Current Applied Science and Technology Vol. 22 No. 2 (March-April 2022)

Research article

Modelling and Forecasting of Tea Production, Consumption and Export in Bangladesh

Farhana Arefeen Mila[1], Mst Noorunnahar[2], Ashrafun Nahar[1]*, Debasish Chandra Acharjee[1], Mst Tania Parvin[1] and Richard J. Culas[3]

[1]*Department of Agribusiness, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur-1706, Bangladesh*
[2]*Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur-1706, Bangladesh*
[3]*School of Agricultural and Wine Sciences & Institute for Land Water and Society, Charles Sturt University, Orange NSW 2800, Australia*

Received: 26 December 2020, Revised: 29 June 2021, Accepted: 30 July 2021

DOI: 10.55003/cast.2022.02.22.009

Abstract

Bangladesh is the world's 9th largest tea producer and the tea industry is a major contributor to the country's economy. In order to provide information about the demand, supply and foreign trade of tea in the future, forecasting plays a vital role in adjusting the gaps and formulating policy. Taking all of this into account, this study aims at modelling and forecasting tea production, consumption and export in Bangladesh using ARIMA modelling for the period of 2019 to 2028. Forty-seven years of time-series data from 1972 to 2018 were obtained from the Bangladesh Tea Board. Forecasts were computed on the basis of models that were selected using three important information criteria such as Akaike's Information Criterion (AIC), Schwarz's Bayesian Information Criterion (BIC) and Correction for Akaike's Information Criterion (AIC). The study identified that the best-fitting models were ARIMA (0, 1, 0), ARIMA (0, 2, 2) and ARIMA (1, 1, 2) for tea production, consumption and export, respectively. Forecasting showed an upward trend for tea production from 83.40 to 94.88 million kg and consumption from 94.35 to 131.71 million kg over the period of 2019 to 2028. On the contrary, the forecast for tea exports shows a decreasing trend. Such forecast results indicate that the government should immediately take action to accelerate the growth of the tea industry in Bangladesh. Otherwise, the economic development of the country will be hampered by reduced export earnings while relying on imports to meet the domestic demand.

Keywords

tea production;
tea consumption;
tea export;
ARIMA Model;
forecasting;
Bangladesh

*Corresponding author: Tel.: +88(02)920531-14 (Ext- 2414) Fax: +88(02)9205333
E-mail: anlaboni@bsmrau.edu.bd

1. Introduction

Tea, collected from the dried leaves of the plant *Camellia sinensis*, is the oldest and most preferred drink for two-thirds of the world's population [1]. It benefits health in many ways such as reducing cardiovascular diseases, working against obesity, conquering tumor cell growth and reducing the risk of atherosclerosis [2-5] . The tea plant is perennial and characterized as unique due to its cultivation and harvesting, and it is well-adapted to a wide range of soil and climatic conditions [2].

In Bangladesh, tea gardening was introduced during the 1840s, albeit its commercial production and expansion started in 1857. Since then, it has been contributing to national economic development through employment generation and poverty reduction [2]. It is an agro-based, labor-intensive, import substitution and export-oriented sector which employs more than 0.12 million people in Bangladesh, of which 50% are women [6, 7]. According to the International Tea Committee, Bangladesh is ranked 9th in the worldwide production and contributed 3% of the world tea production [8, 9]. The substantial production of tea has also brought a recognition of Bangladesh as a key tea exporter in the global market.

Considering the economic importance of tea, a total of 172 tea gardens have been established in Bangladesh to date, and they cover 166 thousand acres of land which is equivalent to 0.87% of total arable land [10]. Approximately 89.65 million kilograms (kg) of tea have been produced as of November 2019, which is higher than the targeted production of 74.14 million kg [7]. However, the export scenario generates disappointment when compared to the past 20 year's data. During the period 1999-2000, the quantity of tea exports was about 12.62 million kg, which had a value of 14.85 million USD (1 USD = Tk. 85 approximately) whereas during 2018-2019, the amount reduced to only 0.64 million kg, equivalent to the export earnings of 2.46 million USD only [10, 7] Such an export figure reveals that even though the quantity exported has been reduced substantially, the per unit earnings were comparatively higher than the years 1999-2000. It further indicates the potential of tea for export as a cash crop in Bangladesh. On the contrary, the local consumption scenario reveals a dramatic change from 35.25 million kg to 49.64 million kg over the same period [7]. Such a change in consumer behavior was perhaps unexpected for Bangladesh households, and was due to the increase in population, rapid urbanization, increase in per capita income, and change in tastes and so forth [6]. If this trend continues, Bangladesh may become a net tea importer in the near future which ultimately will have a negative impact on the economy of Bangladesh [11]. In such a situation, accurate forecasting of production, consumption and export of tea may help the policymakers to take the necessary steps to increase production so as to balance it with demand and also to accelerate export earnings. It may also assist them to allocate the land and other resources accordingly.

Unfortunately, there were not many studies reporting on forecasting of tea production, consumption and export in the context of Bangladesh. In the body of literature, a very old study by Shilpi [12] estimated the export demand for tea and found that the price elasticity of export demand was zero. A similar study was conducted by Das [13] which showed that the export demand of tea was perfectly elastic. Another study by Ahammed [6] determined the investment needs for the tea industry in Bangladesh while Sobuj [14] assessed the socio-economic status of tea garden workers. Even though Nasir and Shamsuddoha [3] examined all of the three highlighted factors in their study, it was not based on forecasting. Rather, it was merely a scenario analysis within a limited time period (up to 2006) and was lacking from a methodological standpoint. Hossain and Abdulla [2] performed forecasting of tea production using ARIMA model that covered 43 years of time-series data. However, the study failed to address consumption and export scenarios relevant to the tea industry in Bangladesh. Although a very new study by Islam *et al.* [15] is similar to our study in some ways, their study used a more limited dataset (1990-2015) than this study. Also, the study had no forecasted values for tea and only revealed the correlation of the variable. From that pursuit, this study contributes as a comprehensive one that relates all three important aspects of tea gardening, as well as analysis of data which cover a period of 47 years (1972- 2018). Thus, our study aims to cover the existing gap in the literature.

As already described, the main aim of this study is to forecast the production, consumption, and export status of tea in Bangladesh using the most widely used ARIMA modelling. Numerous studies used the Box Jenkins methodology for forecasting internal production, consumption, imports and exports of different products [16-20]. This method has the potential to successfully describe the past values of time series data and can make future projections with minimum error. However, forecasting based on the ARIMA model shows only linear relationship in the variables while other factors, i.e. production quality, technologies and environmental factors, cannot be considered in this model [21]. It is expected, however, that our predictions will provide some important information and useful guidelines to the producers, future researchers and policymakers, for the sustainable development of the tea industry in Bangladesh.

2. Methodology

2.1 Data

The time-series data of tea production, consumption and export (million kg) were collected from the Bangladesh Tea Board for the period 1972 to 2018 [7]. The study divided the whole set of data into two data sets. The data from 1972 to 2015 were the training data set whereas the data from 2016 to 2018 were the testing data set. The ARIMA models were then applied to the time series data of each variable (tea production, consumption and export) through R studio [22]. Finally, a forecast was made for the period of ten years from 2019 to 2028.

The trends of tea production, consumption and export for Bangladesh are plotted in Figure 1 and covered 1972 to 2018. Here, the time-series data shows a short-term movement for production, consumption and export of tea.

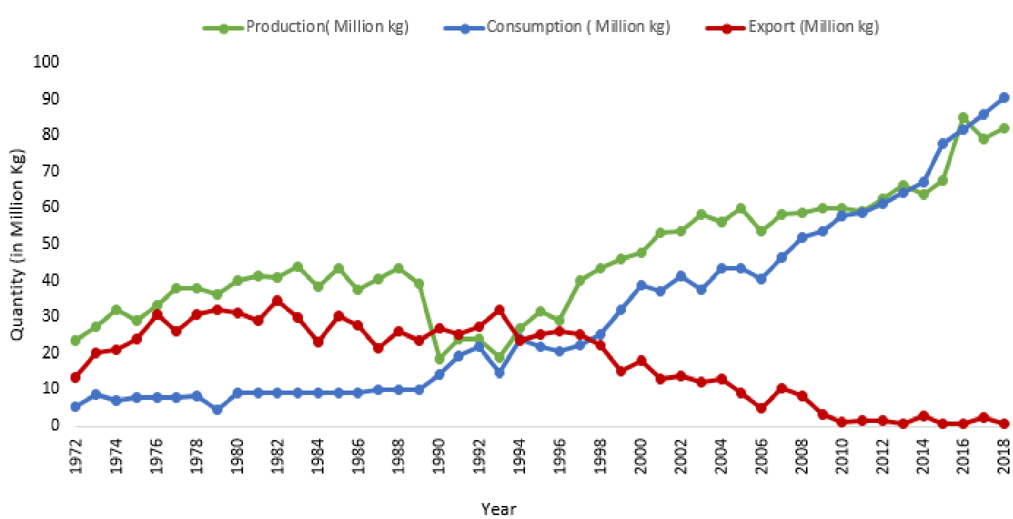

Figure 1. Production, consumption and export of tea in Bangladesh [6, 7]

From the production trends, the maximum production of tea was observed at 85.05 million kg in 2016 whereas the minimum production was 18.36 million kg in 1990. Again, the internal consumption of tea shows an upward trend over the years and the highest amount of tea consumption was 90.45 million kg in 2018. Due to population growth, a high rate of drinking habit, higher income, health awareness, and some aspects of nutritional benefit of tea possibly caused the increase in domestic demand for tea in Bangladesh [6].

Additionally, the minimum amount of tea export was 0.47 million kg in 2016 whereas the highest amount of tea export was observed at 34.42 million kg in 1982. The export of tea shows a declining trend, which was possibly due to increased internal consumption, lack of diversity in the quality of tea, stiff competition in the international markets and total domestic supply dependent only on two to three districts [23].

Furthermore, the domestic price of tea for the study period (1997- 2017) showed a slight increase compared to export prices for the same period, which may be another reason for the decrease in the volume of tea exports (Figure 2).

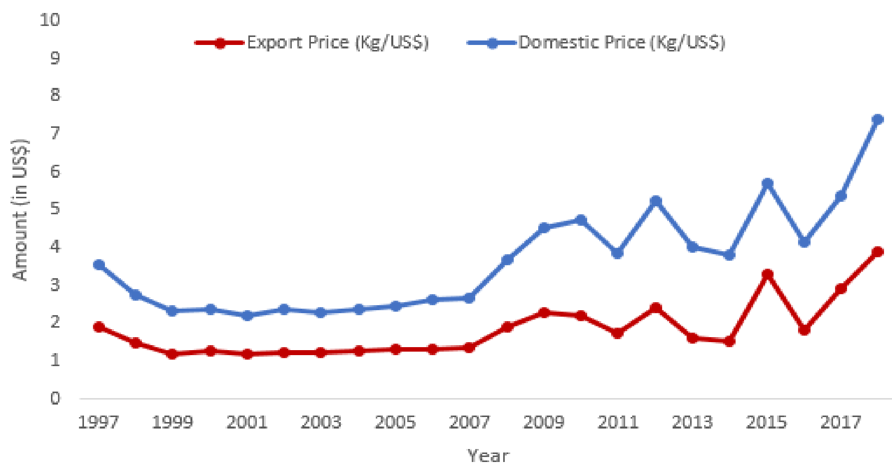

Figure 2. Export price and domestic price of tea in Bangladesh [7]
(Data are taken as nominal values)

2.2 Overall description of the model

The Box Jenkins Approach: The Box Jenkins approach is widely used because it can handle any stationary or non-stationary data series with or without seasonal elements [24]. For building ARIMA model, the study used five steps of the Box Jenkins method. These were test of stationarity, identification, parameter estimation, diagnostic checking and finally forecasting from the best-fitted model [25].

In this study, we used time-series yearly data on production, internal consumption and export (million kg) of tea in Bangladesh. We used non-seasonal ARIMA (p, d, q) models because the data have no seasonality among the variables. So, the equation for the simplest ARIMA model is as follows [24]:

$$Y_t = \delta_0 + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \delta_p Y_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots + \varphi_q \varepsilon_{t-q} \quad (1)$$

Here, Y_t stands for production, consumption and export value of tea with reference to time. This equation has three components. First, AR (p) means auto regression of order p that depends on its own past values. It can be written as

$$Y_t = \delta_0 + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \delta_p Y_{t-p} + \varepsilon_t \tag{2}$$

Second, I denotes integrated (d) indicating differences of the original data series that allows the time series to turn into stationary. When we take difference in time series ARIMA model, the forecasting equation can be written as follows:

No difference (d=0): $y_t = Y_t$
First difference (d=1): $y_t = Y_t - Y_{t-1}$
Second difference (d=2): $y_t = (Y_t - Y_{t-1}) - (Y_t - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

Here, Y_t denotes the original series and y_t denotes the differenced series. Finally, MA (q) is the moving average of order q that combines the dependency between an observation and a random error term. The model can be stated as

$$Y_t = \delta_0 + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots + \varphi_q \varepsilon_{t-q} \tag{3}$$

Where, Y_t= Response (dependent) variable at time t
$Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$= Response variable at time lags $t-1, t-2, \ldots, t-p$, respectively.
δ_0 = Constant or Intercept
$\delta_1, \delta_2, \cdots, \delta_p$ = Coefficients of each parameter p
$\varphi_1, \varphi_2, \cdots, \varphi_q$ = Coefficients of each parameter q
$\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ = Errors in previous time periods that are incorporated in the response Y_t
ε_t = Error term at time t

2.2.1 Test of Stationarity

Before identification of the model, the Augmented Dickey-Fuller (ADF) test can be used to check the data to see if there is any unit root or not [26]. Different ADF tests can be performed depending on the data generating process such as pure random walk, random walk with drift and random walk with drift plus linear trend. The equations are as follows:

$Y_t = Y_{t-1} + \varepsilon_t$; Pure random walk
$Y_t = \theta_0 + Y_{t-1} + \varepsilon_t$; Random walk with drift
$Y_t = \theta_0 + \theta_1 t + Y_{t-1} + \varepsilon_t$; Random walk with drift and deterministic trend

Here, pure random walk predicts the value of Y at time t is equal to the last year value $(t-1)$ plus a white noise error term (ε_t). Likewise, random walk with drift combines pure random walk plus a drift parameter (θ_0) whereas random walk with drift and deterministic trend combines a random walk with a drift component (θ_0) and a deterministic trend $(\theta_1 t)$.

For the ADF unit root test, the null hypothesis is that the data has a unit root whereas for an alternative hypothesis, there is no unit root in the data.

Null hypothesis: H_o = There is a unit root in the data

Alternative hypothesis: H_1 = There is no unit root in the data

If p-value of observed t-statistic is less than the nominated significance level then the null hypothesis will be rejected, i.e. there is no unit root [26].

The KPSS test, named after Kwiatkowski, D., Phillips, PCB., Schmidt, P. and Shin, Y., (1992) can be performed to assess whether the data series is stationary or not [27]. For the KPSS test, if the p-value is less than 0.05, then the null hypothesis will be rejected, i.e., the data is non-stationary [28, 29].

2.2.2 Identification of the model

Following the Box Jenkins approach, at first it is required to ascertain the correct values of p, d, and q for building the model. The study observed the graph of ACF (auto correlation function) and PACF (partial auto correlation function) to identify the order of moving average (MA) and auto regressive (AR) terms. Autoregression of p^{th} order can be derived from the partial autocorrelation function whereas the moving average of q^{th} order can be obtained from the autocorrelation function [24]. The order of the difference expressed as d is the value of the parameter in the model. Hence, for stationary data, the value of d = 0 and ARIMA (p, d, q) can be written as ARIMA (p, q) [30, 31].

2.2.3 Estimation of the model

For estimating the best-fitted parameters of the model, the following diagnostic checking is required: **Akaike's Information Criterion:** Akaike's Information Criterion (AIC), proposed by the statistician Hirotugu Akaike, is used to select the best possible model with minimum AIC value from a set of competing models [30]. The AIC is given by

$$\text{AIC} = \text{n} \log (\text{MSE}) + 2\text{h}$$

where n is the sample size, MSE is the mean square error and h is the total number of estimated parameters.

Bayesian Information Criterion: Again the best-fitted model can be selected through the minimum value of the Bayesian Information Criterion (BIC) that was developed by Gideon E. Schwarz [31]. The BIC is defined as

$$\text{BIC} = \text{n} \log (\text{MSE}) + \text{h} \log \text{n}$$

where n is the sample size, MSE is the mean square error and h is the total number of estimated parameters.

Moreover, AICc is also used when AIC will overfit because of small sample size [32]. AICc is AIC with a correction for small sample sizes, and is calculated as follows:

$$\text{AICc} = \text{AIC} + \frac{2\text{h}^2+2\text{h}}{\text{n}-\text{h}-1}$$

where n denotes the sample size and h denotes the number of parameters [33]. Thus, the model with minimum Akaike's Information Criterion, Akaike's Information Criterion with a correction and Bayesian Information Criterion was the best-fitting model of the study.

2.2.4 Diagnostic checking

After selecting the suitable model, it is required to do diagnostic checking of the model. It can be carried out by examining the autocorrelation plots of the residuals of the selected model. If all the ACF and PACF are small, the model is considered satisfactory and forecasts can be generated. Thus, the residuals are tested to find out if they are white noise or not.

The Ljung-Box test statistics: Again, for checking the absence of serial autocorrelation of the fitted model, the Ljung-Box (1978) statistics can be used. It is a modification of the Box-Pierce test [34]. The Ljung-Box test can be written as

$$Q = n\,(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

Where, $\hat{\rho}$= the residual autocorrelation at lag k,

n= the number of residuals,

h = the number of time lags being tested.

If the calculated p-value is higher than 0.05, the model suggests that there is strong autocorrelation of the residuals. The study then develops a new model until the model proves that there is no autocorrelation among the residuals, i.e. the calculated p-value is less than 0.05.

Mcleod-Li test: The McLeod-Li test is used to assess the conditional heteroscedasticity of the data. The null hypothesis of the McLeod-Li test is that there is no autoregressive conditional heteroscedasticity (ARCH) among the lags considered [35]. Hence,

Null hypothesis: H_0 = There is no heteroscedasticity among the residuals.

Alternative hypothesis: H_1 = There is heteroscedasticity among the residuals.

Therefore, the best model will satisfy the Gauss-Markov properties, i.e. residuals have zero mean, distinct error terms are homoscedastic, and they are uncorrelated [36].

2.2.5 Forecasting

Finally, after checking the diagnostics step, the model can be used to forecast the data.

Evaluation of model accuracy: The mean absolute percentage error (MAPE) is used for measuring the prediction accuracy of a forecasting model [37]. MAPE can be defined as

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where, A_t = the actual value and F_t = the forecast value.

3. Results and Discussion

For building ARIMA (autoregressive integrated moving average) model in this study, the data followed a random walk model because a slow steady change was observed in the chronological data of tea production, consumption and export. The last 47 years of data were considered for model building whereas the last 3 years of data were used for evaluating the prediction accuracy of the model.

3.1 Production

3.1.1 Stationarity test

Here the data generating process shows random walk with drift as the value of current period is equal to the value of last period with a drift and a white noise error term.

At first, the Augmented Dickey-Fuller test was used to check the unit root. Again, the KPSS test was used for checking the stationarity of the time-series data. The ADF and KPSS tests have opposite null hypothesis statement, i.e. the null hypothesis is non-stationary for ADF but it is stationary for KPSS.

The ADF test result showed that for original tea production data, the p-value (0.78) was higher than 0.05 at 5% level of significance indicating unit root, as shown in Table 1. After taking the first difference, the calculated p-value (0.01) was smaller than the tabulated p-value (0.05), showing no unit root. Again, the KPSS test showed that the p-value was less than 0.05 for the original series and higher than 0.05 after the first difference, indicating that the data became stationary at 5% level of significance.

Table 1. ADF and KPSS tests of tea production

Unit Root and Stationarity Test	Series	Test Statistics	p-value
ADF Unit Root Test	Original Series	-0.13	0.78
	First Differenced Series	-4.19	0.01
KPSS Test	Original Series	0.95	0.01
	First Differenced Series	0.15	0.10

3.1.2 Identifying the model

In order to identify any unusual observations, the ACF and PACF of the time series data of tea production are plotted in Figures 3 and 4. Here, the ACF and PACF plots are used to identify the order of MA (q) and the order of AR (p) process, respectively.

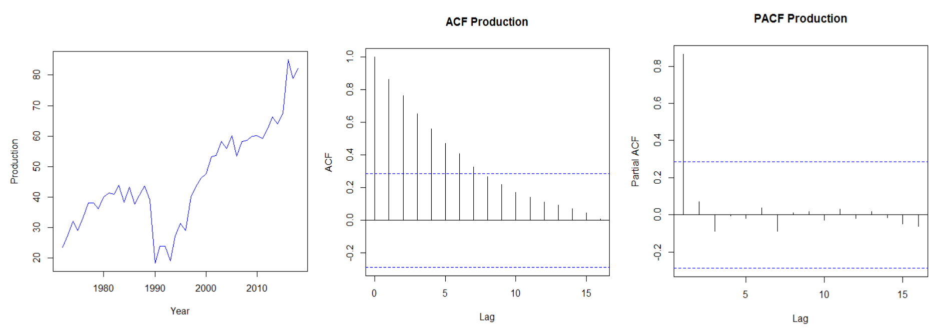

Figure 3. Original series of tea production

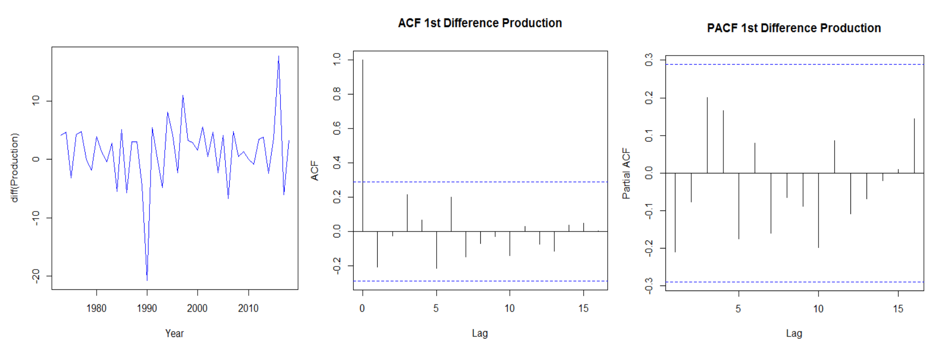

Figure 4. First differenced series of tea production

From Figure 3 it is difficult to say which model actually it is as here ACF shows exponential decaying behavior and decreases sufficiently quickly towards zero; and that's why we have to select the order of differencing d as the first order to know about the model. In PACF we see only one spike which is statistically significant at lag 1.

It has been clear from Figure 4 that the ACF shows one significant spike but the autocorrelation dies off gradually. In addition, no significant spike is observed from PACF, indicating AR (0) and MA (0) process. The tentative model is (0, 1, 0) since the series is integrated of order one.

3.1.3 Fitting the ARIMA model

For tea production, the selected models are ARIMA (0, 1, 0), ARIMA (0, 1, 1), ARIMA (1, 1, 0), ARIMA (1, 1, 1) and ARIMA (0, 1, 2). We found ARIMA (0, 1, 0) is the best-fitting model for tea production among the chosen models because it had the lowest AIC, BIC and AICc values, as shown in Table 2.

Table 2. Selected ARIMA models for tea production with AIC, AICc and, BIC values

Variables	Models	AIC Values	AIC_c Values	BIC Values
Tea Production	**ARIMA (0, 1, 0) with drift**	**292.64**	**292.92**	**296.29**
	ARIMA (0, 1, 1) with drift	292.56	293.13	298.04
	ARIMA (1, 1, 0) with drift	292.68	293.16	298.07
	ARIMA (1, 1, 1) with drift	294.50	295.47	301.81
	ARIMA (0, 1, 2) with drift	294.30	295.27	301.61

3.1.4 Residual analysis

In Figure 5, the residuals of the autocorrelation function and partial autocorrelation function of tea production are observed within the boundary of the permissible limit (95%), indicating that there is no significant correlation among the residuals of the fitted ARIMA model.

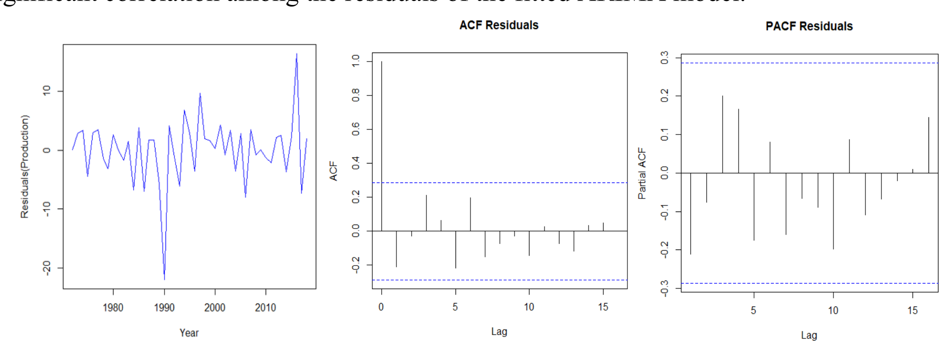

Figure 5. Time series plot, ACF and PACF of the residuals for tea production

In addition, the Ljung-Box test was also used to examine the presence or absence of autocorrelations in the residuals. From Table 3, it is observed that the p-value (0.49) is higher than 0.05, which points to a white noise error in the residuals of the fitted model. It also indicates that there is no autocorrelation among the residuals. Therefore, ARIMA (0, 1, 0) was the best-fitted model for forecasting tea production at 5% level of significance.

Table 3. Residual test of tea production

Variable	Ljung-Box Test	Value
Tea Production	χ-squared Statistic	7.45
	p-value	0.49

In Figure 6, the p-values of the McLeod-Li test statistic suggest the acceptance of the null hypothesis, i.e. p-value is greater than 0.05 at 5% level of significance. Thus, there is no ARCH effect in the residuals which indicates that the selected model is adequate in modelling the series. Consequently, the best fitting model also satisfies the Gauss-Markov properties (i.e. the model has zero mean, no autocorrelation and no heteroscedasticity in the residuals).

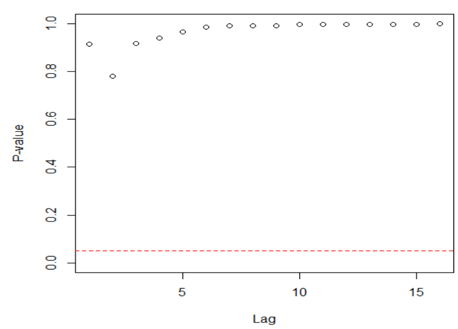

Figure 6. p-values of MCLeod-Li test

3.1.5 Forecasting

The accuracy measurement for tea production is presented in Table 4. Here, the projected values were higher than the real values in 2017. However, in 2016 and 2018 the projected values were lower than the real values.

Table 4. The accuracy measurement of tea production using the selected ARIMA model

Variables	Years of Projection	Real Values	Predicted Values	Percentage Error (±)	Mean Absolute Percent Error
Production	2016	85.05	68.65	19.28	10.31
	2017	78.95	86.32	-9.34	
	2018	82.13	80.22	2.33	

From Table 4, the forecasted values of tea production for the years 2016, 2017, and 2018 were 68.65, 86.32 and 80.22 million kg with the deviations of 19.28%, 9.34%, and 2.33%, respectively. It is clear from the mean absolute percent error (MAPE) that tea production is predicted with roughly 89.69% of accuracy.

Table 5 shows the forecasted values of tea production from 2019 to 2028. It is observed that tea production is increasing over the years, being 83.40 million kg in 2019, and will go up to 94.88 million kg in 2028.

Table 5. Forecasted values of tea production in Bangladesh at 95% confidence interval

Year of Prediction	Forecasted Values of Tea Production (million kg)		
	Forecasted Values	Lower Confidence Limit	Upper Confidence Limit
2019	83.40	72.36	94.45
2020	84.68	69.05	100.31
2021	85.95	66.82	105.09
2022	87.23	65.13	109.33
2023	88.50	63.80	113.21
2024	89.78	62.72	116.84
2025	91.05	61.82	120.29
2026	92.33	61.08	123.58
2027	93.60	60.46	126.75
2028	94.88	59.94	129.82

3.2 Consumption

3.2.1 Stationarity test

The consumption data series imply pure random walk process because there are no drift parameters. For tea consumption data, we also used the same ADF and KPSS test (Table 6) to check the unit root and stationarity.

Table 6. ADF and KPSS test of tea consumption

Unit Root and Stationarity Test	Series	Test Statistics	p-value
ADF Unit Root Test	Original Series	-0.04	0.99
	First Differenced Series	-4.29	0.78
	Second Differenced Series	-5.81	0.01
KPSS Test	Original Series	1.18	0.01
	First Differenced Series	0.68	0.01
	Second Differenced Series	0.06	0.10

For the data above, the findings of the ADF test indicated that there is unit root in both the original and the first differenced data series whereas the data has no unit root after the second difference at 5% level of significance, i.e. the p-value (0.01) is less than 0.05 after taking second difference. In addition, the KPSS test for tea consumption revealed that the p-value is less than 0.05 for both original and first difference but it is greater than 0.05 after second difference, which indicates stationarity.

3.2.2 Identifying the model

From Figure 7, it is clear that ACF shows exponential decaying behavior and decreases sufficiently towards zero and that is why we have to take differencing d twice to know about the model. In the PACF, only one spike is statistically significant at lag 1.

Figure 8 dictates that both the ACF and PACF plots showed two significant spikes significant indicating AR (0) and MA (2) process. The tentative model for tea consumption is (0, 2, 2) as the series is integrated of order two and having almost same pattern.

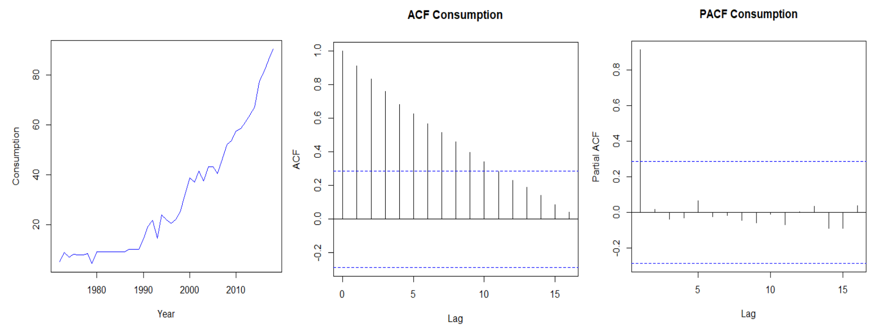

Figure 7. Original series of tea consumption

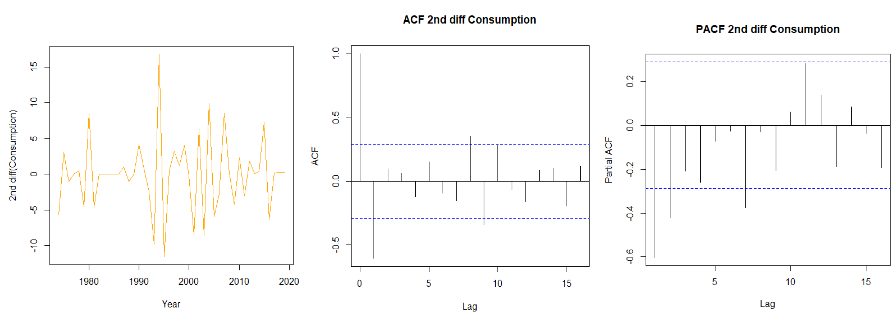

Figure 8. Second differenced series of tea consumption

3.2.3 Fitting ARIMA model

ARIMA (0, 2, 2), ARIMA (1, 2, 2), ARIMA (1, 2, 1), ARIMA (2, 2, 2) and ARIMA (3, 2, 2) were the chosen models for tea consumption. Here, we found the best model for tea consumption was ARIMA (0, 2, 2) based on the least values of AIC, AICc and BIC without considering the other factors of tea consumption (Table 7).

Table 7. Selected ARIMA models for tea consumption with AIC, AICc and BIC values

Variables	Models	AIC Values	AIC$_c$ Values	BIC Values
Consumption	**ARIMA (0, 2, 2)**	**244.60**	**245.19**	**250.02**
	ARIMA (1, 2, 2)	245.82	246.82	253.05
	ARIMA (1, 2, 1)	245.02	245.61	250.44
	ARIMA (2, 2, 2)	248.56	250.10	257.60
	ARIMA (3, 2, 2)	250.56	252.77	261.40

3.2.4 Residual analysis

Figure 9 indicates the residual plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for tea consumption. It can be seen from the Figure that the residuals for both the ACF and PACF are observed within the boundary excepting ACF at lag 0 of the permissible limit for up to 12 lag orders. This means that the residuals have no autocorrelations.

Moreover, it is evident from Table 8 that the p-value of the Ljung-Box test for tea consumption is 0.92 (i.e. p-value is higher than 0.05), indicating that there is no autocorrelation among the residuals. Hence, there is a white noise error in the residuals after fitting the model. This suggests that ARIMA (0, 2, 2) model for tea consumption does not only have the smallest AIC, AICc and BIC values but also the better-behaved residuals at 5% level of significance.

The p-values of McLeod-Li test are illustrated in Figure 10. Most of the p-values of the McLeod-Li test statistic are greater than 0.05 at 5% level of significance, suggesting the acceptance of the null hypothesis. Therefore, the selected model is the best model in the series because it has zero mean, no autocorrelation and the residuals are homoscedastic which also satisfy the Gauss-Markov properties.

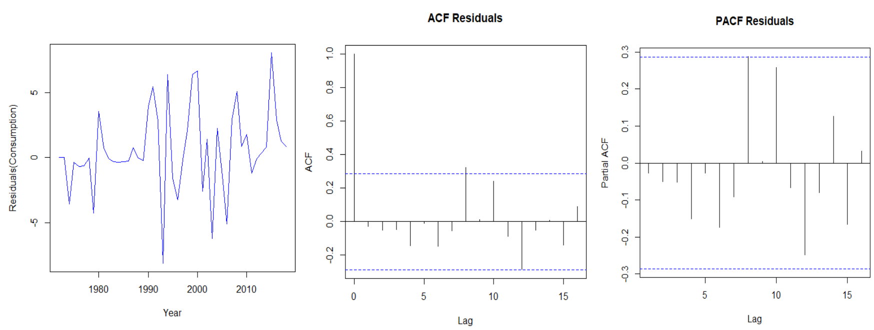

Figure 9. Time series plot, ACF and PACF of the residuals for tea consumption

Table 8. Residual test of tea consumption

Variable	Ljung-Box Test	Value
Tea Consumption	χ-squared Statistic	1.41
	p-value	0.92

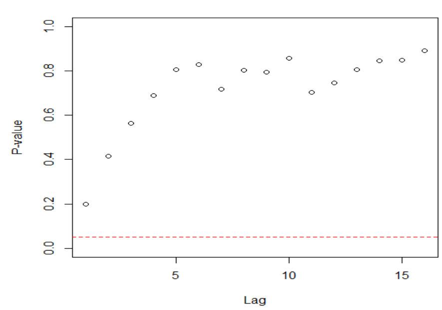

Figure 10. p-values of MC Leod-Li test

3.2.5 Forecasting

Table 9 represents the accuracy measurement values of tea consumption. Here, all the projected values were lower than the real values. During the years 2016-2018, the predicted values of tea consumption were 79.90, 85.19 and 89.82 million kg with deviations of 2.31%, 0.86%, and 0.70%, respectively. Moreover, the mean absolute percent error (MAPE) of tea consumption was projected with approximately 98.77% of accuracy.

The forecasted values of tea consumption from 2019 to 2028 are presented in Table 10. Here, tea consumption showed an upward trend, i.e. 94.35 million kg to 131.71 million kg over the years from 2019 to 2028.

Table 9. The accuracy measurement of tea consumption using the selected ARIMA model

Variables	Years of Projection	Real Values	Predicted Values	Percentage Error (±)	Mean Absolute Percent Error
Consumption	2016	81.64	79.90	2.13	1.23
	2017	85.93	85.19	0.86	
	2018	90.45	89.82	0.70	

Table 10. Forecasted values of tea consumption in Bangladesh at 95% confidence interval

Year of Prediction	Forecasted Values of Tea Consumption (million kg)		
	Forecasted Values	Lower Confidence Limit	Upper Confidence Limit
2019	94.35	87.62	101.08
2020	98.50	89.77	107.23
2021	102.65	91.84	113.46
2022	106.80	93.82	119.79
2023	110.95	95.69	126.21
2024	115.11	97.48	132.73
2025	119.26	99.17	139.34
2026	123.41	100.76	146.05
2027	127.56	102.26	152.85
2028	131.71	103.67	159.74

3.3 Export

3.3.1 Stationarity test

The data series of tea export also indicates pure random walk process as there is no constant or drift parameters. A similar test was also used for tea export data to check both the unit root and stationarity as presented in Table 11 and Figure 11.

The ADF test of tea export data showed that the p-value (0.14 and 0.06) was greater than 0.05, indicating the data had unit roots for both the original and first differenced series. After taking second difference, there is no unit root in the data at 5% level of significance, i.e. the p-value (0.01) is less than 0.05. Again, the KPSS test for tea export showed that the data was non stationary (i.e. the p-value is less than 0.05) for both the original and first differenced series and became stationary after second difference (i.e. the p-value is greater than 0.05).

Table 11. ADF and KPSS test of tea export

Unit Root and Stationarity Test	Series	Test Statistics	p-value
ADF Unit Root Test	Original Series	-3.09	0.14
	First Differenced Series	-3.45	0.06
	Second Differenced Series	-7.57	0.01
KPSS Test	Original Series	1.02	0.01
	First Differenced Series	0.42	0.04
	Second Differenced Series	0.08	0.10

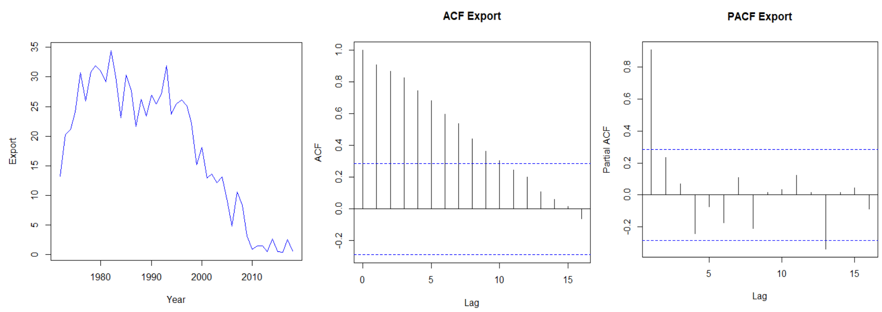

Figure 11. Original series of tea export

3.3.2 Identifying the model

Here, the ACF shows good exponential behavior and the PACF shows only one spike significant at lag 1. So, it's an AR (1) model primarily, but from the ACF, as the autocorrelation decreases sufficiently quickly towards zero, we have to proceed on differencing procedure.

From Figure 12, the ACF indicates good exponential decay and a damped sinewave pattern. Spikes are significant at lag 0,1, 4 and 7. In addition, the PACF shows significant spikes at lag 1, 3 and 12. As both ACF and PACF have the same pattern, this is an ARIMA process. The tentative model is (1,1,2) since the plots are from the first differenced series.

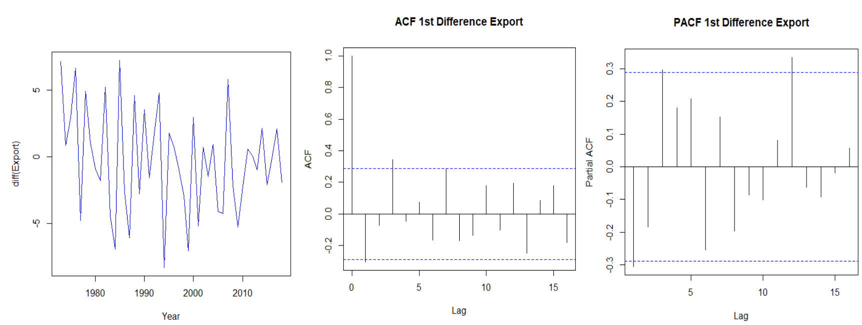

Figure 12. First differenced series of tea export

3.3.3 Fitting the ARIMA model

The selected models for tea export were ARIMA (1, 1, 2), ARIMA (2, 1, 2), ARIMA (1, 1, 3), ARIMA (2, 1, 3) and ARIMA (3, 1, 3). The best-fitted model was ARIMA (1, 1, 2) as the AIC, AICc and BIC showed the lowest values among the other selected models (Table 12) of tea export.

Table 12. Selected ARIMA models for tea export with AIC, AICc and, BIC values

Variables	Models	AIC Values	AIC_c Values	BIC Values
Export	**ARIMA (1, 1, 2)**	**248.44**	**249.41**	**255.75**
	ARIMA (2, 1, 2)	248.91	250.41	258.06
	ARIMA (1, 1, 3)	248.29	249.79	257.43
	ARIMA (2, 1, 3)	247.39	249.54	258.36
	ARIMA (3, 1, 3)	249.26	252.21	262.06

3.3.4 Residual analysis

The plot of the standardized residuals (Figure 13) of ACF and PACF reveals that all the spikes are in the 95% limits for up to 8 lags which implies no autocorrelations.

In addition, the p-value (0.68) of the Ljung-Box test in Table 13 for tea export is higher than 0.05 which completely proposes no autocorrelation in the residuals. Therefore, ARIMA (1, 1, 2) for tea export is the best-fitted model for accurately predicting the future values in the test set.

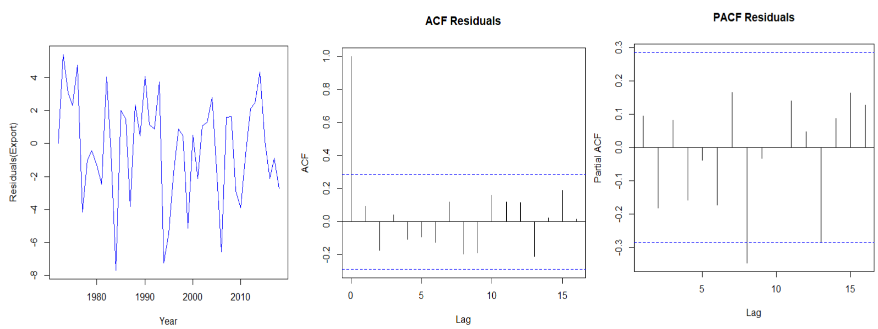

Figure 13. Time series plot, ACF and PACF of the residuals for tea export

Table 13. Ljung-Box test of the residuals of tea export

Variable	Ljung-Box Test	Value
Tea Export	χ-squared Statistic	3.11
	p-value	0.68

The McLeod-Li test illustrates the acceptance of the null hypothesis because most of p-values are greater than 0.05 at 5% level of significance (Figure 14). Therefore, there is no ARCH effect in the residuals which indicates that the selected model is adequate for modelling the series. Hence, the model proves that it has zero mean, no autocorrelation and no heteroscedasticity among the residuals, i.e. the model satisfies the Gauss-Markov properties.

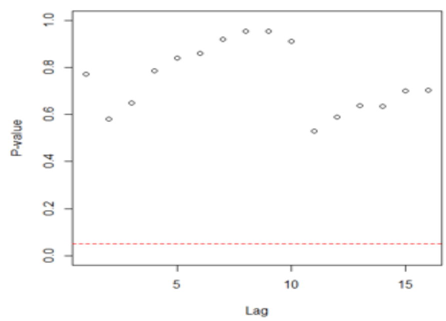

Figure 14. p-values of MC Leod-Li test

3.3.5 Forecasting

The actual and forecasted values of tea exports are presented in Table 14. Here, in 2016 and 2018 the predicted values were higher than the actual values, i.e. 0.72 and 0.95 million kg with a deviation of 53.19% and 46.15%, respectively. In 2017, the forecasted value of tea export was 1.25 which was lower than the actual value with 51.17% deviation. Furthermore, tea exports were predicted with approximately 49.83% of accuracy from the mean absolute percent error (MAPE).

Table 14. The accuracy measurement of tea exports using the selected ARIMA model

Variables	Years of Projection	Real Values	Predicted Values	Percentage Error (±)	Mean Absolute Percent Error
Export	2016	0.47	0.72	-53.19	50.17
	2017	2.56	1.25	51.17	
	2018	0.65	0.95	-46.15	

Table 15 displays the forecasted values of tea exports during the years 2019 to 2028. It is evident from the Table that the tea exports are decreasing gradually over the years. It is also observed from the original data series that the tea exports are gradually declining in Bangladesh [7]. The decreasing trend in exports is mainly caused by increasing domestic demand for tea during this period. These trends imply that, in order to achieve sustainable growth of tea in the future, an accelerated effort to increase tea production is necessary and the agricultural policies of the country should be directed towards achieving this objective.

Table 15. Forecasted values of tea exports in Bangladesh at 95% confidence interval

Year of Prediction	Forecasted Values of tea export (million kg)		
	Forecasted Values	Lower Confidence Limit	Upper Confidence Limit
2019	2.28	-4.22	8.78
2020	1.05	-5.81	7.92
2021	1.03	-8.20	8.58
2022	0.97	-11.13	10.29
2023	0.85	-14.19	12.50
2024	0.63	-17.16	14.87
2025	0.35	-19.97	17.26
2026	0.27	-22.61	19.60
2027	0.19	-25.07	21.85
2028	0.07	-27.37	24.01

4. Conclusions

Tea, being a very popular beverage, has significant potential for the economy of Bangladesh. Considering the economic importance of tea, modelling and forecasting of tea production, consumption and export has been carried out in the study using ARIMA modelling. The aim is to assist different actors (producers, government and policymakers) to undertake the necessary actions to develop the tea industry in Bangladesh in a sustainable way. Based on the forecasting and validation of the results, the production and consumption patterns of tea showed an increasing trend except for the export of tea, which will decrease further from its 2019 level. Though tea exports have declined over the time and will continue to decline as per the forecasted result of this study, the amount of export earning had not reduced substantially over the previous years. Recalling the export earnings from tea cultivation described in introduction section, it can be concluded that tea still has the potential as a cash crop that can contribute to the economic development of the country. Therefore, its export can be promoted in ways consistent with the sectoral and trade policies of the country. Furthermore, tea production and export should also be promoted through marketing and creating opportunities for export so that it will not only fulfil domestic demand but sufficient foreign currency to support the economic development of the country can also be earned.

Even though the ARIMA model used in this study can adequately forecast the production, consumption, and export scenarios using the available secondary data, some other factors (such as quality, certification, production technology, climate change, land use, macroeconomic, trade and policy variables and so forth) can possibly influence such predictions. Therefore, the government should also consider such factors when making policy decisions for the development of the tea industry in Bangladesh. From the further research perspective, the use of ARIMAX model can better accommodate the impact of extraneous variables such as changing environmental condition or other factors on the forecasted values.

References

[1] Rahman, A., 2017. Modeling of tea production in Bangladesh using Autoregressive Integrated Moving Average (ARIMA) Model. *Journal of Applied & Computational Mathematics*, 6(2), 349-357.

[2] Hossain, M.M. and Abdulla, F., 2015. Forecasting the tea production of Bangladesh: Application of ARIMA model. *Jordan Journal of Mathematics and Statistics*, 8(3), 257-270.

[3] Nasir, T. and Shamsuddoha, M., 2011. Tea productions, consumptions and exports: Bangladesh perspective. *International Journal of Educational Research and Technology*, 2(1), 68-73.

[4] Wang, H., Wen, Y., Du, Y., Yan, X., Guo, H., Rycroft, J., Boon, N., Kovas, E. and Mela, D., 2010. Effects of catechin enriched green tea on body composition. *Obesity*, 18, 773-779.

[5] Dhekale, B.S., Sahu, P.K., Vishwajith, K.P., Mishra, P. and Noman, M., 2014. Modelling and forecasting of tea production in West Bengal. *Journal of Crop and Weed*, 10, 94-103.

[6] Ahammed, K.M., 2012. Investment for sustainable development of Bangladesh tea industry – An empirical study. [online] Available at: https://bea-bd.org/site/images/pdf/9.pdf.

[7] Bangladesh Tea Board, 2019. *Statistical Bulletin of Bangladesh Tea Board for the Month of November 2019*. [online] Available at : http://teaboard.portal.gov.bd.

[8] Milad, M., 2020. *Tea Brews Booming Business in Bangladesh*. [online] Available at: https://en.prothomalo.com/bangladesh/good-day-bangladesh/tea-brews-booming-business-in-bangladesh.

[9] Raza, S.M.S., 2019. Prospects and challenges of tea industry in Bangladesh. *The Cost and Management*, 47(2), 119-135.

[10] Bangladesh Bureau of Statistics (BBS), 2018. *Statistical Yearbook of Bangladesh 2018*. Ministry of Planning, Government of the People's Republic of Bangladesh.

[11] Baten, A., Kamil, A.A. and Haque M.A., 2002. Productive efficiency of tea industry: A stochastic frontier approach. *African Journal of Biotechnology*, 9, 3808-3816.

[12] Shilpi, F.J., 1990. *Estimating Income and Price Elasticities of Imports and Exports of Bangladesh*. *Research Report No. 122*. Dhaka: Bangladesh Institute of Development Studies.

[13] Das, N.C., 2006. Export demand function for Bangladesh's tea. *Bangladesh Journal of Agricultural Economics*, 29(1-2), 1-18.

[14] Sobuj, M.A., 2016. *Assessing Socio-economic Status of Tea Garden Workers in Sylhet District*. M.S. Thesis, Department of Agricultural Extension, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh.

[15] Islam, M.A., Sumy, M.S.A., Uddin, M.A. and Hossain, M.S., 2020. Fitting ARIMA model and forecasting for the tea production, and internal consumption of tea (per year) and export of tea. *International Journal of Material and Mathematical Sciences*, 2(1), 8-15.

[16] Muhammed, F., Siddique, M., Bashir, M. and Ahamed, S., 1992. Forecasting rice production in Pakistan using ARIMA models. *Journal of Animal and Plant Science*, 2, 27-31.

[17] Hossain, M.J., Samad, Q.A. and Ali, M.Z., 2006. ARIMA model and forecasting with three types of pulse prices in Bangladesh: a case study. *International Journal of Social Economics*, 33(3-4), 344-353.

[18] Awal, M.A. and Siddique, M.A.B., 2011. Rice production in Bangladesh employing by ARIMA model. *Bangladesh Journal of Agricultural Research*, 1, 51-62.

[19] Debnath, M.K., Bera, K. and Mishra, P., 2013. Forecasting area, production and yield of cotton in India using ARIMA model. *Research & Reviews: Journal of Space Science & Technology*, 2(1), 16-20.

[20] Prabakaran, K., Sivapragasam, C., Jeevapriya, C. and Narmatha, A., 2013. Forecasting cultivated areas and production of wheat in India using ARIMA model. *Golden Research Thoughts*, 3(3), 1-7.

[21] Box, G.E.P. and Jenkin, G.M., 1976. *Time Series of Analysis: Forecasting and Control*. Revised ed. San Francisco: Holden-Day.

[22] RStudio Team, 2020. *RStudio: Integrated Development for R*. [online] Availabel at: http://www.rstudio.com/.

[23] Shabbir, S.M. and Sa'adat, N., 2010. Expansion of tea production and export from Bangladesh: some policy suggestion. *Thoughts on Economics*, 16(3-4), 66-67.
[24] Anderson, T.W., 1971. *The Statistical Analysis of Time Series*. New York: John Wiley.
[25] Gujarati, D.N., and Porter, D.C., 2008. *Basic Econometrics*. 5th ed. Boston: McGraw Hill.
[26] Dickey, D.A. and Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(6), 427-431.
[27] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3), 159-178.
[28] Rahman, N.M.F. and Baten, M.A., 2016. Forecasting area and production of black gram pulse in Bangladesh using ARIMA models. *Pakistan Journal of Agricultural Science*, 53(4), 759-765.
[29] Makridakis, S.G., Wheelwright, S.C. and Hyndman, R.J., 1998. *Forecasting: Methods and Applications*. 3rd ed. New York: Wiley.
[30] Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
[31] Schwarz, G.E., 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
[32] Cavanaugh, J.E., 1997. Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, 31(2), 201-208.
[33] Burnham, K.P. and Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33, 261-304.
[34] Ljung, G.M. and Box, G.E.P., 1978. On a measure of a lack of fit in time series models. *Biometrika*, 65(2), 297-303.
[35] McLeod, A.I. and Li, W.K., 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4, 269-273.
[36] Walters, A.A., 1970. *An Introduction to Econometrics*. New York: W.W. Norton.
[37] Makridakis, S., 1993. Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4), 527-529.