ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

Future Projection Of Production Of Tea In Assam: By Using ARIMA Model

[1]**Pranjeet Borah &** [2]**Sahista Amrin**

[1]Department of Statistics, Dibrugarh University, Dibrugarh-786004, Assam, India
Email:pranjeetborah123@gmail.com

[2]Research Scholar, Department of Statistics, Dibrugarh University, Dibrugarh-786004, Assam, India
Email:amrin.sahista@gmail.com

ABSTRACT

In this paper a best fitted Auto Regressive Integrated Moving Average model commonly known as ARIMA model has been developed for future projection of tea production in the state Assam by using various statistical tools and techniques viz., SPSS, R-software. We have also used various statistical test to study the nature of data and other time series related test. Data are collected from Tocklai Tea Research Institute, which is a pioneer institute of tea research in Assam. The type of data used in this study are secondary in nature. From the study we found that ARIMA (1,1,1) is suitable for the data set and we forecasted upcoming Fifteen years.

Keywords: Auto Regressive Integrated Moving Average (ARIMA), SPSS, R- software.

1. INTRODUCTION:

Tea is most commonly and widely used beverage product in India as well in Assam. Now-a-days more than thirty countries of the world produces tea which includes India, China, Sri Lanka, Kenya, Indonesia etc. India is considered as principal tea producing country of the world. India produces a significant amount of tea, which share 26 percent of world tea production. Tea industry of India is one of the oldest industries having more than 180 years old history. In India most of the tea are produced by the states like Assam, West Bengal, Tamil Nadu and Kerala. Assam is one of the tea producing state of India, it produces almost fifty three percent of India's total tea production having plantation area of about 3.22 lakh hectares which is more than half of the country's total area under tea[1]. Tea industry is one of the most employments generating industry in Assam. This industry plays a vital role in earning foreign currency in the country. The first Indian to start planting of tea was an Assamese nobleman Maniram Dutta Barma, popularly known as Maniram Dewan[5].The growth rate of production of tea in India in general and Assam

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

in particular is not satisfactory as compared to the other tea producing countries like China, Sri Lanka, Kenya etc. India occupied first position till 2005 in terms of world tea production, but China occupied first position in terms of production in the year 2006 which forced India in second position. It is the single largest industry in Assam that provides average daily employment to more than 6.86 lakhs persons in the State. Assam tea is manufactured from the plant Camellia sinensis var. assamica. Assam tea is famous for its malty flavor, bright colour and briskness. In this paper, ARIMA model is used to forecast the future production of Tea in Assam.

2. Materials and Methods:

2.1 Data source:

This study is based on the secondary data which is taken from Tocklai Research Centre. Tocklai Tea Research Center- a trend setter in the world of tea research for the last century. Whenever we hear about Tocklai, the very first name that comes to our mind is TRA (Tocklai Research Association). TRA gave birth to many of the innovations as well as manufacturing techniques in the tea processing.

2.2 Auto Regressive Integrated Moving Average (ARIMA) Model :

ARIMA is an acronym that stands for Auto Regressive Integrated Moving Average. It is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so the equation can be used to forecast future values. An ARIMA model is characterized by three terms: p, d, q which defines: p is the order of the AR term, q is the order of the MA term and d is the number of differencing required to make the time series stationary

Generally, an ARIMA model is denoted by ARIMA (p, d, q) where, p is the number of auto regression parameters, d is the order of differencing need to make the data stationary and q is the number of moving average parameters. The ARIMA model is given by-

$$\varphi(B)(1- B)^{d} y_t = \mu + \theta(B)e_t \qquad (2.1)$$

$\varphi(B)$ and $\theta(B)$polynomials are the autoregressive and moving average components of orders p and q respectfully.

$$\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\ldots\ldots\ldots-\varphi_p B^p \qquad (2.2)$$

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

To choose the best models among the fitted ones, some models choice criteria like AIC, Log-likelihood has also been used here. For the diagnostic checking, the significance of the AR and MA coefficients are carried out through Z test. For model accuracy, the values of Error measures like Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

Error (MAE) etc. are calculated here. To check the normality of the residuals of the ARIMA Models, "Normal Q-Q plot" is also constructed here.

2.3 Jarque Bera Test:

The normality of the data is checking by using Jarque Bera[2] test. Which is goodness of fit based on sample kurtosis and skewness. The test statistic for Jarque-Bera test is given by $\mathbf{JB=\frac{n}{6}\left(s^2+\frac{(k-3)\hat{}2}{4} \right) \sim \chi^2_{(2)}}$, where n is the number of observations, k is the sample kurtosis and s is the sample skewness. The statistic of Jarque-Bera test follow Chi-square with 2 degrees of freedom.

2.4Augmented Dickey-Fuller (ADF) test:

ADF (Augmented Dickey-Fuller) test Augmented Dickey-Fuller) is a statistical significance test which means the test will give results in hypothesis tests with null and alternative hypotheses. As a result, we will have a p-value from which we will need to make inferences about the time series, whether it is stationary or not. This test examines the null hypothesis of an autoregressive integrated moving average (ARIMA) against stationary and alternatively

Explanation of the Dickey-Fuller test: A simple AR model can be represented as:

$$y_t = \rho y_{t-1} + u_t$$

where, y_t is variable of interest at the time t , ρ is a coefficient that defines the unit root, u_t is noise or can be considered as an error term.

The formal version of Dickey Fuller test is explained here:

Consider an AR (1) model: $y_t = \rho y_{t-1} + \varepsilon$ ……. (1)

Dickey Fuller suggest an alternative equation by subtracting $Y_{t-1}$from both sides of equation (1)

$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \varepsilon$

$\Delta y_t = (\rho-1)\, y_{t-1} + \varepsilon$

$\Delta y_t = \Upsilon\, y_{t-1} + \varepsilon$ ……...(2)

where $\gamma = \rho-1$

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

Dickey – Fuller also suggest two alternate forms:

$\Delta y_t = \alpha + \Upsilon y_{t-1} + \varepsilon$ .....(3)

$\Delta y_t = \alpha + \beta t + \Upsilon y_{t-1} + \varepsilon$.....(4)

Testing of hypothesis is done via t-test statistic

$$t = \frac{\hat{\Upsilon} - H0}{SE(\hat{\Upsilon})}$$

If the calculated value of test statistic t is greater than the tabulated value then we reject the null hypothesis otherwise we accept the null hypothesis. If we accept the hypothesis it means that the series is stationary other series is not stationary.

3. Results and Discussions:

To forecast a time series, first of all it is necessary to check the normality and stationarity of the data. To check normality of the data we use Jarque-Bera test and to checking stationarity we use Dicky-fuller test which we have discussed already.

To check normality of the data the Jarque-Bera test is performed here with the help of the software R and the results are shown in Table 1.

Table 1: Jarque-Bera Test for Normality Check

Hypothesis	**Test Statistic**	**Degrees of freedom**	**p-value**
H_0: the data is normally distributed H_1: the data is not normally distributed	Jarque Bera = 2.0807	2	0.3533

Since the p-value greater than 0.05 so, we may accept the null hypothesis that the data is normally distributed.

The time series plot of the data is shown in Fig-1 which exhibits an upward stochastic trend.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

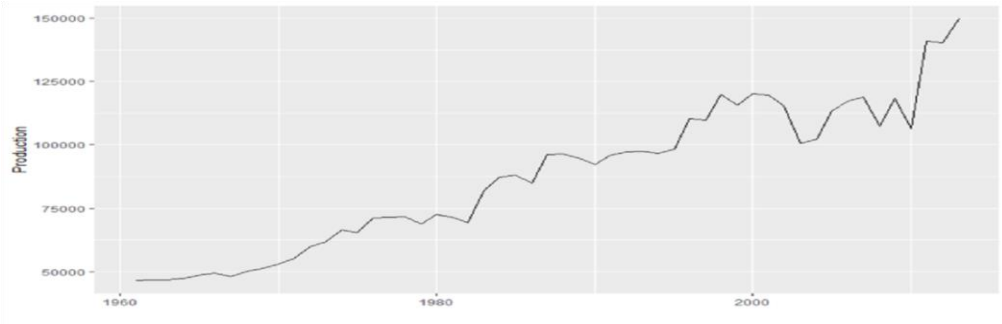

Fig1: Time series plot of tea production of south bank from the year 1960-2013

To test the stationarity of the data graphically the Auto Correlation Function (Acf), Partial Auto Correlation Function (Pacf) is plotted with help of software R and are shown in Fig-2 and Fig-3.

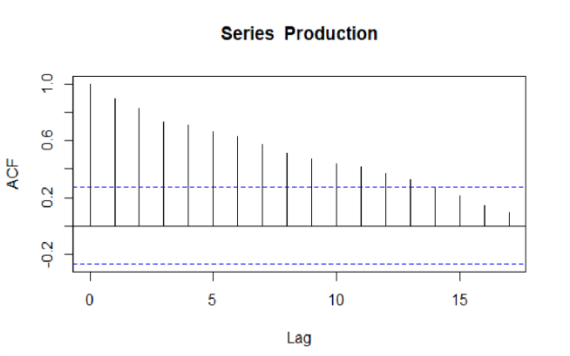

Fig 2: Auto Correlation Function (Acf) for the tea production data

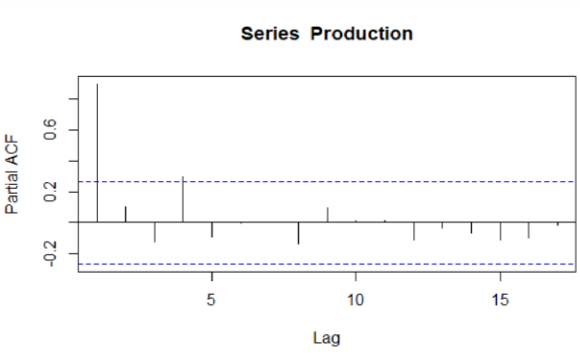

Fig 3: Partial Auto Correlation Function (Pacf) for the production data

Since in Fig 2 most of the bars crosses the upper limit and in Fig 3 two bars crosses the upper limits so our data is not stationary in nature according to both the criteria acf and pacf. To test it statistically we use Augmented Dickey-Fuller Test with the help of the software R and the results are shown in Table 5.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

Table 2: Augmented Dickey-Fuller Test

Hypothesis	Test Statistic	Lag order	p-value
H_0:The data is not Stationary H_1: The data is Stationary	Dickey-Fuller = -2.6766	3	0.3021

Since the p-value is greater than 0.05 so we may accept the null hypothesis that the data is not stationary.

Fitting of ARIMA Model:

Since tea production data becomes stationary after taking the first difference. So the in the ARIMA(p,d,q) model the order of the d is identified as 1. Keeping d=1 as constant five ARIMA models are proposed here to fit the data and the best model will be choosen according to some criteria like AIC and Maximum Likelihood etc. The proposed ARIMA models are,

Model 1= ARIMA (1,1,0): $Y_t = (1+\varphi_1)\, Y_{t-1} - \varphi_1\, Y_{t-2} + \varepsilon_t$…….(1)

Model 2 = ARIMA (0,1,1): $Y_t = Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$…….(2)

Model 3 = ARIMA (1,1,1): $Y_t = (1+\varphi_1)\, Y_{t-1} - \varphi_1\, Y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$…….(3)

Model 4 = ARIMA (2,1,1): $Y_t = (1+\varphi_1)\, Y_{t-1} + (\varphi_2 - \varphi_1) Y_{t-2} - \varphi_2 Y_{t-3} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ …….(4)

Model 5 = ARIMA (2,1,2): $Y_t = (1+\varphi_1)\, Y_{t-1} + (\varphi_2 - \varphi_1) Y_{t-2} - \varphi_2 Y_{t-3} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$ …(5)

After fitting the proposed models in R software we get the estimated models as follows,

Model 1= ARIMA (1,1,0): $Y_t = 0.7722\, Y_{t-1} + 0.2278\, Y_{t-2} + \varepsilon_t$ …….(6)

Model 2 = ARIMA (0,1,1): $Y_t = Y_{t-1} + \varepsilon_t + 0.1588\, \varepsilon_{t-1}$ …….(7)

Model 3 = ARIMA (1,1,1): $Y_t = 0.1489 Y_{t-1} + 0.8511 Y_{t-2} + \varepsilon_t - 0.6241\, \varepsilon_{t-1}$ …….(8)

Model 4 = ARIMA (2,1,1): $Y_t = 0.3826\, Y_{t-1} + 0.8102 Y_{t-2} - 0.1928 Y_{t-3} + \epsilon_t - 0.4952 \varepsilon_{t-1}$…(9)

Model 5 = ARIMA (2,1,2): $Y_t = -0.1561 Y_{t-1} + 0.5933 Y_{t-2} + 0.5628 Y_{t-3} + \varepsilon_t - 1.2471 \varepsilon_{t-1} - \varepsilon_{t-2}$ …… (10)

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

Table 3: AIC and Log Likelihood of the fitted ARIMA models

Model	ARIMA Order	AIC	Log Likelihood
Model 1	(1,1,0)	1074.43	-535.21
Model 2	(0,1,1)	1075.33	-535.66
Model 3	(1,1,1)	1072.19	-533.10
Model 4	(2,1,1)	1072.84	-532.42
Model 5	(2,1, 2)	1068.82	-529.41

From the Table it is observed that the AIC value for Model 3: ARIMA(1,1,1) is slightly lower as compared to the other fitted models. So here Model 3 is used to forecast the data

Diagnostic Checking of the ARIMA (1,1,1) Model :

To get an idea about the significance of the AR and MA coefficients Z test is performed here and the test results are shown in the Table 8.

Table 4: Z test of the AR and MA coefficients

Coefficients	Estimate	Std. Error	Z value	Pr (>|z|)
AR1	-0.85107	0.13662	-6.2296	$4.678x10^{-10}$
MA1	0.62407	0.15853	3.9367	$8.262x10^{-05}$

From the Table 8 it is observed that the p-values are less than 0.01. So we may conclude that both the AR and MA coefficients are highly significant.

Forecasting with the help of ARIMA(1,1,1) Model:

The forecasted values (Point Forecast) along with 80% and 95% Upper and Lower Confidence intervals for next 15 years based on ARIMA(1,1,1) Model are calculated with the help of R software and are given in Table 10 and the it is diagrammatically represented in Fig4

Year	Point Forecast	80% CI(Lower Limit)	80% CI(Upper Limit)	95% CI (Lower Limit)	95% CI(Upper Limit)
2014	143080.8	134316.7	151844.8	129677.34	156484.2
2015	148637.6	137560.5	159714.8	131696.57	165578.7
2016	143908.3	129965.4	157851.3	122584.42	165232.2
2017	147933.3	132319.8	163546.8	124054.51	171812.1

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

2018	144507.8	126847.2	162168.3	117498.25	171517.3
2019	147423.1	128347.7	166498.5	118249.82	176596.5
2020	144941.9	124218.9	165665.0	113248.80	176635.1
2021	147053.6	125068.5	169038.7	113430.32	180676.9
2022	145256.4	121866.1	168646.7	109484.04	181028.8
2023	146786.0	122241.7	171330.3	109248.69	184323.2
2024	145484.2	119698.8	171269.7	106048.78	184919.6
2025	146592.1	14119736.1	173448.1	105519.44	187664.8
2026	145649.2	117671.0	173627.4	102860.26	188438.2
2027	146451.7	117470.8	175432.5	102129.27	190774.1
2028	145768.7	115756.1	175781.3	99868.39	191669.0

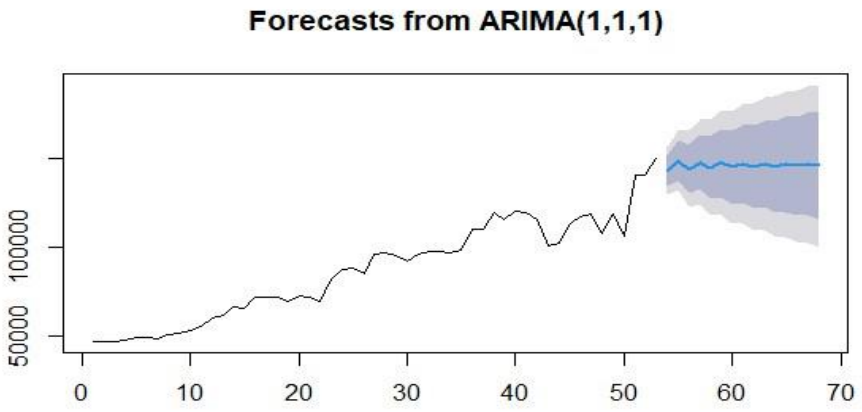

Fig-4: Plot of forecasted Tea Production of South Bank using ARIMA(1,1,1) Model

4. CONCLUSION:

In this study, Time series analysis is used to analyze the data and to forecast the data with the help of suitably fitted ARIMA Model. In analysis we use Jarque-Bera test to check the normality of the data. And the test results indicates that the data are normally distributed. To test the stationarity of the data we have plot the Auto Correlation Function(Acf), Partial Auto Correlation Function and perform Augmented Dickey-Fuller Test and they indicates that the data is not stationary. After taking the first simple difference the data becomes stationary. Since tea production data becomes stationary after taking the first difference. So in the ARIMA(p,d,q)

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

model the order of the d is identified as 1. Keeping d=1 as constant five ARIMA models were proposed here to fit the data and the best model will be chosen according to some criteria like AIC and Maximum Likelihood. ARIMA(1,1,1) is the best fitted model for the tea production data so it is used to forecast next 15 years tea production and a good number of tea productions are predicted using this model.

REFERENCES :

1. Laskar, N., & Thappa, S. (2015). A study on the present scenario of tea industry in Assam-Challenges ahead. *Indian Journal of Applied Research*, *5*(11), 553-537.
2. Mushtaq, R. (2011). Augmented dickey fuller test.
3. Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.
4. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427-431.
5. Narzary, S. (2016). A study on the status of growth and development of tea industry in Assam. *International Journal of Humanities and Social Science*, *3*(4), 1-9.
6. Pujashree Borah, A Study on the Problems and Strategies required for the development of Small Tea Growers in Assam With special reference to Dibrugarh District, *International Journal of Humanities & Social Science Studies (IJHSSS)*, Volume-II, Issue-VI, May 2016