Forecasting Potato 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
Email: mmhmm.justat@gmail.com

Abstract. Bangladesh is recognized as a rice-eating nation; nevertheless, large quantities of potatoes are produced and consumed each year. In Bangladesh, potato has gradually gained popularity. Potato is used as food crop as well as vegetable by both the poor and rich people. Recently, the government has been trying to diversify food habits and encourage potato consumption to reduce pressure on rice. In this regard, potato can play an important role as an alternative and a multipurpose food crop of Bangladesh. This study considered the published secondary data of yearly potato production in Bangladesh over the period 1971 to 2013. The best selected Box-Jenkins ARIMA model for forecasting the potato production for whole Bangladesh is ARIMA(0,2,1). The comparison between the original series and forecasted series shows that the fitted model behaved statistically well and is suitable to forecast the potato productions in Bangladesh.

Keywords: Potato, time series analysis, forecasting, Bangladesh.

1   Introduction

Potatoes (*Solanum tuberosum*) have been cultivated as an important food crop the very beginning of human civilization, and their roots were in the windswept Andes Mountains of South America. In Indian sub-continent the cultivation of potatoes started probably during the 17th century (Ahmed [1]). However, potato production in Bangladesh began in the later part of the 19th century. By the 1920's, the first commercial production of the crop was established (Islam [2]). Potato has occupied an important position in area allocation in Bangladesh because of its high productivity in comparison to other food and vegetables crops. Potatoes grow normally once a year i.e., mid September to mid November is sowing period and mid January to March is harvesting period (BBS [3]). Potato grows in many different environments, but it adapts best to temperate climates (Haverkort [4]). Potato is also frost sensitive, and severe damage may occur when temperature drops below 0 °C (Hijmans [5]).

Bangladesh is recognized as a rice-eating nation; nevertheless, large quantities of potatoes are produced and consumed each year. In Bangladesh, potato has gradually gained popularity. Potato is used as food crop as well as vegetable by both the poor and rich people. People consume potato as cooked food, fries, chips and flour to make breads and biscuits. Besides, there are more than hundred palatable foods made from potatoes (Das [6]). Adequate supply of potato stabilizes the vegetable market all around the year (Moazzem and Fujita [7]). Potato are also used in processing industry such as production of starch, alcohol, glucose, dextrin and citric acid in textile and paper industries, inks, dyes, soap and leather processing (Sarker [8]). Potato also provides as much as 6 percent of the daily per capita calories and protein consumed in rural areas and a much higher percentage of total food intakes during the winter months when they are in seasonally abundant supply. This is also the off season for rice, the major staple for which potatoes are a partial substitute in many households (Scott [9]). Rice and wheat are main food in Bangladesh but their production is not sufficient to meet the increasing requirements for the growing population in the country. It is well recognized that to meet the demand for food for increasing population of Bangladesh, food habit has to be slightly diversified so as to reduce dependence of rice. To achieve this goal, production of maximum amount of nutrient per unit of area and per unit of time is to be emphasized. Potato has high nutritive value. From nutritive point of view, potato is one of the cheapest sources of carbohydrates and furnishes appreciable amount of vitamin B

and C as well as some minerals (Thompson and Kelly [10]). Potato provides twice as much as calories per unit area of land and in a shorter period of time compared to rice and wheat (Ahmed and Kamal [11]). Uddin *et al.* [12] conducted a study to build the union level digital database and maps of potato during 2008-09. Azad *et al.* [13] try to find the optimum conditions for bioethanol production from potato of Bangladesh. There are also some works reported on the potato in Bangladesh (Elias *et al.* [14]; Elias *et al.* [15]; Hakim [16]; Rashid [17]; Arif [18]; Akhter *et al.* [19] and Mukul *et al.* [20]). Previous studies about different agricultural crops in Bangladesh (Hossain and Abdulla [21]; Hossain and Abdulla [22]; Hossain and Abdulla [23]; Abdulla and Hossain [24]; Hossain and Abdulla [25]; Hossain and Abdulla [26]; Hossain and Abdulla [27]; Hossain *et al.* [28]) used ARIMA model to analyze and forecast the production of different agricultural crops like jute, tea, wheat, tomatoes etc.

One of the major problems faced by developing countries in general and Bangladesh in particular, is the ever increasing population. As per current trend, the population in Bangladesh is expected to be around 172.9 million by the year 2020. In order to increase agricultural production further, the only option is to grow high productivity crops, like potato (Azimuddin *et al.* [29]). Also, 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. Recently, the government also has been trying to diversify food habits and encourage potato consumption to reduce pressure on rice. In this regard, potato can play an important role as an alternative and a multipurpose food crop of Bangladesh. Thus it is too much essential to estimate the production of food-grains 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 potato in Bangladesh.

2   Materials and Methods

2.1   Data Source

This study considered the published secondary data of yearly potato production in Bangladesh which was collected over the period 1971 to 2013 from the website of Food and Agricultural Organization (FAO).

2.2   ARIMA Model

If $\left\{\zeta_t\right\}$ is a white noise with mean zero variance σ^2 then $\left\{Y_t\right\}$ is defined by $Y_t = \zeta_t + \beta_1\zeta_{t-1} + \beta_2\zeta_{t-2} + \ldots + \beta_q\zeta_{t-q}$, which is called a moving average process of order q and is denoted by $MA(q)$. The process $\left\{Y_t\right\}$ is given by $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \zeta_t$, which 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 $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \zeta_t + \beta_1\zeta_{t-1} + \beta_2\zeta_{t-2} + \ldots + \beta_q\zeta_{t-q}$, where, Y_t is the original series, for every t, and assume that ζ_t is independent of $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-n}$. A time series $\left\{Y_t\right\}$ is said to follow an integrated autoregressive moving average (ARIMA) model if the dth difference $W_t = \nabla^d Y_t$ is a stationary ARMA process.

2.3   Box-Jenkins Methods

The influential work of Box-Jenkins (Box and Jenkins [30]) shifted professional attention away from the stationary serially correlated deviations from deterministic trend paradigm toward the $ARIMA(p,d,q)$ paradigm. 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 the number of parameters to estimate is clear. 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   Evaluation of Forecast Errors

Before forecasting the production it is necessary to check the accuracy of the fitted model. There are many summary statistics available in the literature for evaluating the forecast errors of any model, time series or econometric. Here, an attempt is made to identify the best models for potato production in Bangladesh using the following contemporary model selection criteria, such as MASE, RMSPE, MPFE and TIC.

Mean Absolute Scaled Error (MASE): Mean Absolute Scaled Error (MASE) is defined as,

$$MASE = \frac{\frac{1}{T}\sum_{t=1}^{T}\left|y_t^a - y_t^f\right|}{\frac{1}{T-1}\sum_{t=2}^{T}\left|y_t - y_{t-1}\right|},$$ where y_t^f is the forecast value at time t and y_t^a is the actual value at time t.

Root Mean Square Percentage Error (RMSPE): Root Mean Square Percentage Error (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 at time t and y_t^a is the actual value at time t.

Mean Percent Forecast Error (MPFE): Hossain *et al.* [31] used the Mean Percent Forecast Error (MPFE) which 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 at time t and y_t^f is the forecast value at time t.

Theil Inequality Coefficient (TIC): Theil [32] 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 at time t and y_t^a is the actual value at 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 $P\left(\left|\tau\right| \geq -6.3608\right) < 0.01$ and Phillips-Perron (PP) test with $P\left(\left|\tau\right| \geq -63.2099\right) < 0.01$ at 5% level of significance adequately declared that the data series are 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 are stationary and suggest that there is no unit root with $P\left(\left|\tau\right| \geq 0.0298\right) > 0.1$ at 5% level of significance. The graphical representations of original and second differenced time series are given in Figure 1(a), (b).

It is clear that the potato production data series show an increasing trend but the variance is not stable (Figure 1(a)) and so we examine   whether the second difference of the potato production data series shows stable variance i.e., the differenced data series becomes stationary (Figure 1(b)). To stabilize the variance and make the data stationary, second difference is enough. The alternative positive and negative ACF (Figure 1(c)) and exponentially decay PACF (Figure 1(d)) indicate an autoregressive moving average process. The PACF with significant spike at lag 1 and ACF with significant spike at lag 1 suggest that first order autoregressive and first order moving average are effective on potato production in Bangladesh.
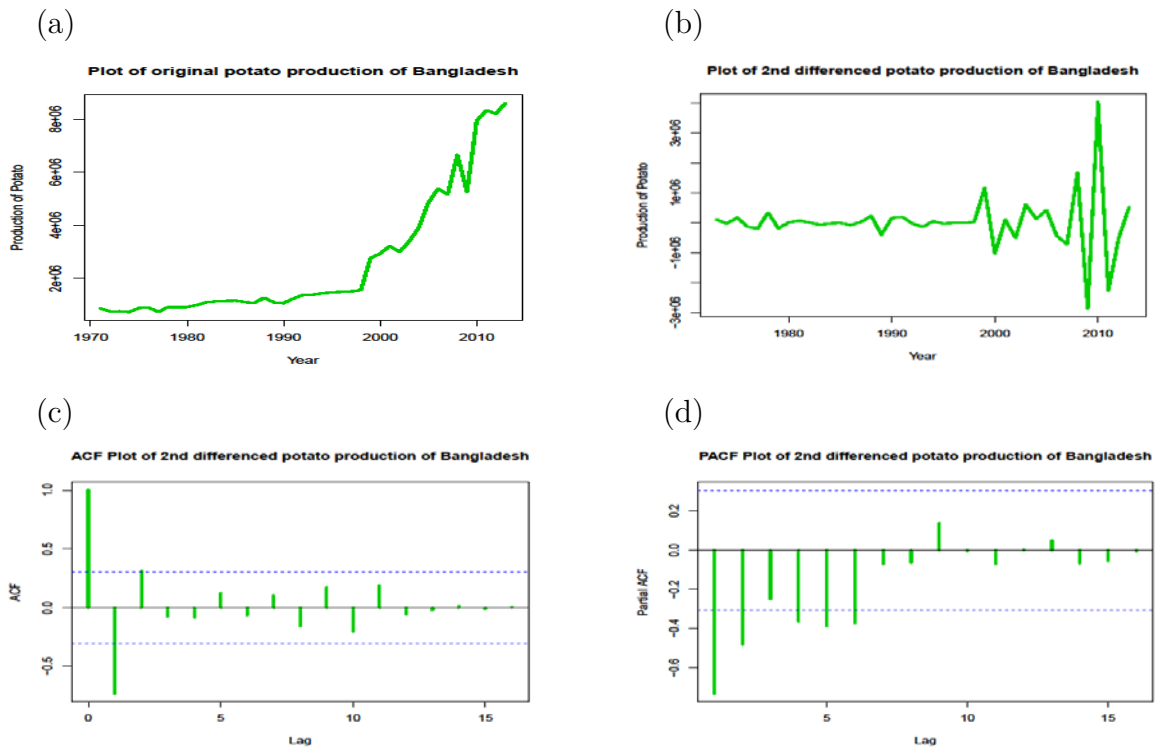
(a)   (b)

(c)   (d)

Figure 1. (a) Time series (original series) plot, (b) Time series (2nd differenced) plot (c) ACF and (d) PACF of 2nd differenced potato production of Bangladesh.

Using the tentative procedure, it is clear that ARIMA(0,2,1) model with $AIC = 1210.46$, $AIC_C = 1210.77$ and $BIC = 1213.89$ is the best selected model for forecasting the potato 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	-0.9133	0.0614	-14.87459	0.02136742

Several graphical tests of the residuals for the fitted ARIMA(0,2,1) model are presented in Figure 2, suggesting that there is no significant pattern and hence, there is no autocorrelation among the residuals. At 5% level of significance, from Swed and Eisenhart [33] Tables we obtain that the lower and upper values of the runs are 16 and 28 respectively and in this study the observed number of runs is 23. So the "Run" test with $P(16 \leq R \leq 28) = 0.95$ at 5% level of significance strongly suggests that there is no autocorrelation among the residuals of the fitted ARIMA(0,2,1) model. 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 2. Histogram

with Normal Curve approximately suggests that 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 is adequate to forecast the potato production in Bangladesh.
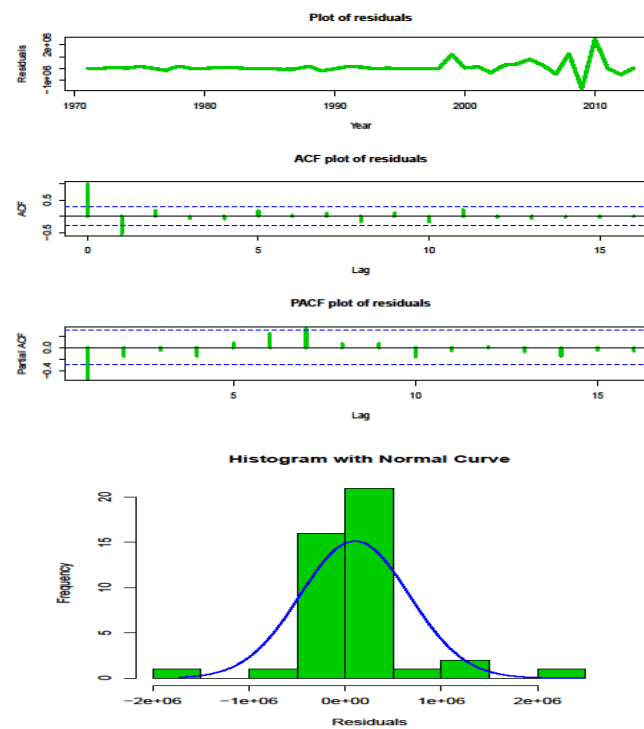

Figure 2. Several plots of residuals and Histogram with Normal Curve.

The values of the most useful "forecasting criteria" of the fitted ARIMA(0,2,1) model are $MASE = 0.8932963$, $RMSPE = 0.1322551$, $MPFE = 0.0288415$, and $TIC = 0.08264775$. By using the best fitted model ARIMA(0,2,1), the forecast value and 95% confidence level for ten years are shown in Table 2.

Table 2. Forecasted potato production (tonnes) in Bangladesh

Year	Forecasted value	LCL	UCL
2014	8961660	7823378	10099942
2015	9320320	7639257	11001384
2016	9678981	7531810	11826151
2017	10037641	7455341	12619940
2018	10396301	7393097	13399504
2019	10754961	7336964	14172958
2020	11113621	7282412	14944831
2021	11472281	7226684	15717879
2022	11830942	7168006	16493878
2023	12189602	7105187	17274016

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 3. It is apparent that the original series (dark-green-color) initially shows equal production and after some time shows a slightly upward tendency and again after 1981 shows a rapidly upward tendency in the potato production in Bangladesh. The forecast series (blue-color) fluctuates from the original series with a very

small amount which shows production in the same manner of the original series (Figure 3). So, the forecasted series gives really better representation of the original potato production series in Bangladesh.
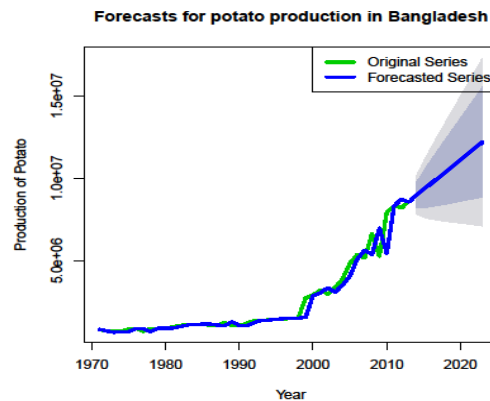

Figure 3. Comparison of original and forecasted potato production in Bangladesh.

4   Conclusions

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 potato production in Bangladesh with the help of the latest available model selection criteria such as AIC, BIC, etc. The best selected Box-Jenkins ARIMA model for forecasting the potato for whole Bangladesh is ARIMA(0,2,1). The comparison between the original series and forecasted series shows the same manner indicating fitted model is statistically well behaved to forecast the potato productions in Bangladesh i.e., the models forecast well during and beyond the estimation period to a satisfactory level. Thus, these models can be used for policy purposes as far as forecasting the potato production in Bangladesh.

Potato in Bangladesh is produced once a year. Farmers face many hardships and make struggle to obtain a satisfactory yield over the cultivation period, but when it comes to marketing, the revenue of their production is unsatisfactory, and as a result the farmer falls into debt again. Most of the farmers are illiterate in Bangladesh. Dissemination of market information should be increased, so that farmers can get fair price of the potato. Potato farming is assuming a greater dimension, however bringing pressure on the government to expand its use as alternative food in the domestic sector while looking for greater export markets in overseas trade.

References

1. Ahmed, K. U. (1977). *Report on Potato Production in Bangladesh, International Potato Course: Production, Storage and Seed Technology*, Report of Participants, International Agricultural Center, Wageningen, The Netherlands.
2. Islam, M. A. (1983). *Report on Potato Production in Bangladesh, International Potato Course: Production, Storage and Seed Technology*, Report of Participants, International Agricultural Center, Wageningen, The Netherlands.
3. BBS (2004). *Statistical Pocketbook of Bangladesh 2002*, Bangladesh Bureau of Statistics, Government of the People's Republic of Bangladesh.

4. Haverkort, A. J. (1990). Ecology of potato cropping systems in relation to latitude and altitude, *Agricultural Systems,* 32(3): 251-272.

5. Hijmans, R. J. (2003). The Effect of Climate Change on Global Potato Production, *Amer J of Potato Res*, 80: 271-280.

6. Das, S. C. (1992). *An Economic Analysis of Potato Growing Farms with Respect to Production and Resource Use Efficiency*, Department of Agricultural Economics, Bangladesh Agricultural University, Mymensing, Bangladesh, M.S. thesis.

7. Moazzem, K. G. and Fujita, K. (2004). Potato marketing system and its changes in Bangladesh: From the perspective of village study in Comilla district, *The Developing Economies*, XLII-1 (March): 63-94.

8. Sarker, K. R. (2004). *Marketing System and Price Behavior of Potato areas of Munshigonj District*, Department of Agricultural Economics, Bangladesh Agricultural University, Mymensing, Bangladesh, M.S. thesis.

9. Scott, G. (1985). *Marketing Bangladesh's Potatoes: Present Patterns and Future Prospects*, Lima: International Potato Center.

10. Thompson, H. C. and Kelly, W. C. (1957). *Vegetable crops*, McGraw-Hill Book Company, Inc., New York.

11. Ahmed, K. U. and Kamal, B. M. (1984). *Bidhatar Daan Alu.* Dhaka: Bunglow No. 2, Krishi Khamar Sharak, Farm Gate, Dhaka, Bangladesh.

12. Uddin, M. A. Yasmin, S. Rahman, M. L. Hossain, S. M. B. and Choudhury, R. U. (2010). Challenges of Potato Cultivation in Bangladesh and Developing Digital Databases of Potato, *Bangladesh Journal* of *Agricultural Research*, 35(3): 453-463.

13. Azad, A. K., Yesmin, N., Sarker, S. K., Sattar, A. and Karim, R. (2014). Optimum Conditions for Bioethanol Production from Potato of Bangladesh, *Advances in Bioscience and Biotechnology*, 5: 501-507. http://dx.doi.org/10.4236/abb.2014.56060

14. Elias, S. M. A. Jubber and N. I. Mondol (1980). *An Economic profile of potato cultivation in Bangladesh,* Agricultural Economics Research Report No. 2 Joydebpur, Agricultural Economics Division, BARRI.

15. Elias, S. M. and Nazrul, M. I. (1982). *Socio- economic Assessment of improved Technology of potato and Identification and Constraints to its higher production,* Agricultural Economics Research Report No. 4. Joydebpur, Agricultural Economics Division, BARRI.

16. Hakim, M. A. (1993). *A Comparative Economic Analysis of Cardinal and Malta varieties of potato in kotwali Thana of Bogra District.* M.Sc. Ag. Econ. thesis submitted to the department of Agricultural Economics, Bangladesh Agricultural University, Mymensing.

17. Rasid, M. H. (1994). *An Economic study of Farmers growing crops with potato in a selected area of Rangpur District.* M.Sc. Ag. Econ. thesis submitted to the department of Agricultural Economics, Bangladesh Agricultural University, Mymensing.

18. Arif, A. (1998). *An Economic study of potato production in selected areas of Comilla Districts.* M.Sc. Ag. Econ. thesis submitted to the department of Agricultural Economics, Bangladesh Agricultural University, Mymensing.

19. Akhter, S., Anwar, M. M. and Asaduzzaman, M. (2001). *Potato production in some selected areas of Bangladesh* . TCRC, BARRI, Joydebpur, Gazipur, Bangladesh.

20. Mukul, A. Z. A., Rayhan, S. J. and Hassan, M. M. (2013). Farmer's Profitability of Potato Cultivation at Rangpur District: The Socio-economic Context of Bangladesh, *International journal of Economics, Finance and Management Science*, 1(3): 136-144.   doi: 10.11648/j.ijefm.20130103.12

21. Hossain, M. M. and Abdulla, F. (2015). Jute Production in Bangladesh: A Time Series Analysis, *Journal of Mathematics and Statistics*, 11(3): 93-98.

22. Hossain, M. M. and Abdulla, F. (2015). Forecasting the Tea Production of Bangladesh: Application of ARIMA Model, *Jordan Journal of Mathematics & Statistics*, 8(3): 257-270.

23. Hossain, M. M. and Abdulla, F. (2015). On the production behaviors and forecasting the tomatoes production in Bangladesh, *Journal of Agricultural Economics and Development*, 4(5): 66-74.

24. Abdulla, F. and Hossain, M. M. (2015). Forecasting of Wheat Production in Kushtia District & Bangladesh by ARIMA Model: An Application of Box-Jenkin's Method, *Journal of Statistics Applications & Probability*, 4(3): 465-474.

25. Hossain, M. M. and Abdulla, F. (2015). Forecasting the Sugarcane Production in Bangladesh by ARIMA Model, *Journal of Statistics Applications & Probability*, 4(2): 297-303.

26. Hossain, M. M. and Abdulla, F. (2015). A Time Series Analysis for the Pineapple Production in Bangladesh, *Jahangirnagar University Journal of Science*, 38(2): 49-59.

27. Hossain, M. M. and Abdulla, F. (2015). Forecasting the Garlic Production in Bangladesh by ARIMA Model, *Asian Journal of Crop Science*, 7(2): 147-153. DOI: 10.3923/ajcs.2015.147.153

28. Hossain, M. M., Abdulla, F. and Majumder, A. K. (2016). Forecasting of Banana Production in Bangladesh, *American Journal of Agricultural and Biological Sciences*, 11(2): 93-99. DOI: 10.3844/ajabssp.2016.93.99.

29. Azimuddin M., Alam, Q. M. and Baset, M. A. (2009)**.** Potato for Food Security in Bangladesh, *International Journal of Sustainable Crop Production*, 4(1): 94-99.

30. Box, G. and Jenkins, G. (1970). *Time series analysis: Forecasting and Control*, San Francisco: Holden-Day.

31. Hossain, M. Z., 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(4): 344-353. DOI 10.1108/03068290610651652

32. Theil, H. (1966). *Applied Economic Forecasting*, North-Holland Publishing Company, Amsterdam.

33. Swed, Frieda S. and Eisenhart, C. (1943). Tables for Testing Randomness of Grouping in a Sequence of Alternatives, *The Annals of Mathematical Statistics*, 14(1): 66-87.