoregressive integrated moving average, PACF - partial autocorrelation function, Groundnut.

India is one of the largest producers of oilseeds in the world and occupies an important position in the Indian agricultural economy. It is estimated that nine oilseeds namely groundnut, rapeseed-mustard, soybean, sunflower, safflower, sesame, Niger, castor and linseed, accounted for an area of 23.44 million hectares with the production of 25.14 million tones. Groundnut is called as the 'King' of oilseeds. It is one of the most important food and cash crops of our country. While being a valuable source of all the nutrients, it is a low-priced commodity. Groundnut is also called as wonder nut and poor men's cashew nut. Groundnut is one of the most important cash crops of our country. It is a low-priced commodity but a valuable source of all the nutrients. Groundnut is grown on 26.4 million ha worldwide with a total production of 37.1 million metric t and an average productivity of 1.4 metric t/ha. Over 100 countries worldwide grow groundnut. Developing countries constitute 97 per cent of the global area and 94 per cent of the global production of this crop. The production of groundnut is concentrated in Asia and Africa (56 per cent and 40 per cent of the global area and 68 per cent and 25 per cent of the global production, respectively).

The major groundnut producing countries in the world are India, China, Nigeria, Senegal, Sudan, Burma and the United States of America. Out of the total area of 18.9 million hectares and the total production of 17.8 million tonnes in the world, these countries account for about 69 percent of the area and 70 percent of the production. India occupies the first place, both in regard to the area and the production in the world. About 7.5 million hectares are put under it annually and the production is about six million tonnes. Seventy percent of the area and seventy five percent of the production has been concentrated in the four states of Gujarat, Andhra Pradesh, Tamil Nadu and Karnataka. Andhra Pradesh, Karnataka, Tamil Nadu and Orissa have irrigated areas primarily during the rabi season. The irrigated areas form about six percent of the groundnut area in India. In these states groundnut production is mainly depends on rainfall.

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 groundnut crop in India. This would enable to predict expected groundnut production for the years from 2015 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 groundnut 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.

Material and Methods

The time series data of groundnut production in India has been collected from the website of Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture from 1950-51 to 2013-14. ARIMA stochastic modeling is used on the groundnut production of India for forecasting purpose.

Development of ARIMA models

The development of ARIMA models was based on the methodology described in the classic work of Box and Jenkins. Univariate ARIMA models use only the information contained in the series itself. Thus, models are constructed as linear functions of past values of the series and/or previous random shocks (or errors). Forecasts were generated under the assumption that the past history could be translated into predictions for the future. Forecasts were generated under the assumption that the past values of the series and / or previous random shocks (or errors). Forecasts were generated under the assumption that the past history could be translated into predictions for the future. ARIMA modeling was developed following the standard three steps procedures. (i) Identification of the model; (ii) Parameter estimation and (iii) Diagnostic and verification of the model. The identification step determines (i) whether the process is sta- tionary and the possible transformations to obtain stationarity and (ii) whether the form of the process is autoregressive (AR), moving average (MA) or both (ARMA), and its orders. Three parameters are used in summarizing an ARIMA model are the AR parameter p, integration parameter d and MA parameter q. Parameters p and q denote the order of AR and MA, while d denotes the degree of differencing the series to obtain stationarity. The autocorrelation function (ACF) and partial autocorrelation functions (PACF) of a series together are the most powerful too, usually applied to reveal the correct values of the parameters. The ACF gives the autocorrelations calculated at lags 1, 2 and so on, while PACF gives the corresponding partial autocorrelations, controlling the autocorrelations at intervening lags. Parameter estimation of tentative models was determined using maximum-likelihood methods. The final results included the parameter estimates, standard errors, estimates of residual variance, standard error of the estimate, natural log likelihood, Akaike's Information Criterion (AIC), and Schwartz's Bayesian criterion (SCB) or Bayesian Information criterion (BIC). Model selection was based on the minimization of AIC and BIC. These criteria are descriptors of the model's parsimony as they simultaneously account for the model's fit onto the observed series alongside number of parameters used in the fit. The ability to forecast using ARIMA models was tested by applying the ARIMA methodology to available data (1950-51 to 2013-14).

The autoregressive moving average (ARMA) model, denoted by ARMA (p,q), is given by

$$y_t = \phi_1 y_{t-1} + \phi_1 y_{t-2} + \ldots\ldots + \phi_p y_{t-p} + \varepsilon t - \theta_1\ \varepsilon_{t-1} - \theta_{2\ t\text{-}}\varepsilon_2 - \ldots\ldots\ldots - \theta_{q\ t}\varepsilon_{\text{-}q} \qquad (1)$$

or equivalently by

$$\varphi\ (B)yt = \theta\ (B)\ \varepsilon_t \qquad (2)$$

where $\varphi\ (B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots\ldots - \varphi_p B^p$ and

$\theta\ (B) = 1 - \theta_1 B - \theta_2 B^{\ q}$

B is the backshift operator defined by $By_t = y_{\ t\text{-}p}$

A generalization of ARMA models which incorporates a wide class of nonstationary time-series is obtained by introducing 'differencing' into the model. The simplest example of a nonstationary process which reduces to a stationary one after differencing is 'Random Walk'. A process {yt} is said to follow autoregressive integrated moving average (ARIMA), denoted by ARIMA (p,d,q), if $\blacktriangledown^d\ y_t = (1 - B)^d\ \varepsilon_t$ is ARMA(p,q). The model is written as

$$\varphi\ (B)(1 - B)^d\ y_t = \theta\ (B)\ \varepsilon_t \qquad (3)$$

where ε_t are identically and independently distributed as N $(0, \sigma^2)$. The integration parameter d is a nonnegative integer. When d = 0, the ARIMA (p,d,q) model reduces to ARMA (p,q) model.

Estimation of Parameter

Estimation of parameters for ARIMA model is generally done through Nonlinear least squares method. Several software packages are available for fitting of ARIMA models. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) values are used for parameter estimation.

Results and Discussion

Building ARIMA model for Groundnut Production data in India

To fit an ARIMA model requires a sufficiently large data set. In this study, we used the data for groundnut production for the period 1950-1951 to 2013-14. As we have earlier stated that development of ARIMA model for any variable involves three steps: identification, estimation, diagnostic and forecasting. Each of these three steps is now explained for groundnut production. The time plot of the groundnut production data is presented in Figure 1.

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 coefficients of various orders of Xt are computed (Table 1). Using the expert modeler we found the best ARIMA model for the dataset, it is ARIMA (0,1,1) consisting of non seasonal part. The autocorrelation and partial autocorrelation functions are used, as basic instruments, to identify stationary of time series. Figure 2 and 3 show the ACF and PACF functions of the logarithmic transformed production data.

The estimate of the parameters with corresponding standard error for ARIMA (0,1,1) is given 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 up to 16 lags were computed and the same along with their significance which is tested by Box-Ljung test. 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.

Table 3 shows a goodness of the fit statistics for the given data set. R-squared represents an estimate of the proportion of the total variation in the series that is explained by the model. Largest values (upto a maximum value of 1) indicate better fit. A value of 0.412 means that the model does an excellent job of explaining the observed variations in the series.

Mean percentage error (MAPE) for the model is 16.025 per cent. A measure of how much a dependent series varies from its model-predicted level. Root Mean Square Error (RMSE), i.e., the square root of mean square error is a measure of how much a dependent series varies from its model- predicted level, expressed in the same units as the dependent series. Maximum Absolute Percentage Error (MaxAPE) represents the largest forecasted error, expressed as a percentage. The result in table 3 shows a good performance of the ARIMA model (0, 1, 1). The Ljung-Box statistics indicate that the model is specified correctly, i.e., a significance value is greater than 0.05.

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 groundnut production was found to be ARIMA (0,1,1). The forecasts of groundnut 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. ARIMA model being stochastic in nature, it could be successfully used for modeling as well as forecasting the groundnut production of India. The model demonstrated a good performance in terms of explaining variability and predicting power. The forecasting of groundnut production can help to both farmers as well as the planner for future planning.

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 G. M. Jenkins. (1970). Time series analysis: forecasting and Control, Holden Day, San Francisco, CA.

Box, G.E.P. and G.M. Jenkins, 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,

Iqbal, N., 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.

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

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

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

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

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

Makridakis, S, 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.

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

Prindycke, R.S. and D.L. Rubinfeld, 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., 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.

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.

Graph of Groundnut Production data

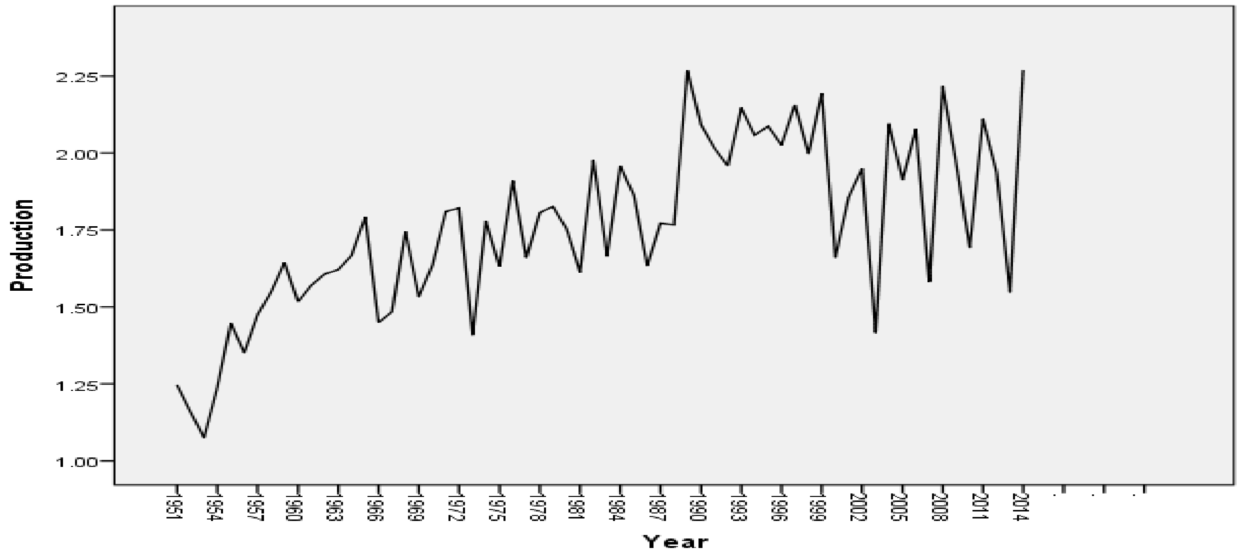

Figure 1: Time plot of groundnut production data

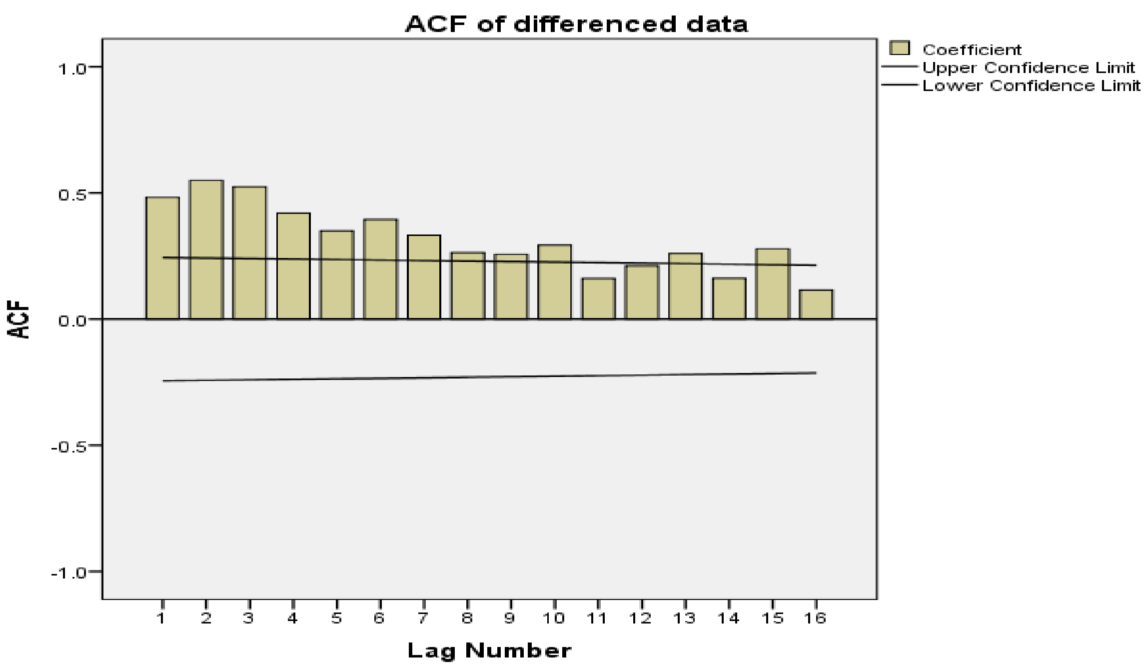

Figure 2: ACF of differenced data

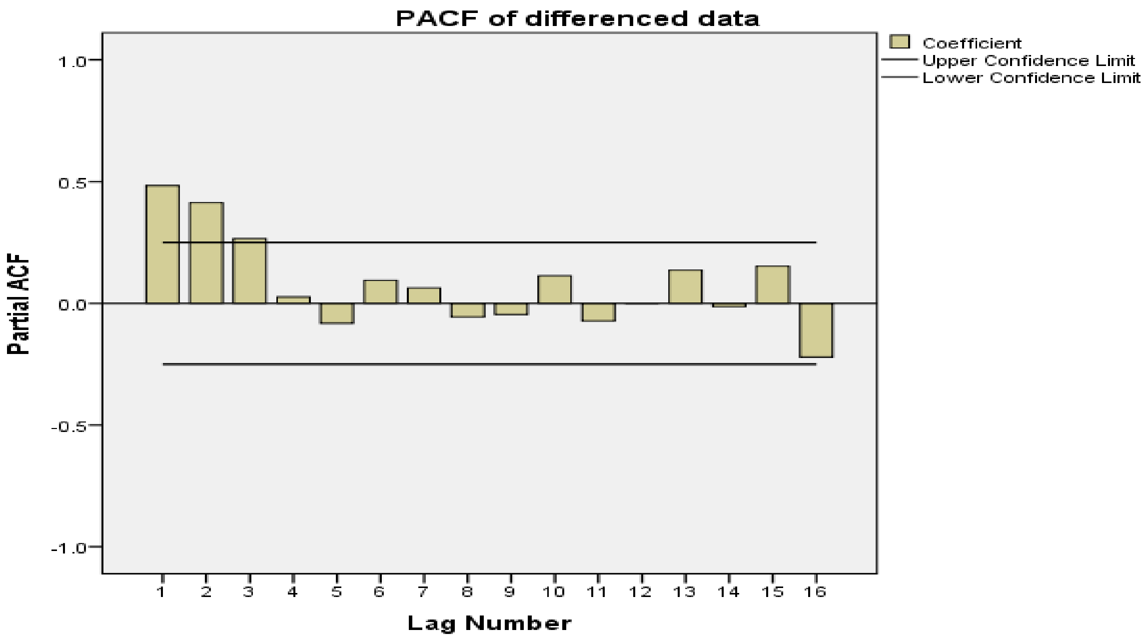

Figure 3: PACF of differenced data

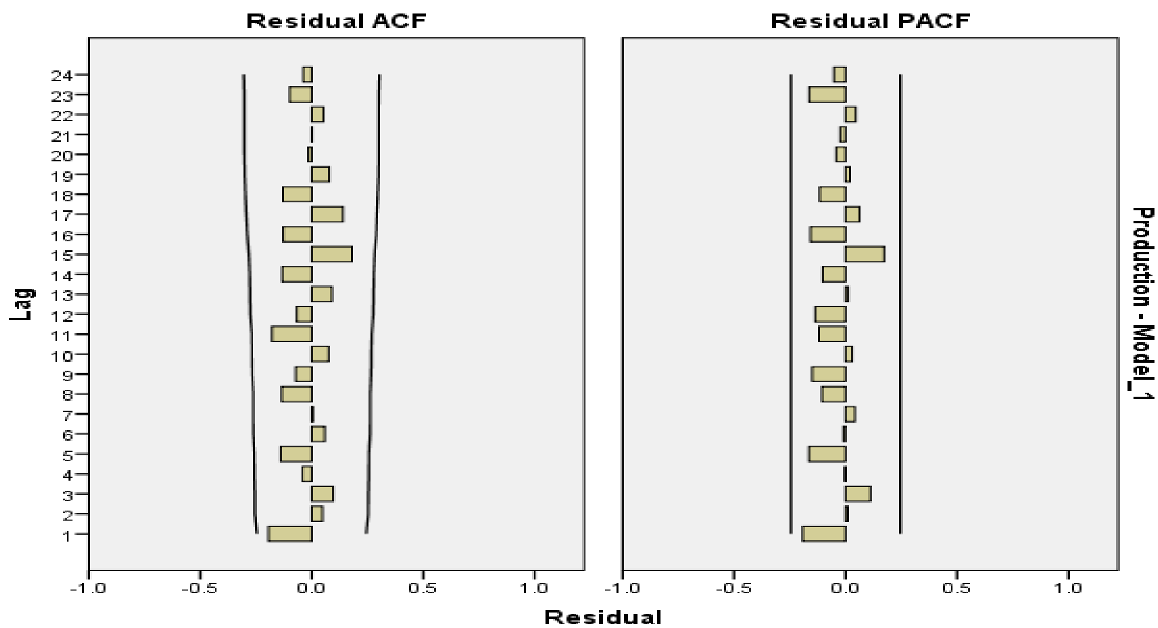

Figure 4: ACF and PACF of residuals of fitted ARIMA model

Table 1: Autocorrelations and partial autocorrelations

Lag	Autocorrelation	Std.error	Lag	Partial Autocorrelation	Std.error
1	0.483	0.122	1	0.483	0.125
2	0.550	0.121	2	0.413	0.125
3	0.524	0.120	3	0.265	0.125
4	0.420	0.119	4	0.026	0.125
5	0.350	0.118	5	-0.082	0.125
6	0.394	0.117	6	0.094	0.125
7	0.332	0.116	7	0.063	0.125
8	0.264	0.115	8	-0.056	0.125
9	0.257	0.114	9	-0.046	0.125
10	0.293	0.113	10	0.113	0.125
11	0.161	0.112	11	0.073	0.125
12	0.211	0.111	12	-0.000	0.125
13	0.261	0.110	13	0.136	0.125
14	0.162	0.109	14	-0.014	0.125
15	0.279	0.108	15	0.152	0.125
16	0.116	0.107	16	-2.21	0.125

Table 2: Parameter estimate of the fitted ARIMA model

Parameters	Estimate	SE
Constant	0.011	0.005
MA1	0.801	0.082

Table 3: Model fit of the fitted ARIMA model

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	0.467		0.467	0.467	0.467	0.467	0.467	0.467	0.467	0.467	0.467
R-squared	0.412		0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412
RMSE	1.262		1.262	1.262	1.262	1.262	1.262	1.262	1.262	1.262	1.262
MAPE	16.025		16.025	16.025	16.025	16.025	16.025	16.025	16.025	16.025	16.025
MaxAPE	85.754		85.754	85.754	85.754	85.754	85.754	85.754	85.754	85.754	85.754
MAE	0.923		0.923	0.923	0.923	0.923	0.923	0.923	0.923	0.923	0.923
MaxAE	3.533		3.533	3.533	3.533	3.533	3.533	3.533	3.533	3.533	3.533
Normalized BIC	0.597		0.597	0.597	0.597	0.597	0.597	0.597	0.597	0.597	0.597

Table 4: Forecasts for Tur Production (2015-16 to 2024-2025)
(Million Tonnes)

Years	Forecasted Production	Lower limit	Upper limit
2015-2016	7.07	4.58	10.45
2016-2017	7.04	4.49	10.53
2017-2018	7.01	4.42	10.61
2018-2019	6.99	4.35	10.68
2019-2020	6.97	4.29	10.75
2020-2021	6.95	4.22	10.81
2021-2022	6.92	4.16	10.87
2022-2023	6.90	4.11	10.92
2023-2024	6.88	4.06	10.97
2024-2025	6.86	4.01	11.00