Vol.7.Issue.4.2019 (Oct-Dec)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

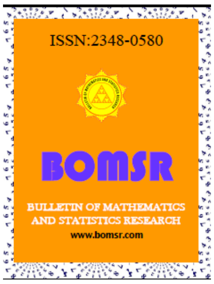
http://www.bomsr.com
Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

FORECASTING THE PRODUCTION OF SUGAR CANE BASED ON TIME SERIES MODELS IN BANGLADESH

MD. ZIAUL HASSAN[1*], MD. SIRAJ-UD-DOULAH[2] AND MD. KAMRUL HASAN[2]
[1*]Department of Statistics, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh.
[2]Department of Statistics Begum Rokeya University, Rangpur, Bangladesh.
DOI: 10.33329/bomsr.74.24



ABSTRACT

Around 70% of the world's sugar is produced from sugarcane. The production of sugarcane is fluctuated from year to year due to fluctuation of area under sugarcane cultivation. According to FAO, sugar requirement per capita/day is 29g and Bangladesh requires 1.0-1.2 million tons of sugar/year to meet the demand of domestic consumption. To meet the demand of domestic consumption of sugar, it is too much essential to estimate the production of sugar since sugar is produced mainly from sugarcane in Bangladesh which leads us to do this project. The main purpose of this project is to forecast the production of sugarcane in Bangladesh. Here, we applied different methods of time series and analyze the yearly sugarcane production data in Bangladesh over the period 1972-1973 to 2016-2017. We applied Box-Jenkin's methodology to identify the actual model based on different model selection criterion. In this research, we found that ARIMA (1, 2, 2) model is suitable for the forecasting the sugarcane productions in Bangladesh. Comparing between the original series and forecasted series we found that the production of sugarcane is slightly decreasing.

Key words: Sugarcane, Box-Jenkin's methodology, ARIMA model, Forecasting.

1. INTRODUCTION

In Bangladesh, per hectare yield is about 10 ton only where as it is about 70 t ha-1 in other sugar producing countries(BER, 2014).. Recovery per cent in Bangladesh was 7.0 to 8.5 only, but 8.5 to11.0 in other countries (EU, Brazil, Australia, Thailand and Mauritius) even the neighboring states of India and Pakistan were also higher substantially (Humbert, 1968). This low yield and recovery caused mainly for management factors in production level. Environmental factors like weather of Bangladesh and sunshine period are not favorable for high recoverable sugar varieties as the country lies in the subtropical zone but other high sugar producing countries lie in the temperate zone (BBS, 2015 & BSFIC, 2015). Although now sugar industry is in loosing concern but still the sugarcane crop is contributing 0.67 percent to the national GDP. The highest contribution comes from jiggery sector (0.39%) followed by sugar (0.15%), chewing and juice (0.04%), seed (0.03%), cattle feed and fuel (0.03%) and by-product (0.03%) respectively (Bangladesh Bank, 2014 & Ministry of Finance, 2015). It is still an insurance crop under rain fed condition as well as water logged situation in northern and western regions where most of the sugarcane is cultivated. During the last ten years (1999-00 to 2008-09) sugarcane was grown on an average 160.54 thousand (0.16 million) hectare of land producing 6564.91 thousand (6.5 million) ton of sugarcane (Miah, 1992 & Sheikh and Haque, 1986).

1.1 Objectives of the study

The main objective of this study is to forecast the future production of Sugarcane in Bangladesh. Here, the study time series analysis on exporting Sugarcane. On the basis of this analysis the study attempts to-

i) To fit accurate model for Sugarcane of Bangladesh.

ii) To check the model adequacy for different fitted model for Sugarcane of Bangladesh.

iii) To forecast future GDP obtain form Sugarcane of Bangladesh.

iv) To see the forecasting performances of the selected ARIMA models.

2. MATERIALS AND METHODS

One of the important types of data used in empirical analysis is time series data. The empirical work based on time series data assumes that the underlying time series is stationary. The time series analysis based on the stationary time series data. In this section we briefly discuss on stationary and non-stationary time series. A stochastic process is said to be stationary if its mean and variance are constant over time. Otherwise it will be non-stationary. Why are stationary time series so important? Because if a time series is non-stationary, we can study its behavior only for the time period under consideration. Each set of time series data will therefore be for a particular episode. As a consequence, it is not possible to generalize it to other time periods. Therefore, for the purpose of forecasting, such (non-stationary) time series may be of little practical value. How do we know that a particular time series is stationary? There are several tests of stationary. Here we used graphical and analytical recognized test. Graphical test: if we depend on common sense, it would seem that the time series depicted in figure is non-stationary, at least in the mean value. Here we applied most widely used popular formal test over the past several years are Autocorrelation function (ACF), Partial Auto-correlation function (PACF) and Augmented dickey-fuller test.

2.1 Auto correlation function (ACF)

Seasonal patterns of time series can be examined via correlograms. The correlogram (auto correlogram) displays graphically and numerically the autocorrelation function (ACF), that is, serial correlation coefficients (and their standard errors) for consecutive lags in a specified range of lags (e.g., 1 through 30). Ranges of two standard errors for each lag are usually marked in correlograms but typically the size of auto correlation is of more interest than its reliability because we are usually interested only in very strong autocorrelations (Box, et al., 1994). While examining correlograms one should keep in mind that autocorrelations for consecutive lags are formally dependent. Mathematically the auto-correlation function is,

$$\rho_k = \frac{\dfrac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n}}{\dfrac{\sum (Y_t - \bar{Y})}{n}}$$

2.2 Partial Auto-correlation function (PACF)

Another useful method to examine serial dependencies is to examine the partial autocorrelation function (PACF) - an extension of autocorrelation, where the dependence on the intermediate elements (those within the lag) is removed (Brockwell, et al., 2002). Mathematically the partial auto-correlation function is,

$$\phi_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^{p} \phi_{pj} r_{p+1-j}}{1 - \sum \phi_{pj} r_j}$$

Where, $\phi_{p+1,j} = \phi_{pj} - \phi_{p+1,p+1}\phi_{p,p-j+1}$

(iii) Augmented Dickey-fuller test

For theoretical and practical reasons the Dickey Fuller test (Box, et al., 1994) is applied to regression model of the following form:

$$\Delta y_t = \delta y_{t-1} + u_t \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (a)$$

$$\Delta y_t = \beta_1 + \delta y_{t-1} + u_t \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (b)$$

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + u_t \dots\dots\dots\dots\dots\dots\dots (c)$$

Where t is the time or trend variable. In each case the null hypothesis is that$\delta = 0$, that is there is a unit root. The difference between (a) and the other two regression lines are in the inclusion of the constant (intercept) and the trend term.

If the error term u_t is auto correlated, one modified (c) as follows:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta y_{t-1} + \varepsilon_t \dots\dots\dots\dots\dots\dots (d)$$

Where $for\ example, \Delta y_{t-1} = (y_{t-1} - y_{t-2}), \Delta y_{t-2} = (y_{t-2} - y_{t-3})\ etc.$ i.e. one uses lag difference terms.

The null hypothesis is still that $\delta = 0$ or$\delta = 1$, i.e., a unit root exists in y (i.e. is non-stationary) when the Dickey Fuller test is applied to models like (d), it is called augmented Dickey-Fuller (ADF) test. The ADF the test statistic has same asymptotic distribution as the DF statistic, so the same critical values can be used.

2.3 Model selection

Here, we applied Box-Jenkin's Methodology to identify the appropriate model for analyzing and forecasting the production of Aman rice in Bangladesh. The Box-Jenkin's Methodology (Box, et al., 1994) is shown below:

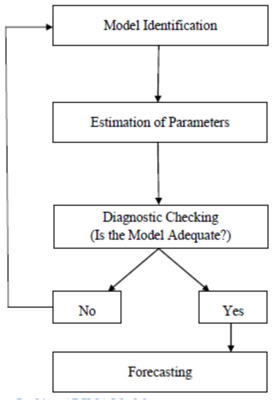


Figure 1: Box-Jenkin's Methodology

2.4 Forecasting

A planning tools that helps management in its attempts to cope with the uncertainty of the future, relying mainly on data from the past and present and analysis of trends. Predictions of future events and conditions are called forecasts, and the act of making such predictions is called forecasting (Brockwell, et al., 2002).

3. RESULTS AND DISCUSSION

We have used the secondary data regarding sugarcane production in Bangladesh which are collected over the period 1972-1973 to 2016-2017from Department, Bangladesh Bank. Here we applied different techniques for analyzing the data and the results of this data are shown below:

3.1 Identification of Sugarcane Production

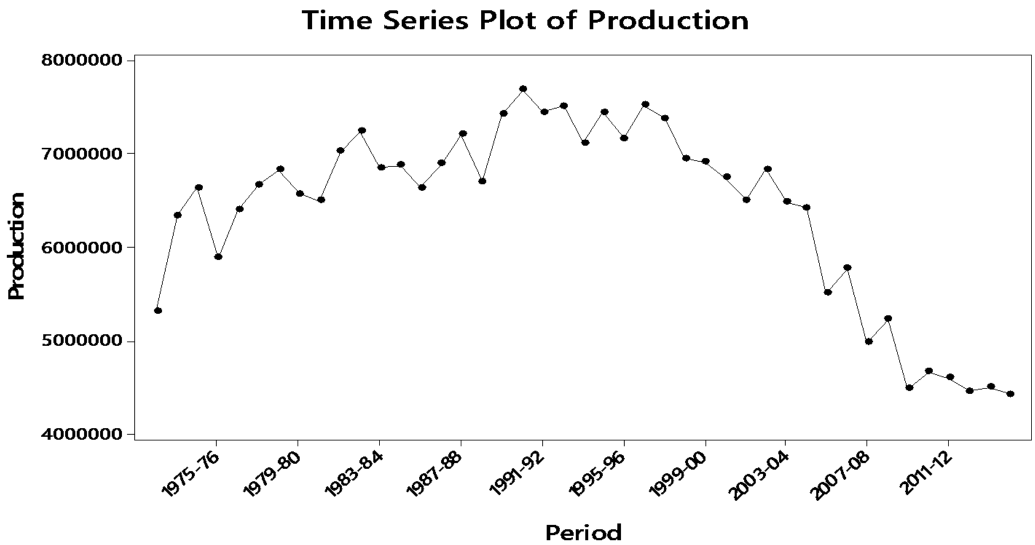


Figure 2: Time plot of sugarcane Production data (in tons)

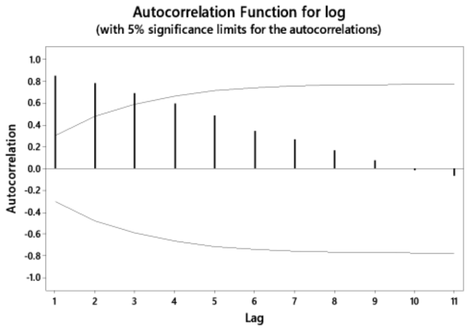


Figure 3: ACF for sugarcane Production

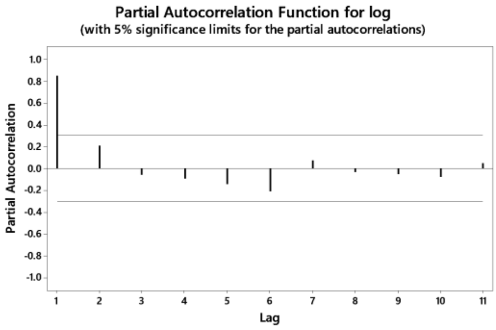


Figure 4: PACF for Sugarcane Production

From Figure 2 it is clear that the data must take account of the obvious increasing trend. Hence the original series is not stationary.From the Figure 3 and 4, we found that ACF and

PACF show a typical pattern of non-stationary series.

Table 1. Q-test and LB Test for Sugarcane Production.

Statistic	*Tabulated Value*	*Calculated Value*
Q-Statistic	15.98	33.655
LB Statistic	15.98	35.649

From Table 1 we observed that the series of sugarcane production is correlated. Therefore the series is non-stationary.

Now we perform Unit-root that means Augmented Dickey-Fuller (ADF) test to check the stationary of sugarcane Production data.

Table 2. ADF Test statistic

ADF Test statistic		*1*	*-3.600987*
0.532615	**Critical Value (%)**	5	-2.935001
		10	-2.605836

From Table 2, we found that 0.532615(ADF value) >-3.119910(Negative Critical Value) accept the null hypothesis. So, the series is non-stationary.

Transforms: Natural log, difference (1)

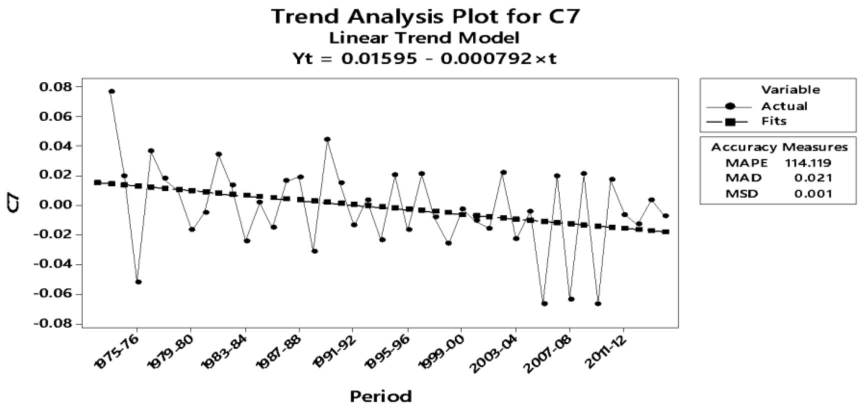


Figure 5: Sugarcane Production after first difference

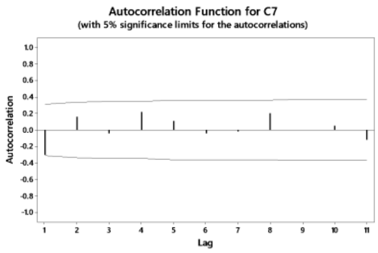


Figure 6: ACF of log for sugarcane Production after first difference.

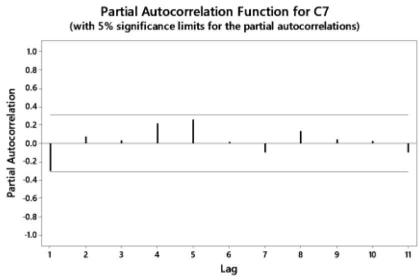


Figure 7: PACF of log for sugarcane Production after first difference.

From Figure 5, 6 & 7 we observed that the time series of the sugarcane production is non-stationary.

Table 3. Q-test and LB Test for Sugarcane Production after first difference

Statistic	*Tabulated Value*	*Calculated Value*
Q-statistic	15.98	4.0850
LB Statistic	15.98	5.817

From Table 3 we observed that the series of sugarcane production with first differenceis correlated. Therefore the series is stationary.

Table 4. ADF test statistic

ADF Test statistic		*1*	*-3.600987*
-9.505294	**Critical Value (%)**	5	-2.935001
		10	-2.605836

From Table 4 we observed that -9.505294(ADF value) <-3.119910(Negative Critical Value) reject the null hypothesis. So, the production of Aman rice is stationary.

Now we consider the different types of tentative models as much as possible from which we select the best model using the model selection criterion. Since the characteristics of a good ARIMA model is parsimonious ignoring the higher order of p and q, the tentative models on the basis of model selection criterion are as follows:

Table 5. Different ARIMA Models

Model	*SSE*	*AIC*	*NBIC*	*RMSE*
ARIMA(0,2,1)	0.0645	1182.8	1186.22	455407.5010
ARIMA(1,2,2)	**0.3175**	**1178.28**	**1185.13**	**464564.124**
ARIMA(2,2,2)	0.4853	1180.27	1188.83	462408.747

From the above Table 5, we see that for the model ARIMA (1, 2, 2); SSE, AIC, SIC are smaller than other models. So the model ARIMA (1, 2, 2) is the best tentative model and we use this model for our forecasting purposes.

3.2 Out of Sample Model Adequacy of Sugarcane Production

We will compare different model with each other by the following statistical measure of criterion.

Table 6. Statistical Summary Measures of Model Forecast

Model	*RMSE*	*MAE*	*MAPE*
ARIMA(0,2,1)	0.031	0.024	0.352
ARIMA(1,2,2)	**0.030**	**0.022**	**0.030**
ARIMA(2,2,2)	0.076	0.049	1.715

We observed from the above Table 6 that RMSE, MAE, MAPE, are smaller for ARIMA (1,2,2,) So we can conclude that the ARIMA (1, 2, 2) model is the best fitted model among all the tentative models.

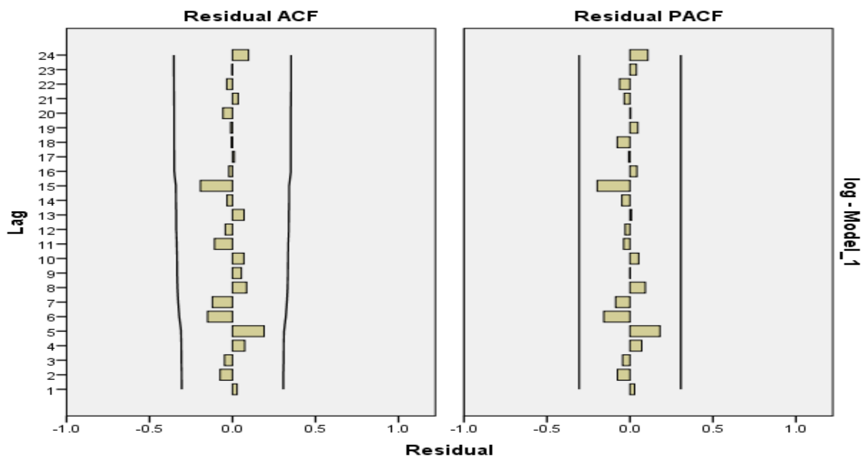


Figure 7: Residual ACF and Residual PACF for the Model ARIMA (1, 2, 2)

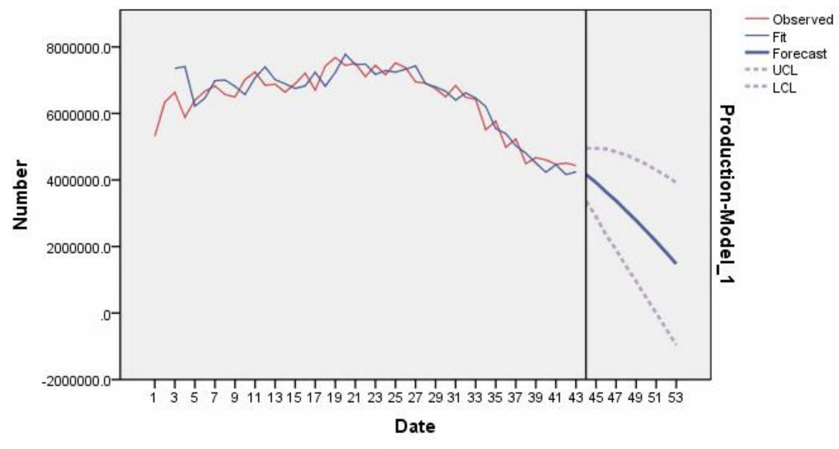


Figure 8: Plot for Forecasts for Sugarcane production for Bangladesh

From Figure 8 we observed that the forecasted values are gradually decreasing. Therefore, we may conclude that if our sugarcane production is stay present situation, in future we will lose our sugarcane production.

Table 7. Forecasted Value of Sugarcane Production Data using ARIMA (1,2,2,)

Period	Production	LCL	UCL
2015-16	4242732.7	3729746.9	4807378.3
2016-17	4069727.6	3459130.5	4758541.1
2017-18	3884664.1	3158321.8	4730783.4
2018-19	3710488.1	2933034.1	4634570.4
2019-20	3531574.9	2700287.6	4543262.9
2020-21	3360066.0	2503603.3	4421555.3
2021-22	3188036.3	2309330.4	4298818.8
2022-23	3021928.6	2136390.6	4160443.5
2023-24	2857802.9	1969759.3	4019718.5
2024-25	2699097.2	1817787.7	3870674.2
2025-26	2543902.0	1673423.9	3719803.5

From the above Table 7 we observed that if we go one step ahead, we get a good forecasted value. But as we goes multi-step ahead forecasting error decreases.

4. CONCLUSION

The best selected Box-Jenkins ARIMA model for forecasting the sugarcane productions in Bangladesh is ARIMA(1,2,2). From the comparison between the original series and forecasted series shows the same manner indicating fitted model are statistically well behaved to forecast sugarcane productions in Bangladesh i.e., the models forecast well during and beyond the estimation period which reached at a satisfactory level. Thus, this model can be used for policy purposes as far as forecasts the sugarcane production in Bangladesh.

REFERENCES

[1]. Brockwell Peter J., Richard A. Davis, 2002, Introduction to Time Series and Forecasting, 2nd ed. Springer-Verlang.

[2]. Box, G.E.P., Jenkins G.M., Reinsel G.C., 1994, Time Series Analysis, Forecasting and Control, 3rd edn. Prentice Hall, Englewood Clifs, NJ.

[3]. BER, 2014, Bangladesh Economic Review 2014, Ministry of Finance, Dhaka.

[4]. BBS, 2015, Statistical Yearbook of Bangladesh 2014. Bangladesh Bureau of Statistics, Statistics Division

[5]. Bangladesh Bank, 2014, Annual Report 2014. Bangladesh Bank, Dhaka.

[6]. BSFIC, 2015, MIS Report. Bangladesh Sugar and Food Industries Corporation. Motijheel Commercial Area, Dhaka, Bangladesh, 2008

[7]. Humbert R.P., 1968, The Growing of Sugar Cane. Elsevier publishing company, Amsterdam, 1968.

[8]. Ministry of Finance, 2015, Bangladesh Economic Review 2014. Finance Division.

[9]. Miah, M.A.M., 1992. *An economic study on sugarcane cultivation with several intercrops in Joypurhat Sugarmills areas*. An unpublished M. Sc. Thesis. Dept. of Agril. Economics, BAU, Mymensingh.

[10]. Sheikh M.N., Haque M.M., 1986, Production consumption and demand for sugar in Sugar Industry in Bangladesh, *BSFIC*, Shilpa Bhaban, Motijheel Commercial Area, Bangladesh.