TEA PRODUCTION IN ASSAM: FORECASTING WITH ARIMA MODEL

Sakuntala Deka[1,*], P.J. Hazarika[2] and A. N. Patowary[3]

[1,2] Department of Statistics, Dibrugarh University, Dibrugarh–786004, Assam, India.
[3]College of Fisheries, Assam Agricultural University, Raha–782103, Assam, India.
Corresponding Author: sakuntaladeka456@gmail.com

Abstract

This study outlines the development of a conventional time series model viz. autoregressive integrated moving average (ARIMA) model for the annual tea production of Assam, the largest tea producing state in India. The model is used to forecast the tea production covering the period 1957 to 2016 likely to occur in future**.** The analyzed data are secondary and obtained from the Tea Board of India, Ministry of Commerce from 1957 to 2016 with 60-year data. A significant upward trend is observed in the yearly tea production in Assam from 1957-2016. We investigated and found that ARIMA (0, 2, 1) model is suitable for the given data set. The forecasted tea production from the model in Assam for the upcoming ten years also exhibits an upward trend.

Key Words**:** ARIMA, forecasting, tea production

1. INTRODUCTION

Tea is a perennial commercial crop cultivated widely in the plains of North-East India, which is one of the most important tea growing regions globally. India is the second-largest tea producing country after China. The popularity of tea as a beverage in the world is well known. Its popularity as a beverage encompasses all age groups, geographical locations and social levels. Tea is considered the most important crop in Assam. Assam produces around 51% of the tea produced in

India, and it contributes 17% of the tea produced in the world. Though Assam produces mainly black tea and orthodox tea, it also produces small quantities of green tea. There are more than 850 big tea estates and more than 25000 small tea growers in Assam. The tea plantation area in Assam is about 3.22 lakh hectares (Tea board of India, 2014). The tea industry in Assam is about 180 years old and plays a vital role in the state economy and the national economy (Arya,2013).. The world's largest CTC tea auction centre is in Guwahati, Assam, and it is also the world's second-largest auction centre. In India, the discovery of tea is attributed to Robert Bruce. Singphu chief Bessa Gam introduced it to him, and tea was called *phalap* by them. In 1840, the first tea garden was established in Chabua in the Dibrugarh district of Assam. By the beginning of the 20th century, tea spread through the length & breadth of the globe. Assam Tea Company began first commercial tea production in this region. By the 19th century, Assam became the leading tea-producing region globally. Gijo, E.V.(2013) has applied the Box-Jenkins seasonal autoregressive integrated moving average (ARIMA) model for a tea packaging company in India and shown that his model has helped the organisation to plan the production activities more efficiently. Verma *et al.* (2016) developed an ARIMA model to forecast coriander prices for a mandi in Rajasthan. The research work concludes that such a study will enable the farmers to make proper cropping choices. Weerapura and Abeynayake (2013) used the ARIMA model to forecast tea production in Sri Lanka during 2011 and compared the results with Exponential Smoothing techniques.

2. MATERIALS AND METHODS

2.1 Data Source

The data considered in this study are completely secondary in nature. Yearly tea production data of Assam were collected for the period of 1957 to 2016 from Tea Board of India (ITB), Ministry of Commerce.

2.2 ARIMA Model

The Box and Jenkins (1970) methodology called Autoregressive Integrated Moving Average (ARIMA) were used. It is an advanced technique of forecasting which requires long term time series data. This model decomposes historical data into an Autoregressive (AR) process, where there is a memory of past values, an Integrated (I) process accounted for making the data stationary, and a Moving-Average (MA) process, which accounts for previous error terms making it easier to forecast. The steps involved in Box and Jenkins (1970) methodology are given below,

Phase 1

(a) Data preparation: Transform data to stabilize variance and difference data to obtain stationary series

(b) Model Selection: Examine autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify potential models.

Phase 2

(a) Estimation: Estimate parameters in potential models. Select the best model using suitable criteria.

(b) Diagnostic: Check ACF and PACF of residuals. Examine residuals follow white noise or not. Select another model by model selection criterion if it does not follow white noise.

Phase 3

(a) If residuals follow white noise, use the model for forecasting.

Box and Jenkins (1970) autoregressive integrated moving average (ARIMA) model is given by

$$\phi(B)\nabla^d x_t = \mu + \theta(B)e_t, \qquad (2.1)$$

Where, e_t is the usual white noise process. The general model is denoted by ARIMA(p, d, q). The ordinary autoregressive and moving average components are represented by the following polynomials and of orders p and q, respectively,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p \qquad (2.2)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q \qquad (2.3)$$

and the different components are represented by $\nabla^d = (1-B)^d$.

3. RESULTS AND DISCUSSION

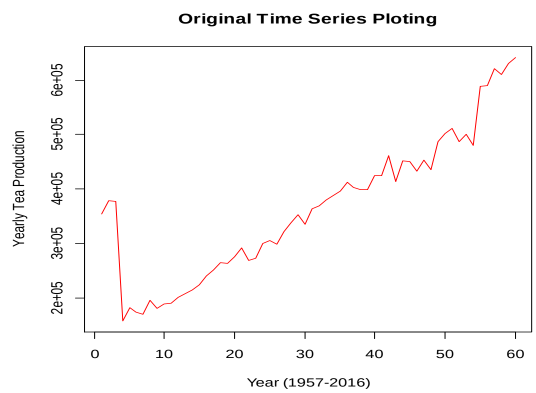

Figure 1: Original Time Series Ploting

From Figure 1, it is clear that the yearly tea production of Assam has shown an increasing trend with some fluctuation over the period from 1957- 2016 i.e. the data series is not stationary. Therefore to make the data stationary and to stabilize the variance the researcher decides to take the differences of data.

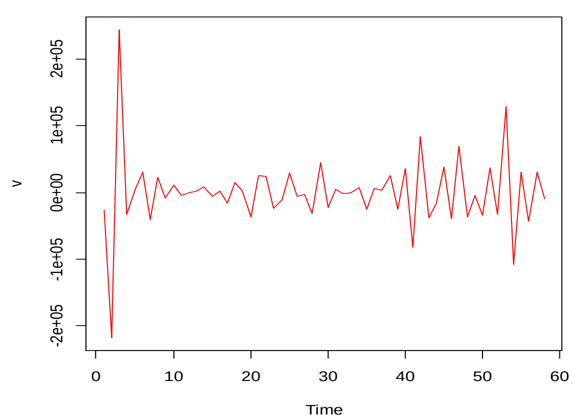

Figure 2: Second Difference of the Original Data

After taking difference, it is clear from the above figure (Figure 2) that the second differenced tea production data of Assam shows stable variance which leads the data becomes stationary. However, for better result Augmented-Dickey-Fuller unit root test is used to check whether the data series is stationary or not.

A clear non stationary behavior of the second differenced data has been observed in Figure 2. Similarly, Augmented Dickey-Fuller test tells the same truth (p-value<0.05). Further, ACF and PACF of the second differenced of the original data are plotted in Figure 3.

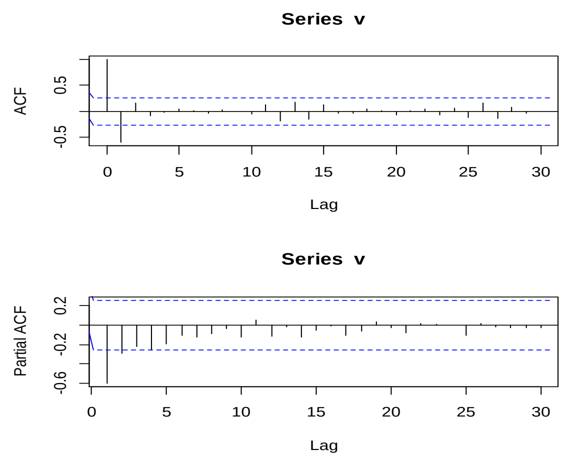

Figure 3: ACF and PACF of the second difference of the original data

In Figure 3, it is observed that there is exponential decay in the first few lags of PACF and one prominent significant spike at lag 1 in the ACF indicating ARIMA (0, 2, 1) model. Also, auto.arima() function in R-software tells the same truth. We have also compared AIC and BIC of the models for different combination of p, d, q and found that ARIMA (0,2,1) model has low AIC and BIC. Therefore, the researcher acquire a rough idea that ARIMA (0, 2, 1) model, may be tentatively appropriate to total yearly tea production data which is given by

$$(1-B)^2 Y_t = (1-\theta_1 B)e_t \tag{3.1}$$

For the ARIMA (0, 2, 1) model obtained above Eq. (3.1), the parameter of the model is estimated by maximum likelihood method obtained from R software is as follows:

Table 1: Parameter Estimation of ARIMA (0, 2, 1)

Parameter	Estimate	Std. Error	Z	p-value
$\hat{\theta}_1$	-0.9501	0.0591	-16.1200	0.00001

In Table 1, it is observed that the coefficient is highly significant. The constant term of the model is omitted due to differencing. Then the estimated model can be written as

$$(1-B)^2 Y_t = (1+0.9501\times B)e_t \quad (3.2)$$

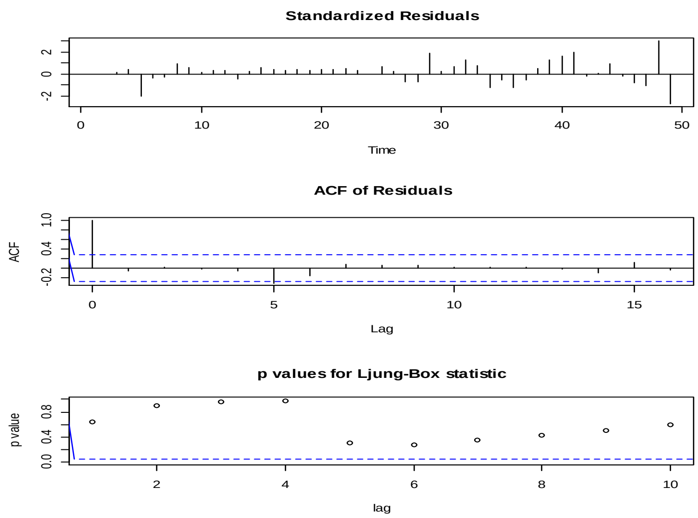

Figure 4: standardized residuals, the ACF of the residuals, and the p-values of the Q-statistic

Figure 4 displays a plot of the standardized residuals, the ACF of the residuals, and the p-values of the Q-statistic at lag 1 through 12. From the standardized plot of residuals, it is observed that residuals are lie within the limit of -3 and +3. Also, Figure 4 shows that none of the autocorrelations is individually statistically significant and nor the Ljung–Box–Pierce Q-statistics are statistically significant. Here, the Ljung-Box test statistic is 12.22, and the p-value is 0.2011, so the null hypothesis of independence of the residual series cannot be rejected. Using the white noise test, it is obtained that p value is 0.4711 which indicates that the residuals series is white noise (with mean 0 and variance σ^2).

To check the normality assumptions of the residuals of the fitted model, quintile-quintile (q-q) plot of the residuals is drawn which is given in Figure 5.

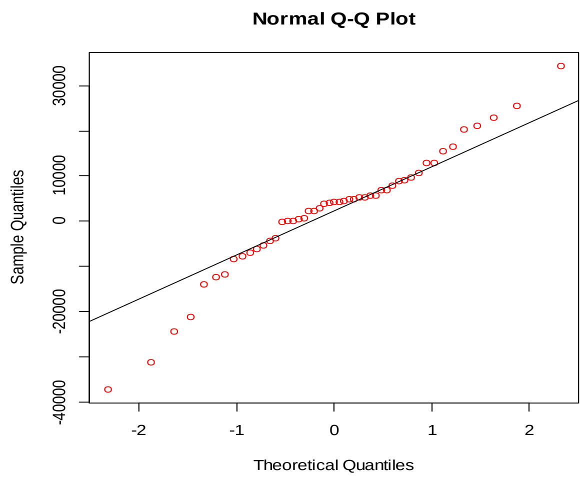

Figure 5: Q-Q plot of the residuals

Figure 5 shows that the residuals are more or less normally distributed. Moreover, by applying Komogorov-Smirnov (K-S) test to the residuals it is found that D=0.08 and p-value =0.6187 which indicates that the residuals follow normal distribution well. Now, to check the validity of the fitted model, the actual observations are plotted with predicted values from 2007-2016 (10 years) in Figure 6. It is observed from the figure that the yearly predicted tea production of Assam for the period 2007-2016 is almost equal and exact pattern with the actual data. Therefore, it indicates that the fitted model behaved statistically well to the yearly tea production of Assam for the period 1957-2016.

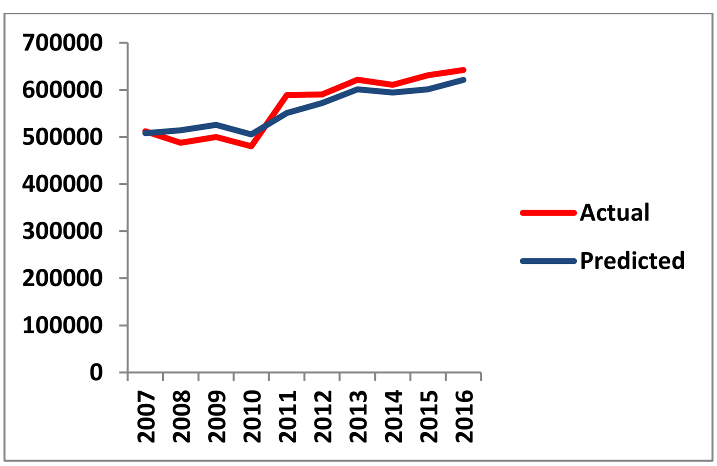

Figure 6: Plotting of Predicted values with Actual observations (1957-2016)

Moreover, the forecasted tea production of Assam for the upcoming 10 years which is given in Table 2 along with 95% confidence interval. Also, Figure 7 represents forecasted values along with 99% and 95% Confidence Interval (CI).

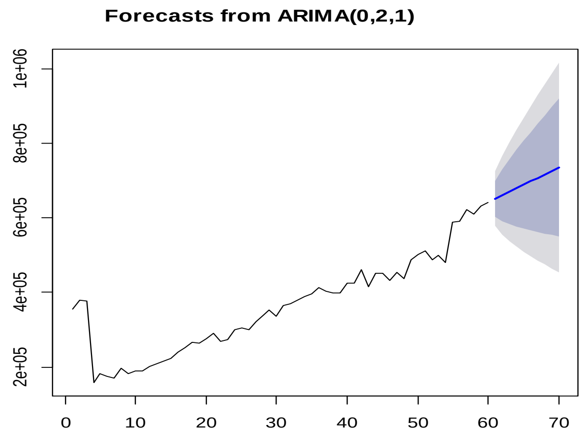

Figure 7: Trend of Tea production of Assam for upcoming 10 years

Table 2: Forecasting Tea Production of Assam for Upcoming 10 Years

Year	Point Forecast	95% CI	
		Lower Limit	Upper Limit
2017	651435.6	578392.8	724478.4
2018	660691.3	554905.6	766476.9
2019	669946.9	537319.8	802574
2020	679202.5	522495.1	835909.9
2021	688458.1	509248	867668.3
2022	697713.8	496987.5	898440
2023	706969.4	485372	928566.7
2024	716225	474185.3	958264.8
2025	725480.6	463281.3	987680
2026	734736.3	452557.2	1016915.4

4. CONCLUSION

This paper aimed to study the yearly tea production of India using the Autoregressive Integrated Moving Average (ARIMA) approach. The best selected model for forecasting the tea production of Assam is ARIMA (0, 2, 1). It gives ten years of yearly tea production data which revealed an increasing pattern during upcoming 10 years. The forecasting of tea can help tea garden owners as well as the policy makers for future planning.

REFERENCES

[1] Arya, N. (2013). Indian Tea Scenario. *International Journal of Scientific and Research Publications*, 3(7), 1–10.

[2] Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. holden-day, San Francisco.

[3] Gijo, E. V. (2011). Demand forecasting of tea by seasonal ARIMA model. *International Journal of Business Excellence*, 4, 111–124.

[4] Tea Board of India. (2014). Global tea statistics. https://www.teaboard.gov.in/pdf/ Global_Tea.pdf.

[5] Verma, V. K., Kumar, P. and Singh, H. (2016). Use of ARIMA modeling in forecasting coriander prices for Rajasthan. *International Journal of Seed Spices*, 6(2), 40–45.

[6] Weerapura, W. H. E. B. P. and Abeynayake, N. R. (2013). Forecasting of Tea Production Using Time Series Models. *Proceedings of 12th Agricultural Research Symposium*, 351–355.