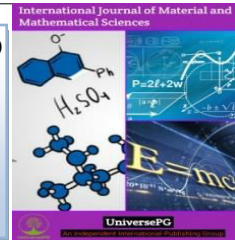




Publisher homepage: [www.universepg.com](http://www.universepg.com), ISSN: 2707-4625 (Online) & 2707-4617 (Print)  
<https://doi.org/10.34104/ijmms.020.08015>  
**International Journal of Material and Mathematical Sciences**  
 Journal homepage: [www.universepg.com/journal/ijmms](http://www.universepg.com/journal/ijmms)



## Fitting ARIMA Model and Forecasting for the Tea Production, and Internal Consumption of Tea (Per year) and Export of Tea

Md. Aliul Islam<sup>1\*</sup>, Mst. Sharmin Akter Sumy<sup>1</sup>, Md. Alhaz Uddin<sup>2</sup>, and Md. Sazzad Hossain<sup>1</sup>

<sup>1</sup>Dept. of Statistics, Islamic University, Jhenidah-Kushtia, Bangladesh; and <sup>2</sup>Dept. of Statistics, University of Rajshahi, Rajshahi, Bangladesh

\*Correspondence: [aliulstatiu2010@gmail.com](mailto:aliulstatiu2010@gmail.com)

### ABSTRACT

Every year, in Bangladesh the production has increased by 1.84 % and contributes 1.37 in export in the world tea trade and earns nearing 1775 million Taka. The internal consumers of the country are presently consuming about 98% of its produce. The consumption is increasing day by day mainly due to the rapid increase in population. This study considered the published secondary data of yearly tea production in Bangladesh over the period 1990 to 2015 by the Bangladesh Tea Board. We are forecasting the production of tea and internal consumption of tea by using Auto-Regressive Integrated Moving Average (ARIMA) for the next 5 years. Run test, Jarque and Bera test criteria are used for the adequacy of the fitted model which followed by the residual analysis. The comparison between the original series and forecasted series are showing the same trends which indicate the fitted model is statistically well and appropriate for the forecast the productions of tea in Bangladesh.

**Keywords:** ARIMA Model, Tea production, Forecasting, Consumption, Internal consumption, and Export.

### INTRODUCTION

Tea is one of the best export items of Bangladesh. But exporting tea of the country has been declined day by day for the calm growth of production rising consumption and strong competition from other tea exporting countries in the world. Bangladesh attester substantial changes for exporting tea to several countries over time (Uddin *et al.*, 2020). It is necessary to search new customers and explore an incursive policy for cope with similar changes in the world tea market.

This paper gives several suggestions and examines present situation for raising tea export. Domestic tea consumption has increased steadily. Likely to remain this trend and even may increase day by day. In the face of rising domestic consumption, a sustained increase in tea production can be stable -

level of tea export maintained (Islam *et al.*, 2019). Various aspects of tea production and export have been dealt with, among others, by Khalid (1975), Majid (1991) and Sabur (2002).

### Objectives of the Study:

- Investigate different properties of time series plot.
- To test the non-stationarity of the series by Graphical methods
- Examination of the present status of production, domestic consumption and export of tea of the country.
- Select an appropriate multivariate time series model for the data.

**METHODOLOGY**

**Dickey-Fuller unit root test** - Dickey and Fuller suggested that under the null hypothesis the estimated coefficient of  $y_{t-1}$  in the model above follows the tau ( $\tau$ ) statistic is known as Dickey-Fuller test.

Allowing the various possibilities DF test is estimated in three different forms, under different null hypothesis

$y_t$  is a random walk:  $\Delta y_t = y_{t-1} + u_t$

$y_t$  is a random walk with drift:  $\Delta y_t = \beta_1 + \delta y_{t-1} + u_t$ .

$y_t$  is a random walk with drift around a stochastic trend-  $\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + u_t$ .

Now we have to test the following hypothesis-

$H_0 : \delta = 0$  i.e., the time series data of production of rice of X is non-stationary,

$H_a : \delta < 0$  i.e., the time series data of production of rice of X is stationary.

Where, X is a particular series concerned in this report.

The appropriate test statistic is tau statistic. This test is also known as Dickey-Fuller (DF) test and is

defined by,  $\tau = \frac{\hat{\delta}}{se(\hat{\delta})}$

**Decision:** If the computed  $|\tau| \geq$  the DF critical value, then we do not accept the null hypothesis, otherwise accept.

**Augmented Dickey-Fuller test** - In case of the error term  $u_t$  are correlated. Dickey and Fuller have developed a test named as the Augmented Dickey-Fuller (ADF) test. The following regression model is estimate by the ADF test:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^m \Delta Y_{t-i} + \varepsilon_t$$

Where  $\varepsilon_t$  is a pure white noise error and where

$$\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2}), \Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$$

We estimate the model (3.3) and test the following hypothesis

$H_0 : \delta = 0$  i.e. the time series is non-stationary.

$H_a : \delta \leq 0$  i.e. the time series is stationary.

The appropriate test statistic under  $H_0$ ,

$$\tau = \frac{\hat{\delta}}{se(\hat{\delta})}$$

**Decision:** If the computed  $|\tau| \geq$  the DF critical value, then we is not accept the null hypothesis, otherwise accept.

**Box-Jenkins Methodology** - The puissant work of Box-Jenkins (1970) expelled professional intentness away from the stationary serially correlated deviations from deterministic trend paradigm toward the  $ARIMA(p, d, q)$  paradigm. It can manage any series, stationary or not with or without seasonal elements that's why it is popular.

The following five basic steps consist in the Box-Jenkins methodology:

- 1. Identification of appropriate model:** We test the Correlogram to conclude on the suitable orders of the AR and MA components, for this we used the different procedure to get a stationary time series. After a point, the Correlogram of a MA process is zero, that of a AR process declined geometrically. Different patterns are showed by the Correlogram of ARMA process (but all dampers after a while). On this basis, one arrives at a tentative ARMA model. This step includes more of a judgment procedure than the use of any clear-cut rules.
- 2. Estimation of the model:** The next step is the estimation of the tentative ARMA model identified in step-2. The estimation of AR model is straight forward. We can estimate OLS by minimizing the error sum of squares  $\sum Z_t^2$ . In case of MA models, Box-Jenkins expressed a grid-search procedure. In that procedure we compute  $\hat{Z}_t$  by successive substitution for each value of the MA parameters and choose the set of values of the parameters that minimizes the error sum of squares  $\sum \hat{Z}_t^2$ . For ARMA models, both the AR and MA parts are to

be estimated with the procedure discussed above.

- 3. Diagnostic Checking:** It is an advisable to check that the model does actually given an adequate statement on the date when a AR, MA, ARMA has been fitted to bestowed time series. There are two measures after used such as reflecting the nearness of fir and the number of parameters estimated. One is called the Akaike information criterion (AIC) and the other one is Schewatz Bayesian information criterion (BIC). If  $s$  is the total numeral estimated parameters

$$ACI(s) = n \log \hat{\sigma}^2 + 2s$$

and  $BIC(s) = n \log \hat{\sigma}^2 + s \log n$

Here  $n$  is the sample size.

If  $RSS = \sum \hat{Z}_t^2$  is the residual aggregate squares, then  $\hat{\sigma}^2 = \frac{RSS}{n - p}$

If we consider different ARMA models we select the one which has the smallest AIC of BIC. The two criteria can lead to several conclusions.

- 4. Forecasting:** Imagine that we estimate the model which has  $n$  observations, and then we want to forecast  $Y_{n+k}$ . That is told a  $k$ -periods ahead forecast. First we want to write out the evolution for  $Y_{n+k}$  and then replace all future values  $Y_{n+k}$  ( $0 < j < k$ ) with their factors and  $Z_{n+j}$  ( $j > 0$ ) by zero (since the expected value is zero). We also replace all  $Z_{n-j}$  ( $j \geq 0$ ) by the predicted residuals.

**Moving average process or MA (q) -** A series  $\{y_t\}$  is said to be a moving average method of order  $q$  if,

$$y_t = z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_q z_{t-q}$$

Where,  $\{z_t\} \sim WN(0, \sigma^2)$  and  $\theta_1, \theta_2, \dots, \theta_q$  are constants. This series is also called as MA (q) process

**Autoregressive process or AR (p) process - A**

series  $\{y_t\}$  is said to be a autoregressive method of order  $p$  if,

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + z_t$$

Where,  $\{z_t\} \sim WN(0, \sigma^2)$  and

$\Phi_1, \Phi_2, \dots, \Phi_p$  are constants. The series is said as AR (p) process

**ARMA (p,q) process -** The time series  $\{y_t\}$  is an ARMA (p,q) process if it stationary and satisfies for every  $t$ ,

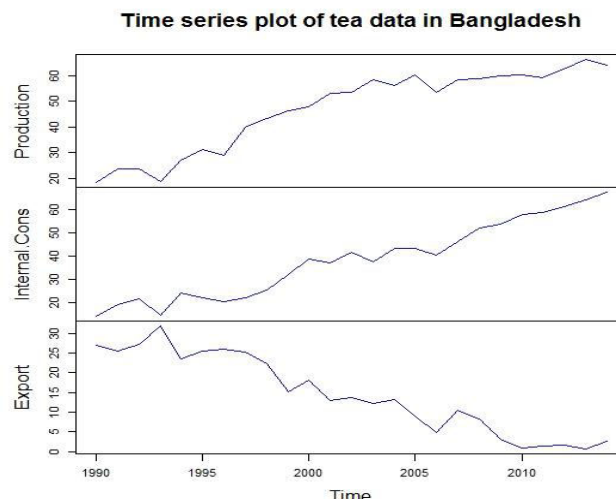
$$y_t - \Phi_1 y_{t-1} - \dots - \Phi_p y_{t-p} = z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \dots + \theta_q z_{t-q}$$

or,  $y_t - \sum_{i=1}^p \Phi_i z_{t-i} = z_t + \sum_{i=1}^q \theta_i z_{t-i}$

Where,  $\{z_t\} \sim WN(0, \sigma^2)$  and the polynomials  $(1 - \Phi_1 z - \dots - \Phi_p z^p)$  and  $(1 + \theta_1 z + \dots + \theta_q z^q)$  have no common factors.

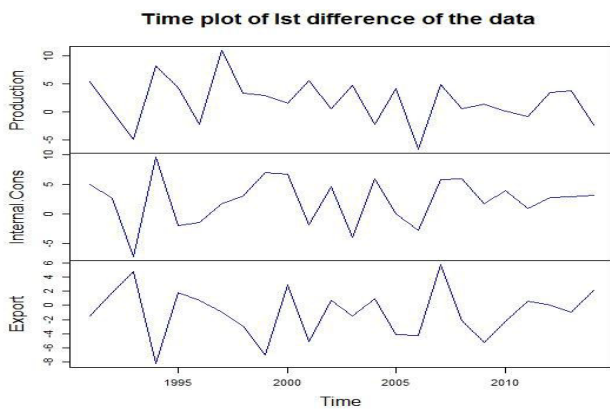
**ARIMA (p,d,q) process -** If the ARMA (p,q) process is also integrated process than it is known as ARMA process of order (p,d,q). Here  $d$  represents the number of difference the needed to be stationary. ARIMA process is used when data is non-stationary.

**Analysis and Findings -**A time series is told to be stationary if it's mean, variance and covariance function does not depend on time, and otherwise it is non-stationary. The time series plot of the data is given below:



**Fig 1:** Time series Plot of the data.

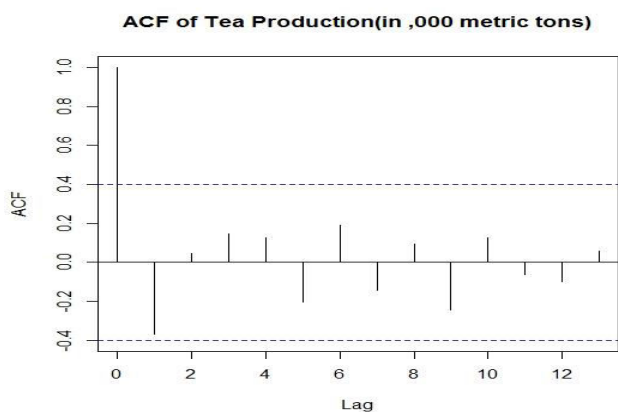
**Comments:** From the above time plot we see that the mean and variance of the tea production, internal consumption, export and export price of Bangladesh from 1990-2014 is not same over time. That is the mean and variance has changed over time. So from the properties of non-stationary time series we can tell that the above time series is non-stationary process. We know that in the time series analysis it is possible to make the non-stationary time series into stationary time series. In order to make the above non stationary time series into stationary process we have to make some differences of the tea production.



**Fig 2:** Time series plot of first difference of the data.

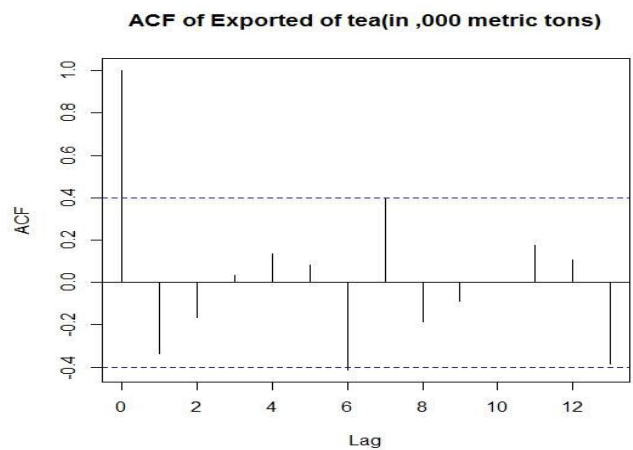
**Comments:** From the above figure we see that variance are almost stable here and the production of tea changes equally over time i.e, mean and variance are about to same. So, we can ask that the series is Stationary.

**The Auto-Correlation Function of ARMA -** We get the following acf of  $i^{st}$  difference of tea production,



**Fig 3:** ACF of Tea Production after taking one difference.

We get the following acf of  $i^{st}$  difference of Export of tea,

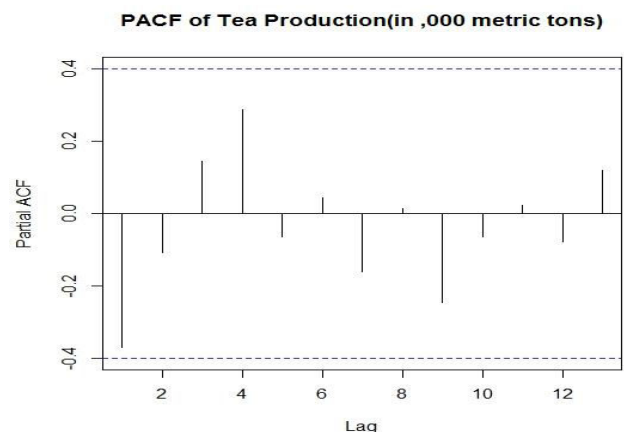


**Fig 4:** ACF of export of Tea after taking one difference.

**The Partial Auto-Correlation Function of an ARMA**

**Comments:** From the above figure of ACF curve of first difference on tea production is statistically significant at the lag 1. So we can mention that the production of tea follows AR (1) Process.

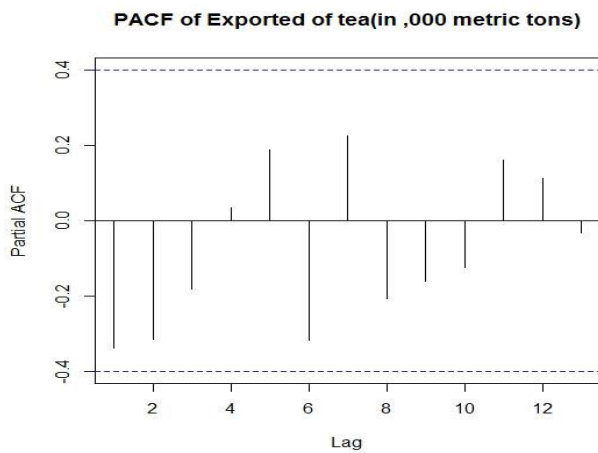
We get the following pacf of  $i^{st}$  difference of tea production,



**Fig 5:** PACF of tea production after taking one difference (thousand metric tons).

**Comments:** From the above figure of Partial auto-correlation function of tea production we can mention that the series is a Moving Average Process of order 0.

We get the following pacf of export of tea,



**Fig 6:** PACF of exported of tea (thousand metric tons).

**Comments:** From the above figures it is evident that of ACF and PACF curve of first difference on tea production is Statistically significant at the lag 1. Therefore the time series of production of tea follows ARIMA (1,1,0). Again, From the above figures it is evident that of ACF and PACF curve of first difference on export tea from Bangladesh is Statistically significant at the lag 1. Therefore the time series of production of tea follows ARIMA (1,1,0).

**Unit root test** - The stationarity of a time series can be tested directly with a unit root test. Two tests are available, such as DF test and ADF test. DF is useful when errors are uncorrelated but when errors are correlated then ADF test is useful.

**Dickey Fuller Test** – Data = Tea production

Consider the hypothesis  $H_0 : \delta = 0$   
 $H_0 : \delta < 0$

At 5% level of significance we will conduct the test.  
 P-Value of the Dickey-Fuller test.  
 P-Value: 0.01  
 This is less than level of significance (0.05)  
 So, we are not accepting the null hypothesis.

**Comments:** Since the null hypothesis is not accepted so we can say that the series of tea production after taking first differences is stationary.

**Data = Export tea**

Consider the hypothesis:  $H_0 : \delta = 0$   
 $H_0 : \delta < 0$

At 5% level of significance we will conduct the test.  
 We get the following P-Value of the Dickey-Fuller test.

P-Value: 0.0106

This is less than level of significance (0.05)

So, we are not accepting the null hypothesis.

**Comments:** Since the null hypothesis is not accepted so we can ask that the series of export tea after taking first differences is stationary.

**Augmented Dickey- Fuller Test** - Consider the hypothesis,

$H_0$  : The tea production is non-stationary.

$H_1$  : The tea production is stationary.

**Data: Tea production**

Dickey-Fuller = -0.9184, Lag order = 1, p-value = 0.9326

At 5% level of significance the p-value of Augmented Dickey-Fuller Test is 0.9326, which is greater than 0.05. So the null hypothesis is not rejected.

That means, the tea production is non-stationary.

**Augmented Dickey- Fuller Test** - Consider the hypothesis,

$H_0$  : The export of tea is non- stationary.

$H_1$  : The export of tea is stationary.

Data: Export of Tea

Dickey-Fuller = -3.529, Lag order = 2, p-value = 0.04172

At 5% level of significance the p-value of Augmented Dickey-Fuller Test is 0.04172, which is smaller than 0.05. So the null hypothesis is rejected.

That means, the export of tea is stationary.

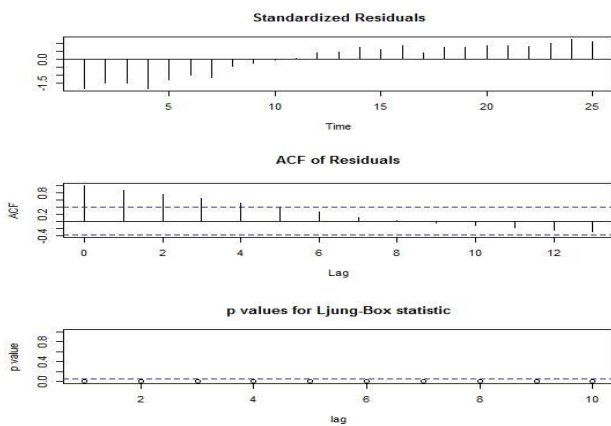
**Fitted ARIMA Model for tea data** - The fitted ARIMA (1,1,0) model for the production of tea which we find from R-Console is given below:

$$X(t) = 46.9224 + 3.0592 * X(t-1)$$

The fitted ARIMA (1,1,0) model for export of tea which we find from R-Console is given below:

$$X(t) = 14.486 + 1.975 * X(t-1)$$

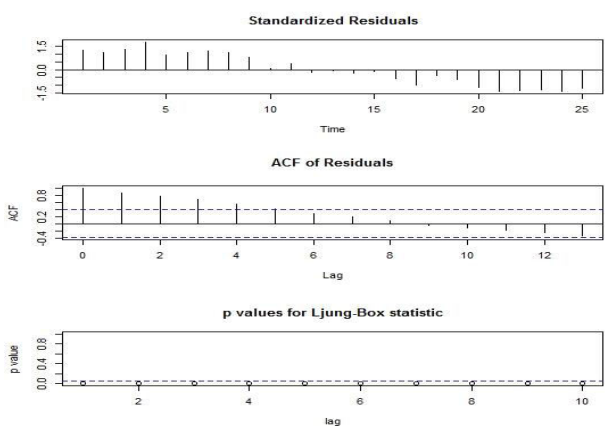
**Diagnostic checking for tea production in Bangladesh** - At this time we have to perform diagnostic checking for the fitted model. Now from R3.0.2 we have found the diagnostic checking plot, which is given below:



**Fig 7:** Diagnostic checking for the tea production in Bangladesh.

**Comments:** From this plot we can observe that the plot of the standardized residual and the ACF plots are correctly specified. From the plot of P-value we can view that all p-values are above the level of significance. So, we can interpret that the residual is white noise, hence the model is also correctly estimated.

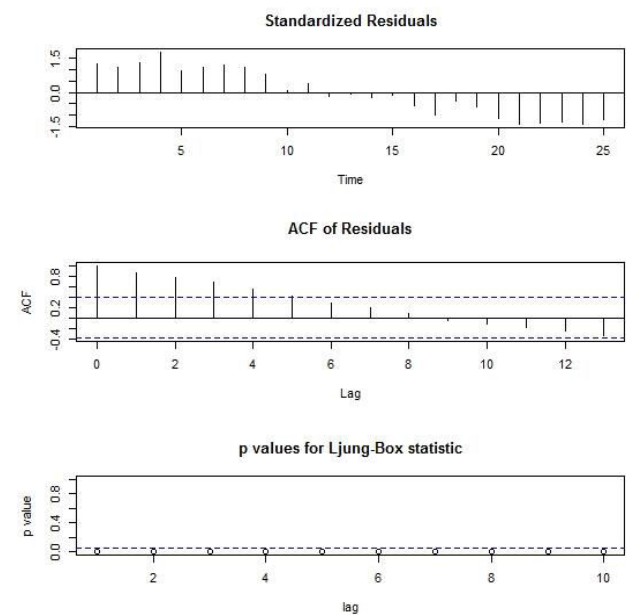
**Diagnostic checking for the internal consumption of tea** - At this point we have to perform diagnostic checking for the fitted model. Now from R3.0.2 we have found the diagnostic checking plot, which is given below:



**Fig 8:** Diagnostic checking for the fitted model of internal consumption of tea.

**Comments:** From this plot we can observe that the plot of the standardized residual and the ACF plots are correctly specified. From the plot of P-value we can see that all p-values are above the level of significance. So we, can interpret that the residual is white noise, hence the model is also correctly estimated.

**Diagnostic checking for the exported tea from Bangladesh** - At this point we have to perform diagnostic checking for the fitted model. We have found the diagnostic checking plot, which is given below:



**Fig 9:** Diagnostic checking for the fitted model of exported tea.

**Comments:** From this plot we can notice that the plot of the standardized residual and the ACF plots are correctly specified. From the plot of P-value we can observe that all p-values are above the level of significance. So we, can interpret that the residual is white noise, hence the model is also correctly estimated. From the above analysis we may say that the production of tea and the amount of exported tea are non-stationary process.

After taking successive differences we have made all the non-stationary time series into stationary process. In these analysis, production of tea follows ARIMA (0,1,1) model, export of tea follows ARIMA (0,1,1) model. After diagnostic checking for all types of data of tea we have found that all the models are appropriate.

## CONCLUSION

The main concept of this work is forecasting a group of associated variables thus making use of all possible information such as correlation. The correlations among the export variables may be high or comparatively low (Islam *et al.*, 2019). In this project, we focus better forecast which can give strong correlation among these associated export variables. In multivariate time series models, the vector autoregressive model is used for forecasting. For this we find out the forecasted values. We measure the forecasting accuracy by using mean error (ME), mean square error (MSE), Root Mean Squared Error (RMSE), mean absolute error (MAE), mean percent error (MPE), mean absolute percent error (MAPE), mean absolute percent error (MSPE) and so on. We also select a proper model for forecasting on the basis of model selection criterions such as AIC and BIC, HQ and SC of the model. The others highly correlated variables forecast in multivariate forecast are good. The forecasting accuracy ME, MSE, RMSE, MAE, MPE, MAPE and so on all are not so very distance in multivariate forecast. Finally we conclude that before forecasting we should check correlations among the variables, if existing highly correlation then multivariate time series is better than univariate time series. So all over we conclude that if we want to forecast, at first checking correlations among the variables. If there existing highly correlated variables then we should use multivariate forecast for these variables. If there do not exist highly correlated variables then we should use univariate forecast for these variables. Since our case study export variable are highly correlated to production and internal consumption for the selected export variable than univariate time series analysis using ARIMA model.

## ACKNOWLEDGEMENT

We are gratefully acknowledged to my Supervisor, Dept. of Statistics, Islamic University, Jhenidah-Kushtia, Bangladesh for logistically supported and Co-author who encouraged and supported to the successful research.

## CONFLICTS OF INTEREST

The authors declared no conflicts of the interest with respect to the present research work.

## REFERENCES

1. Anderson, T.W. (1984). "An introduction to Multivariate Statistical Analysis", 2<sup>nd</sup> ed., New York: John Wiley and Sons Inc.  
<https://www.wiley.com/en-us/An+Introduction+to+Multivariate+Statistical+Analysis%2C+3rd+Edition-p-9780471360919>
2. Bangladesh bureau of statistics. Ministry of planning, government of People's Republic of Bangladesh, pp. 81 & 327. Dhaka, Bangladesh.
3. Bangladesh tea research institute, government of people's republic of Bangladesh, pp. 94-98, Srimangal, Moulvibazar, Bangladesh.
4. Box, G.E.P., Jenkins, G.W, and Reinsel G. (1994). "Time Series Analysis, Forecasting and Control", 3<sup>rd</sup> edition, Englewood Cliffs, N.J.; Prentice-Hall.
5. Brokwell P. J. and R. A. Davis (1996) "Introduction to Time Series and Forecasting, Springer, New York.
6. Chatfield, C. (1993). Calculating Interval Forecasts, *Journal of Business and Economic Statistics*, 11(2), 121-35.  
[https://econpapers.repec.org/article/besjnlbes/v3a11\\_3ay\\_3a1993\\_3ai\\_3a2\\_3ap\\_3a121-35.htm](https://econpapers.repec.org/article/besjnlbes/v3a11_3ay_3a1993_3ai_3a2_3ap_3a121-35.htm)
7. Cromwell J.B., Labys, W.C., & Terraza, M. (1994). "Univariate tests for time series models", SAGE, London.
8. Diebold, F.X. (2007) "Elements of Forecasting, 3<sup>rd</sup> edition, Thompson South-Western.
9. Gebhard K. and J. Wolters (2007). "Introduction to Modern Time Series Analysis", Springer, New York.  
<https://www.springer.com/gp/book/97833642334351>
10. Gebhard Kirchgässner, Jürgen Wolters (2008). "Introduction to Modern Time Series Analysis". ISBN 3540687351, 9783540687351, Springer Science & Business Media.
11. Gujarati, D. (2002). "Basic Econometrics", 4<sup>th</sup> edition, McGraw-Hill.
12. Hamilton, J. (1969). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, p. 57, 357-384.  
<https://doi.org/10.2307/1912559>
13. Helmut Lütkepohl (2005). Department of Economics, European University Institute "New

- Introduction to Multiple Time Series Analysis*", Springer, New York.
14. Islam KA, Deeba F, and Hassan MKA. (2019). Dust Ion Acoustic Solitary Waves in Multi-Ion Dusty Plasma System with Adiabatic Thermal Change, *Aust. J. Eng. Innov. Technol.*, 1(5), 1-5. <https://doi.org/10.34104/ajeit.019.0105>
  15. Islam MT and Hossain MS. (2019). Hybridization of Vigenere Technique with the Collaboration of RSA for Secure Communication, *Aust. J. Eng. Innov. Technol.*, 1(6), 6-13. <https://doi.org/10.34104/ajeit.019.06013>
  16. Lütkepohl, H. (2005). "New Introduction to Multiple Time Series Analysis", Berlin: Springer.
  17. Lütkepohl, H. and Poskitt, D.S. (1996). "Specification of Echelon-Form VARMA Models", *Biometrika*.
  18. Uddin MM, Karim R, Kaysar MA, Dayan MAR, and Islam KA. (2020). Low-Cost Jute-Cotton and Glass Fibre Reinforced Textile Composite Sheet, *Int. J. Mat. Math. Sci.*, 2(1), 1-7. <https://doi.org/10.34104/ijmms.020.01007>
  19. Wayne A. Fuller (1995). "Introduction to Statistical Time Series", 2<sup>nd</sup> ed., first published: Print ISBN: 9780471552390.
  20. Wilfredo Palma, (2007). "Long-Money Time Series Theory and Methods." ISBN-13: 978-0470114025, ISBN-10: 0470114029, Wiley-Interscience, 1<sup>st</sup> edition.

**Citation:** Islam MA, Sumy MSA, Uddin MA, and Hossain MS. (2020). Fitting ARIMA model and forecasting for the tea production, and internal consumption of tea (per year) and export of tea, *Int. J. Mat. Math. Sci.*, 2(1), 8-15. <https://doi.org/10.34104/ijmms.020.08015> 