Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040214

Forecasting the Sugarcane Production in Bangladesh by ARIMA Model

Md. Moyazzem Hossain[1,*] *and* Faruq Abdulla[2].

[1] Department of Statistics, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh
[2] Department of Statistics, Islamic University, Kushtia-7003, Bangladesh

Received: 29 Mar. 2015, Revised: 7 Jun. 2015, Accepted: 9 Jun. 2015
Published online: 1 Jul. 2015

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 tonnes 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 research. The main purpose of this research is to identify the Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the production of sugarcane in Bangladesh. This study considered the published secondary data of yearly sugarcane production in Bangladesh over the period 1971 to 2013. The best selected Box-Jenkins ARIMA model for forecasting the sugarcane productions in Bangladesh is ARIMA (0,2,1). 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 to a satisfactory level.

Keywords: Sugarcane, ARIMA Model, Forecasting, Bangladesh

1. Introduction:

Sugarcane (*Saccharum officinarum* L.) is an important sugar crops in the world. Sugarcane is grown around the world between tropical and sub-tropical climate (North latitude = 35^0 C and South latitude = 35^0 C). Sugarcane is cultivated in more than 100 countries of the tropical and sub-tropical regions of the world [1]. Different species likely originated in different locations, with *Saccharum barberi* originating in India and *S. edule* and *S. officinarum* in New Guinea [2]. It is theorized that sugarcane was first domesticated as a crop in New Guinea around 6000 BC [3]. New Guinean farmers and other early cultivators of sugarcane chewed the plant for its sweet juice. Early farmers in Southeast Asia, and elsewhere, may have also boiled the cane juice down to a viscous mass to facilitate transportation, but the earliest known production of crystalline sugar began in northern India. The exact date of the first cane sugar production is unclear. The earliest evidence of sugar production comes from ancient Sanskrit and Pali texts [4]. Around 70% of the world's sugar is produced from this crop [5]. Sugarcane is cultivated on an area of about 0.16 million hectare of land which almost 50% area is located in the sugar mills zone, and the remaining 50% is grown in the non-mills zone, where sugarcane is mostly diverted for jaggary and juice production. Presently, 15 sugar mills are in operation under Bangladesh Sugar and Food Industries Corporation (BSFIC) with a capacity of 0.21 million tons of sugar production per year [6]. Sugarcane is the second most important cash crop, which is grown in almost all districts of Bangladesh. It concentrates mainly in the greater districts of Rajshahi, Kushtia, Jessore, Rangpur, Dinajpur, Bogra, Pabna, Faridpur, Barisal, Dhaka, and Mymensingh [7]. Sugarcane is a long duration crop, which occupies the land for 10-14 months from planting to harvesting. Small and medium farmers, who are mainly sharecroppers in Bangladesh, cannot afford to wait for such a long period due to poor financial conditions as well as higher demand for food and vegetables for their family members. As a result, they tend to reduce cane cultivation and increase other corps [8].

*Corresponding author e-mail: mmhmm.justat@gmail.com

The production of sugarcane is fluctuated from year to year due to fluctuation of area under sugarcane cultivation [9]. Yield and production could not be increased to the desired level due to various bottlenecks in production and marketing of sugarcane. Sugar is an important energy-supplying substitute in our dietary allowance. Sugar is produced mainly from sugarcane. There are 16 sugar mills in Bangladesh providing a large amount of sugar of our requirements [10]. According to FAO, sugar requirement per capita/day is 29 g and Bangladesh requires 1.0 - 1.2 million tonnes of sugar/year to meet the demand of domestic consumption [11]. The figure will be larger in near future to keep pace with the country's demographic growth of 2.0 million/year and expected change in quality of life. Three sugar mills have been laid off and now there are 14 sugar mills operating in Bangladesh which have a total production capacity of 0.21 million tonnes/year. The sugar mills could not reach its present production capacity, as it does not get the required amount of sugarcane at the mill gates and the sugar recovery rate is also poor [12].

One of the main aims of the Millennium Development Goals (MDG) of Bangladesh by the year 2015 is to eradicate hunger, chronic food insecurity, and extreme destitution. 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 research. The main purpose of this research is to identify the Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the production of sugarcane in Bangladesh.

2 Materials and Methods

2.1 Data Source:

This study considered the published secondary data of yearly sugarcane production in Bangladesh which was collected over the period 1971 to 2013 from the Food and Agricultural Organization (FAO) website (http://faostat3.fao.org/).

2.2 ARIMA Model:

Suppose that $\{\zeta_t\}$ is a white noise with mean zero variance σ^2, then $\{Y_t\}$ is defined by $Y_t = \zeta_t + \beta_1\zeta_{t-1} + \beta_2\zeta_{t-2} + \ \dots + \beta_q\zeta_{t-q}$ is called a moving average process of order q and is denoted by $MA(q)$. If the process $\{Y_t\}$ is given by $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ \dots \ + \alpha_p Y_{t-p} + \zeta_t$ is called an auto-regressive process of order p and is denoted by $AR(p)$. Models that are combination of AR and MA models are known as $ARMA$ models. An $ARMA(p,q)$ model is defined as $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ \dots \ + \alpha_p Y_{t-p} + \zeta_t + \beta_1\zeta_{t-1} + \beta_2\zeta_{t-2} + \ \dots \ + \beta_q\zeta_{t-q}$, where, Y_t is the original series, for every t, we assume that ζ_t is independent of $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$. A time series $\{Y_t\}$ is said to follow an integrated autoregressive moving average (ARIMA) model if the d^{th} difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. If $\{W_t\}$ follows an $ARMA(p,q)$ model, we say that $\{Y_t\}$ is an $ARIMA(p,d,q)$ process. Fortunately, for practical purposes, we can usually take $d = 1$ or at most 2. Consider then an $ARIMA(p,1,q)$ process. With $W_t = Y_t - Y_{t-1}$, we have, $W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \ \dots \ + \alpha_p W_{t-p} + \zeta_t + \beta_1\zeta_{t-1} + \beta_2\zeta_{t-2} + \ \dots \ + \beta_q\zeta_{t-q}$.

2.3 Box-Jenkins Method:

The influential work of Box-Jenkins [13] shifted professional attention away from the stationary serially correlated deviations from deterministic trend paradigm toward the $ARIMA(p,d,q)$ paradigm. It is popular because it can handle any series, stationary or not with or without seasonal elements.

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

Preliminary Analysis: Create conditions such that the data at hand can be considered as the realization of a stationary stochastic process.

Identification of a Tentative Model: Specify the orders p, d, q of the ARIMA model so that it is clear the number of parameters to estimate. Empirical autocorrelation functions play an extremely important role to recognize the model.

Estimation of the Model: The next step is the estimation of the tentative ARIMA model identified in step-2. By maximum likelihood method we estimate the parameters of the model.

Diagnostic Checking: Check if the model is a good one using tests on the parameters and residuals of the model.

Forecasting: If the model passes the diagnostics step, then it can be used to interpret a phenomenon, forecast.

2.4 Ljung-Box Test:

Ljung-Box [14] test can be used to check autocorrelation among the residuals. If a model fit well, the residuals should not be correlated and the correlation should be small. In this case the null hypothesis is $H_0 : \rho_1(e) = \rho_2(e) = ... = \rho_k(e) = 0$ is tested with the Box-Ljung statistic $Q^* = N(N+1)\sum_{i=1}^{k}(N-k)\rho_k^2(e)$, where, N is the no of observation used to estimate the model. This statistic Q^* approximately follows the chi-square distribution with $(k-q)$ df, where q is the no of parameter should be estimated in the model. If Q^* is large (significantly large from zero), it is said that the residuals autocorrelation are as a set are significantly different from zero and random shocks of estimated model are probably auto-correlated. So one should then consider reformulating the model.

2.5 *Evaluation of Forecast Error*:

Before performing growth analysis it is necessary to estimate the growth model that best fits the time series. There are many summary statistics available in the literature for evaluating the forecast errors of any model, time series or econometric. We often do not compute all the statistics mainly because one of them is the function of the other. Thus, here an attempt is made to identify the best models for potato production in Bangladesh using the following contemporary model selection criteria, such as RMSPE, MPFE and TIC.

Root Mean Square Error Percentage (RMSPE): Root Mean Square Error Percentage (RMSPE) is defined as,

$$RMSPE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{Y_t^f - Y_t^a}{Y_t^a}\right)^2}$$

, where Y_t^f is the forecast value in time t and Y_t^a is the actual value in time t.

Minimum Phone Frame Error (MPFE): Minimum Phone Frame Error (MPFE) is defined as,

$$MPFE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{Y_t^a - Y_t^f}{Y_t^a}\right)$$

, where Y_t^a is the actual value in time t and Y_t^f is the forecast value in time t.

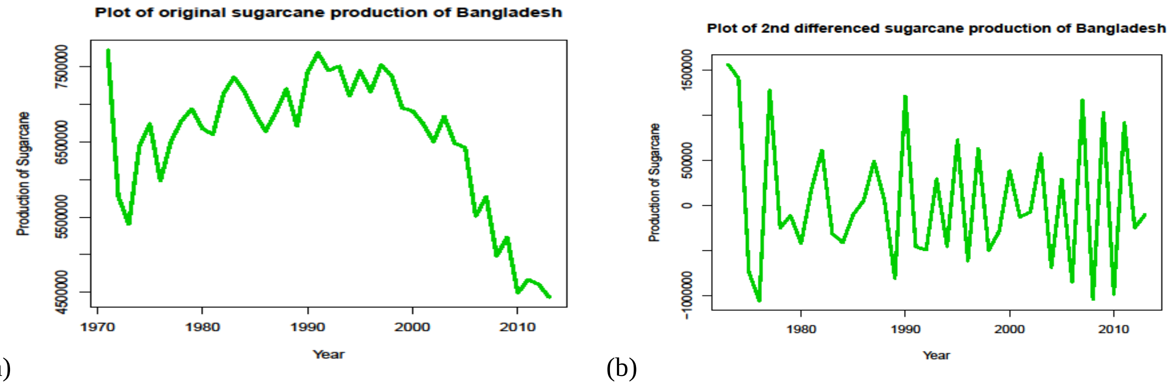
Theil Inequality Coefficient (TIC): Theil [15] Inequality Coefficient (TIC) is defined as

$$TIC = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^f - Y_t^a\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^f\right)^2}}$$

, where Y_t^f is the forecast value in time t and Y_t^a is the actual value in time t.

3 Results and Discussion

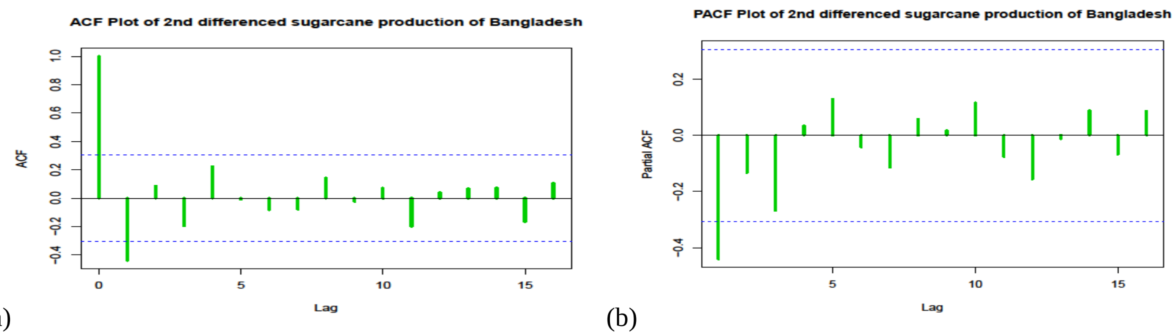
In this study Augmented-Dickey-Fuller (ADF) unit root test, Phillips-Perron (PP) unit root test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root test are used to check whether the data series is stationary or not. After second differencing the Augmented-Dickey-Fuller (ADF) test with $\Pr(|\tau| \geq -6.0158) < 0.01$ and Phillips-Perron (PP) test with $\Pr(|\tau| \geq -48.5686) < 0.01$ at 5% level of significance adequately declared that the data series is stationary and suggest that there is no unit root and also the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root test declared that the data series is stationary and suggest that there is no unit root with $\Pr\left(|\tau| \geq 0.1349\right) > 0.1$ at 5% level of significance. The graphical representations of the original and second differenced series are presented in Figure 1(a), (b).



(a) (b)

Figure 1: Time series plot of original series and 2nd differenced sugarcane production of Bangladesh.

It is clear that the sugarcane production data series shows initially a suddenly decreasing trend for a short time and then gradually an increasing trend and after sometimes again gradually it shows a decreasing trend that is the variance is not stable which leads the sugarcane production data series is not stationary (Figure 1(a)). However, it is clear that the second differenced sugarcane production data series shows stable variance which shows the data becomes stationary. To stabilize the variance and to make the data stationary second difference is enough that is difference order is 2 and it is said that integrated of order 2 (Figure 1(b)). The alternative positive and negative ACF (Figure 2(a)) and PACF (Figure 2(b)) indicates an autoregressive moving average process. The PACF with significant spike at lag 1 and ACF with significant spike at lag 1 suggest the first order autoregressive and first order moving average are effective on sugarcane production in Bangladesh.



(a) (b)

Figure 2: ACF and PACF plot of 2nd differenced sugarcane production in Bangladesh.

Using the tentative procedure, it is clear that ARIMA(0,2,1) model with $\text{AIC}=1201.59$, $\text{AIC}_C = 1201.91$ and $\text{BIC} = 1205.02$ is the best selected model for forecasting the sugarcane production in Bangladesh. The estimates of the parameters of the fitted ARIMA(0,2,1) model are shown in Table 1.

Table 1: Summary statistics of the fitted ARIMA(0,2,1) model

Coefficients	Estimates	Std.Error	t-value	p-value
ma1	-1.0000	0.0989	-10.11122	0.03137881

Several graphical test of the residuals of the fitted ARIMA(0,2,1) model are presented in Figure 3, suggest that there is no significant pattern and hence there is no autocorrelation among the residuals.

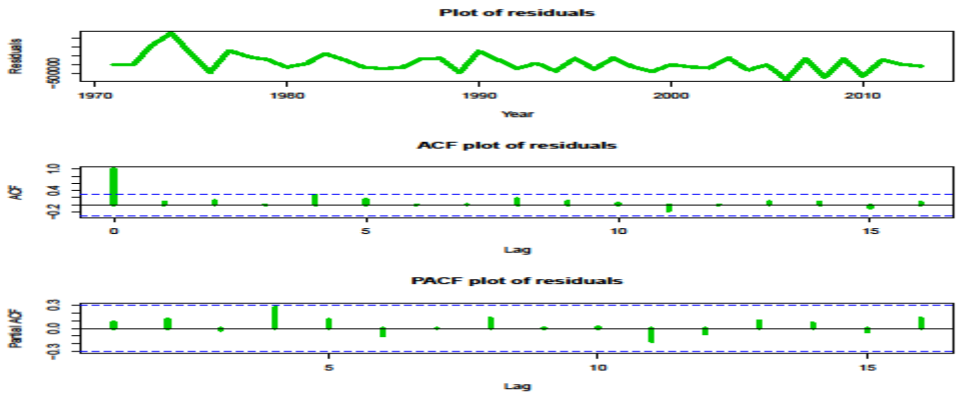


Figure 3: Several residual plots.

The "Box-Pierce" test with $\Pr\left(\left|\chi_1^2\right| \geq 0.2918\right) = 0.5891$ and the "Box-Ljung" test with $\Pr\left(\left|\chi_1^2\right| \geq 0.3126\right) = 0.5761$ at 5% level of significance strongly suggest to accept that there is no autocorrelation among the residuals of the fitted ARIMA(0,2,1) model. Also, here "Histogram with Normal Curve" is used to check the normality assumption of the residuals of the fitted model. The Histogram with Normal Curve of the residuals of the fitted ARIMA(0,2,1) model is given in Figure 4.

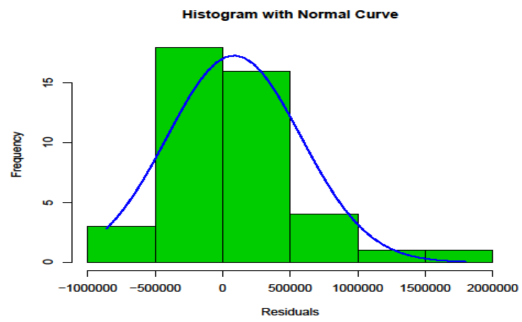


Figure 4: Histogram with Normal Curve.

Histogram with Normal Curve approximately suggests to accept the normality assumption that is the residuals of the fitted ARIMA(0,2,1) model are normally distributed. Therefore, it is clear that our fitted ARIMA(0,2,1) model is the best fitted model and adequately used to forecast the sugarcane production in Bangladesh. Also, the value of the most useful "forecasting criteria" of the fitted ARIMA(0,2,1) model are shown in Table 2.

Table 2: The forecasting criteria of the fitted ARIMA(0,2,1) model

Model	RMSPE	MPFE	TIC
ARIMA(0,2,1)	0.08151437	0.01091984	0.03792229

By using the best fitted model ARIMA(0,2,1), the forecast value and 95% confidence level for ten years are shown in Table 3.

Table 3: Forecasted sugarcane production in Bangladesh

SUGARCANE PRODUCTION (tonnes)			
Year	**Forecast**	**LCL**	**UCL**
2014	4355765	3346250.2	5365279
2015	4277530	2833355.1	5721704
2016	4199295	2410562.6	5988026
2017	4121059	2032786.3	6209332
2018	4042824	1682822.6	6402826
2019	3964589	1351978.9	6577199
2020	3886354	1035168.7	6737539
2021	3808119	729128.1	6887109
2022	3729884	431624.9	7028142
2023	3651648	141059.0	7162238

Note: LCL= Lower Confidence Limit and UCL=Upper Confidence Limit

The graphical comparison of the original series and the forecast series is shown in Figure 5. It is apparent that the original series (dark-green-color) initially shows suddenly a downward tendency for a short time and then gradually an upward tendency and after sometimes again gradually it shows a downward tendency. The forecast series (blue-color) fluctuate from the original series with a very small amount that is it shows production in the same manner of the original series (Figure 5). Also, in the forecasting plot, in sample forecasting part shows a downward trend and similarly the out sample forecasting part also shows a downward trend. Therefore, the forecast series is really better representation of the original sugarcane production series in Bangladesh.

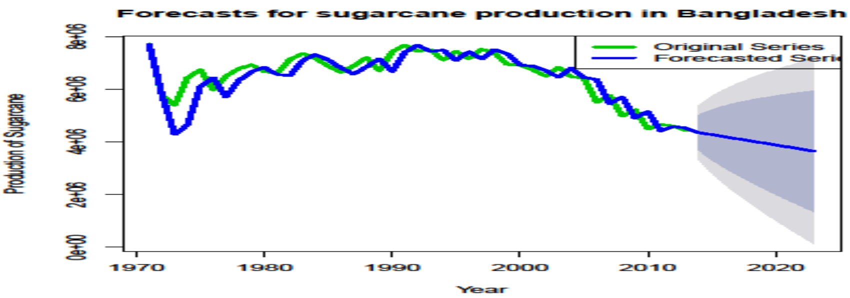


Figure 5: Comparison between original and forecasted sugarcane production in Bangladesh.

4 Conclusion

A time series model is used to identify the patterns in the past movement of a variable and uses that information to forecast the future values. This study tried to fit the best model to forecast the sugarcane production in Bangladesh with the help of the latest available model selection criteria such as AIC, BIC, etc. The best ARIMA model is selected by Box-Jenkins methodology for forecasting the sugarcane productions in Bangladesh and the selected model is ARIMA (0,2,1). 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.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] R.P Humbert, The Growing of Sugar Cane. Elsevier publishing company, Amsterdam, 1968.
[2] Sharpe, Peter, Sugar Cane: Past and Present", 1998. (http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1388&context=ebl&sei-redir=1). Southern Illinois University. Retrieved 2012-04-02.
[3] Kew, Sugar cane-history, 2004. (http://www.kew.org/plant-cultures/plants /sugar_cane_history.html). Royal Botanical Gardens
[4] George Watt, The Economic Products of India, W.H. Allen & Co. **6(II)**: 29–30, (1893).
[5] M.K.U Chowdhury and I.K Vasil, Theoretical and Applied Genetics **86 (2-3)**: 181-188 (1993).
[6] BSFIC, MIS Report. Bangladesh Sugar and Food Industries Corporation. Motijheel Commercial Area, Dhaka, Bangladesh, 2008.
[7] M. Kamruzzaman and M. Hasanuzzaman, Bangladesh Journal of Agricultural Research **32(3)**: 433-444 (2007).
[8] M.A.M Miah, An economic study on sugarcane cultivation with several intercriops in Joypurhat Sugarmills areas. An unpublished M. Sc. Thesis. Dept. of Agril. Economics, BAU, Mymensingh, 1992.
[9] M.N Sheikh and M.M Haque, Production consumption and demand for sugar in Sugar Industry in Bangladesh Eds. Shalauddin *et. al.*, BSFIC, Shilpa Bhaban, Motijheel Commercial Area, Bangladesh, 1986.
[10] Md. Golam Ferdous Chowdhury, Md. Sarwar Hossain, M. A. Sattar and Md. Shirazul Islam, Bangladesh Journal of Agricultural Research **35(1)**: 167-177 (2010).

[11] M.M Hassan, Problems of sugar mills and remedial measures. The Daily Bangladesh Observer, October 24, 2003.
[12] S.M.I Hossain, M.A Zaman, M.A Rabbani and M.S Hossain, Progress. Agric. **17(1)**: 283-293 (2006).
[13] Box, George and Jenkins, Gwilym, Time series analysis: Forecasting and control. San Francisco: Holden-Day, 1970.
[14] G.M Ljung and G.E.P Box, Biometrika **65 (2)**: 297–303 (1978). DOI:10.1093/biomet/65.2.297
[15] H. Theil, Applied Economic Forecasting. North-Holland Publishing Company, Amsterdam, 1966.