Original Research Paper

Forecasting of Banana Production in Bangladesh

[1]Md. Moyazzem Hossain, [2]Faruq Abdulla and [1]Ajit Kumar Majumder

[1]*Department of Statistics, Jahangirnagar University, Bangladesh*
[2]*Department of Statistics, Islamic University, Bangladesh*

Abstract: In Bangladesh, banana is a very popular fruit and cultivated almost everywhere round the year. Rank of Bangladesh is 14 among the top 20 banana producing countries in the world. It is a commercial fruit, but in Bangladesh it is grown in limited area commercially. The demand of banana is increasing day by day in Bangladesh. Thus, this paper attempts to identify the Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the production of banana in Bangladesh. This study considered the secondary data of yearly banana production in Bangladesh over the period 1972 to 2013. The best selected ARIMA model to forecast the banana productions in Bangladesh is ARIMA (0,2,1). The graphical comparison between the observed and forecasted banana production indicate the fitted model behaved statistically well during and beyond the estimation period.

Keywords: Banana, ARIMA Model, Forecasting, Bangladesh

Introduction

Banana (*Musa paradisiaca*, family Musaceae) is a central fruit crop of the tropical and subtropical regions of the world grown on about 8.8 million hectares (Mohapatra *et al*., 2010). It is possibly the world's oldest cultivated plants (Kumar *et al*., 2012). Bangladesh produces nearly 1.00 million tonnes of bananas annually (Hossain, 2014). It is also a nutritious fruit crop in the world and grown in many tropical areas where they are used both as a staple food and dietary supplements (Assani *et al.*, 2001). The total per capita consumption of banana in Bangladesh is about 4.7 kg. This is very much lower than that consumed by Europe especially Belgium (26.7 kg), Sweden (16.7 kg) and Germany (14.5 kg) while USA consumed 13.1 kg and UK at 10.5 kg (Siti Hawa, 1998).

Banana is mainly cultivated for it's ripen fruits, cooked vegetables and leaves in India and many other countries including Bangladesh (Khanum *et al*., 2000). It is the second largest produced fruit after citrus, contributing about 16% of the world's total fruit production (FAO, 2009). Banana is highly nutritious (Sharrock and Lustry, 2000) and is more easily digestible than many other fruits including apple (Mohapatra *et al*., 2010). Banana is cultivated almost everywhere in Bangladesh round the year. The foremost banana growing areas in Bangladesh are Narsingdi, Gazipur, Tangail, Rangpur, Bogra, Natore, Pabna, Noakhali, Faridpur and Khulna. Also, Sylhet, Moulvibazar, Netrokona, Rangamati, Khagrachhari and Bandarban are wild grown banana area in Bangladesh. In 2010-2011, the total production of banana in Bangladesh was 800840 metric tons and the cultivated area was about 130589 acres (BBS, 2012). The banana fruit is variable in size, color and firmness, but is usually elongated and curved, with soft flesh rich in starch covered with a rind which may be green, yellow, red, purple, or brown when ripe. The fruits grow in clusters hanging from the top of the plant. As a diet, banana is an affluent source of carbohydrate with calorific value of 67 calories per 100 g fruit and is one of the most well-liked and widely traded fruits across the world (Emaga *et al*., 2008; Kumar *et al*., 2012). Banana is a rich source of calories, as well as most of the vitamins essential for human nutrition. Bananas are also rich in carbohydrate, potassium and vitamins, including A, C and B6. They are a good source of fat-free dietary fiber. Banana is often the first solid food fed to infant. Ripe banana mixed with rice and milk is the traditional dish for Bangladeshi (Hossain, 2014). Several studies have been conducted to analyze the banana production in Bangladesh (Ahmad *et al*., 1973; 1974; Haque, 1984; Islam and Hoque, 2005; Hoque, 2006; Roy *et al.*, 2006; Ara *et al*., 2011; Mukul and Rahman, 2013; Mohiuddin *et al.*, 2014; Hossain *et al.*, 2015).

Hamjah (2014) fitted ARIMA Model to forecast the different types of major fruits productions in Bangladesh. He found that ARIMA(2,1,3), ARIMA(3,1,2) and ARIMA(1,1,2) are the best model to forecast the Mango, Banana and Guava productions respectively in Bangladesh. Casinillo and Manching (2015), determine the trend of banana of two classes namely, Class A and Class B, using the Box and Jenkins methodology. The identified models for class A and class B bananas were MA(12) and ARIMA(1,6,2) respectively and was identified to be well fitted to the series showed by some statistical tests. Banana production provides suitable options for subsistence and income generation in Bangladesh. It is a commercial fruit all over the world but in Bangladesh, it is grown in limited area commercially. Moreover, a large number of people were involved in the production and marketing of banana in Bangladesh. Also, the demand of banana is increasing day by day in Bangladesh. A small quantity of banana is exported to the Middle-East and European countries. Although bananas are important export commodities of some developing countries in Africa, Latin America and the Asia, unfortunately Bangladesh is not an exporting country. Thus it is necessary to estimate the banana production in Bangladesh so that we can meet our country demand and export it for earning the foreign currency. The main purpose of this paper is to identify the Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the banana production in Bangladesh.

Materials and Methods

Data Source

This study considered the published secondary data of yearly banana production in Bangladesh which was collected over the period 1972 to 2013 from the Food and Agricultural Organization (FAO) website.

ARIMA Model

If $\{\zeta_t\}$ is a white noise with mean zero variance σ^2 then $\{Y_t\}$ is called a moving average process of order q denoted by MA(q) and is defined by:

$$Y_t = \zeta_t + \beta_1 \zeta_{t-1} + \beta_2 \zeta_{t-2} + ... + \beta_q \zeta_{t-q} \tag{1}$$

The process $\{Y_t\}$ is called an auto-regressive process of order p and is denoted by AR(p), is defined by:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \zeta_t \tag{2}$$

Models that are combination of AR and MA models are known as ARMA models. An ARMA(p,q) model is defined as:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p}$$
$$+ \zeta_t + \beta_1 \zeta_{t-1} + \beta_2 \zeta_{t-2} + ... + \beta_q \zeta_{t-q} \tag{3}$$

where, Y_t is the original series, for every t and assume that ζ_t is independent of Y_{t-1}, Y_{t-2},..., Y_{t-p}. A time series $\{Y_t\}$ 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. If $\{W_t\}$ follows an ARMA(p,q) model, $\{Y_t\}$ is said to be an ARIMA(p,d,q) process. An ARIMA(p,1,q) process is defined as:

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + ... + \alpha_p W_{t-p}$$
$$+ \zeta_t + \beta_1 \zeta_{t-1} + \beta_2 \zeta_{t-2} + ... + \beta_q \zeta_{t-q} \tag{4}$$

where, $W_t = Y_t - Y_{t-1}$.

Box-Jenkins Method

The influential work of Box and Jenkins (1970) 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.

Ljung-Box Test

Ljung-Box (Ljung and Box, 1978) test can be used to check autocorrelation among the residuals. In this case the null hypothesis is $H_0: \rho_1(e) = \rho_2(e) = ... = \rho_k = 0$ is tested with the Ljung-Box statistic,

$$Q^* = N(N+1)\sum_{i=1}^{k}(N-k)\rho_k^2(e) \qquad (5)$$

where, N is the number of observation used to estimate the model. This statistic Q^* approximately follows the chi-square distribution with (k-q) degrees of freedom, where q is the number of parameter should be estimated in the model. If Q^* is large (significantly large from zero), it is said that the residuals of the fitted model are probably auto-correlated. Thus, one should then consider reformulating the model.

Evaluation of Forecast Error

Before forecasting it is necessary to estimate the Time Series model and evaluating the performance of the best fitted model. Here, an attempt is made to identify the best models for banana 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:

$$RMSPE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{Y_t^f - Y_t^a}{Y_t^a}\right)^2} \qquad (6)$$

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

Mean Percent Forecast Error (MPFE)

Mean Percent Forecast Error (MPFE) is defined as:

$$MPFE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{Y_t^a - Y_t^f}{Y_t^a}\right) \qquad (7)$$

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 (1966) Inequality Coefficient (TIC) is defined as:

$$TIC = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^f - Y_t^a\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^f\right)^2}} \qquad (8)$$

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

Results and Discussion

In order to make forecasting a time series it is necessary to check the time series is stationary or not first. During the study period the average Banana production in Bangladesh is around 680,753 tonnes per annum with a standard deviation 10,4967.17 tonnes. The maximum production was 1004,520 tonnes occurred at 2007 and the minimum production was 562,000 tonnes was in 1999. Here, the data set is divided into two parts namely training (1972-2003) and test (2004-2013). On the basis of the training data set we build the model and compare the forecasting with test part. To test the stationary of the data series, this paper considered Augmented-Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root test. After second differencing, the ADF test with $\Pr(|\tau|\geq-4.5348)<0.1$ and KPSS unit root test with $\Pr(|\tau|\geq0.0275)>0.1$ at 5% level of significance adequately declared that the data series is stationary which suggest that there is no unit root. The graphical representations of the original and second differenced series are presented in Fig. 1a and 1b.

It is clear that the yearly banana production in Bangladesh fluctuated over the study period 1972 to 2003. It started at about 586 thousand tonnes in 1972, reached a peak in 1986 of around 759 thousand tonnes. There is an increasing trend up to 1986 and after that the production fall dramatically. Between the periods 1989 to 1998, the production of banana was all most equal and then again falls for a short time. After the year 2000, there was a dramatic increasing of banana production i.e., banana production data series is not stationary (Fig. 1a). However, it is clear that the second differenced banana production data series becomes stationary. To make the data stationary second difference is enough (Fig. 1b). So, the difference order is 2 and it is said that the banana production is integrated of order 2. The alternative positive and negative ACF (Fig. 1c) and exponentially decay PACF (Fig. 1d) indicates 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 banana production in Bangladesh. The iterative procedure is used to select the best ARIMA model with the help of AIC, AIC_C and BIC. It is clear that ARIMA(0,2,1) model with AIC = 712.77, AIC_C = 713.22 and BIC = 715.57 is the best selected model for forecasting the banana production in Bangladesh. The estimated 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

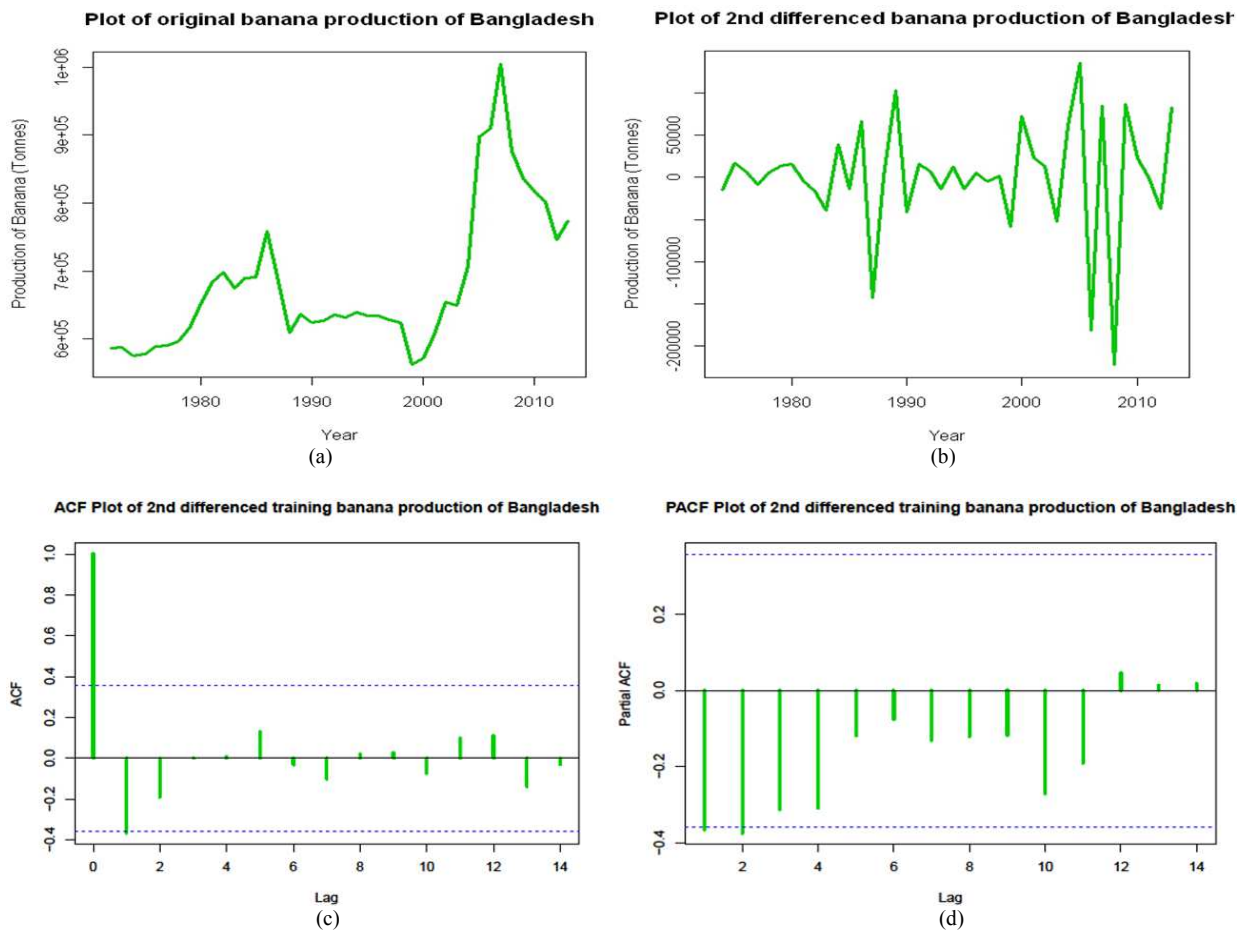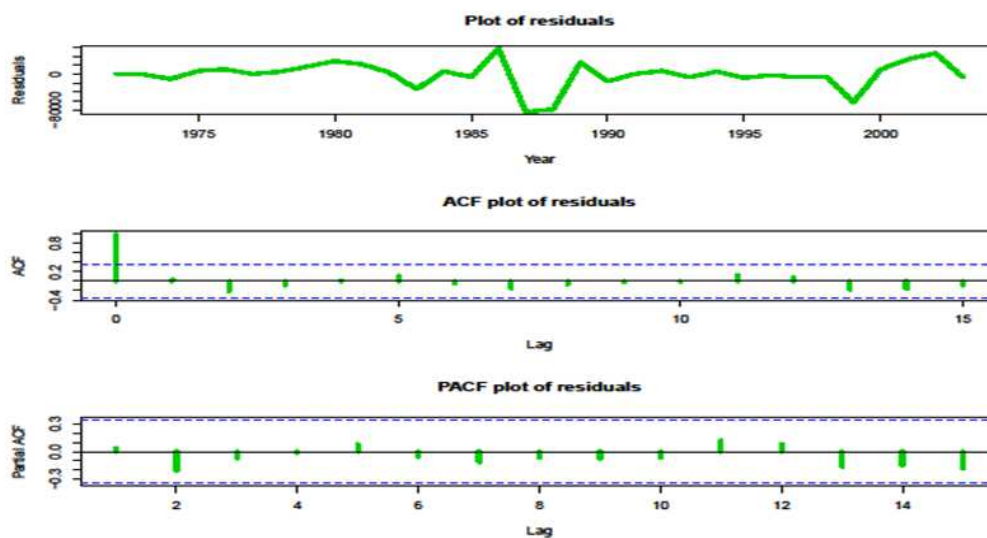| Coefficients	Estimates	Std.Error	t-value	p-value
Ma1	-1.0000	0.1103	-9.066183	0.03496823

(a)

(b)

(c)

(d)

Fig. 1. (a) Time series (original series) plot, (b) Time series (2nd differenced) plot (c) ACF and (d) PACF of 2nd differenced banana production in Bangladesh over the period 1972-2003
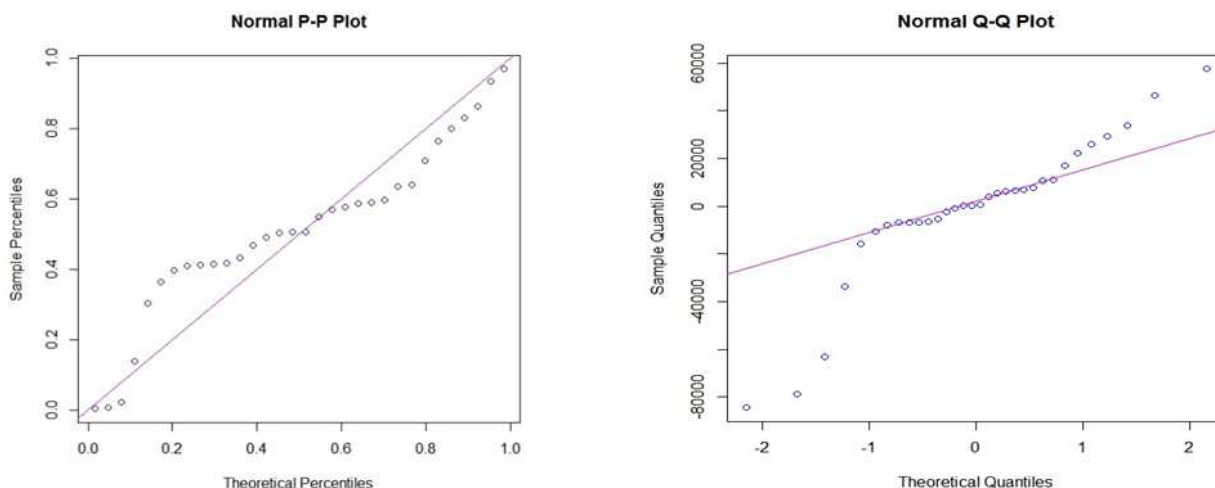
Fig. 2. Time Series plot, ACF and PACF of residuals, Normal P-P plot and Q-Q plot
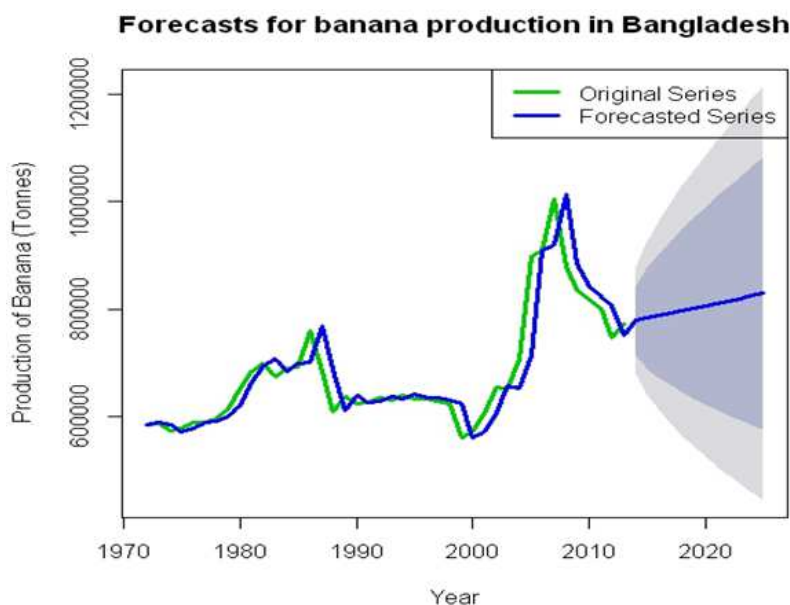

Fig. 3. Comparison of original and forecasted banana production in Bangladesh

Several graphical representations of the residuals of the fitted ARIMA(0,2,1) model are presented in Fig. 2 and suggest that there is no significant pattern. Also, the "Box-Pierce" test with $\Pr\left(\left|\chi_i^2\right| \geq 0.0475\right) = 0.8274$ and the "Ljung-Box" test with $\Pr\left(\left|\chi_i^2\right| \geq 0.0521\right) = 0.8194$ at 5% level of significance strongly suggest that there is no autocorrelation among the residuals of the fitted ARIMA(0,2,1) model. Here, the Normal Q-Q plot and Normal P-P plot are used to check the normality assumption of the residuals of the fitted model. The Normal Q-Q plot and the P-P plot of the fitted ARIMA(0,2,1) model are also presented in Fig. 2. From

P-P and Q-Q plots we may conclude that the errors of the fitted model are approximately normally distributed. Therefore, it is clear that the fitted ARIMA(0,2,1) model is the best fitted model and adequately used to forecast the banana production in Bangladesh.

The value of the most useful "forecasting criteria" of the fitted ARIMA(0,2,1) model are *RMSPE* = 19.577, MPFE = 17.60806 and *TIC* = 0.103167. The graphical comparison of the observed and the forecast banana production is presented in Fig. 3. It is observed that the forecast banana production (blue-color) fluctuated from the observed banana production (dark-green-color) with a very small amount which shows the fitted model for banana production is well (Fig. 3). Therefore, the forecasted

banana production is really better representation of the original banana production in Bangladesh.

Conclusion

The selected Box-Jenkins ARIMA model for forecasting the banana productions in Bangladesh is ARIMA(0,2,1). The graphical comparison between the observed and forecasted banana production shows little variations which indicate the fitted model behaved statistically well to forecast banana productions in Bangladesh i.e., the models forecast well during and beyond the estimation period. Thus, with the help of the fitted model we can forecast the banana production which helps to make decision of the decision makers about the demand of banana in Bangladesh. They can easily identify whether they import or export the banana meeting the country needs. Thus, this model can be used for policy purposes about banana production in Bangladesh.

Acknowledgment

Author's Contributions

Md. Moyazzem Hossain: Design, Collected data and checked the analyzed data and wrote the manuscript, contributed in interpreting data and the final art work.

Faruq Abdulla: Coordinated the study, analyzed data and helped in drafting the manuscript.

Ajit Kumar Majumder: Supervised the works, reviewed the manuscript and contributed in interpreting data.

Ethics

The authors declare no competing interests.

Reference

Ahmad, K. M.A. Matin and M.A. Quasem, 1973. Performance of some banana varieties when grown on damp land. Bangladesh Hort., 1: 70-72.

Ahmad, K., A.K.M.A. Hossain and B. Hossain, 1974. A comparative study on four table varieties of banana. Bangladesh Hort., 2: 5-11.

Assani, A.R., G. Haicour, F. Wenzel, F. Cote and F. Bakry *et al*., 2001. Plant regeneration from protoplasts of dessert banana cv. Grande Naine (*Musa* spp., Cavendish sub-group AAA) via somatic embryogenesis. Plant Cell Rep., 20: 482-488. DOI: 10.1007/s002990100366

Ara, N., M.K. Basher and M.F. Hossain, 2011. Growth, yield and quality of banana (*Musa sapientum* L) influenced by different banana varities/lines and planting time. Tropical Agric. Res. Extens., 14: 45-51.

BBS, 2012. Statistics year book of Bangladesh. Bangladesh bureau of statistics. Ministry of Planning, Government of the people's Republic of Bangladesh, Dhaka.

Box, G. and G. Jenkins, 1970. Time Series Analysis: Forecasting and Control. 1st Edn., Holden-Day, San Francisco.

Casinillo, M.A.C. and I.C.R. Manching, 2015. Modeling the monthly production of banana using the box and Jenkins analysis.

Emaga, T.H., R.H. Andrianaivo, B. Wathelet, J.T. Tchango and M. Paquot, 2008. Effects of the stage of maturation and varieties on the chemical composition of banana and plantain peels. Food Chem., 103: 590-600. DOI: 10.1016/j.foodchem.2006.09.006

FAO, 2009. Production Year Book. Food and Agriculture Organization of the United Nations, Rome, Italy.

Hamjah, M.A., 2014. Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. J. Econom. Sustainable Dev., 5: 96-107.

Haque, M.A., 1984. Effect of planting time on the production of Banana in Bangladesh. Bangladesh J. Agric. Res., 9: 23-32.

Hossain, M.M., M.A. Alam and M.K. Uddin, 2015. Application of stochastic frontier production function on small banana growers of kushtia district in Bangladesh. J. Stat. Applic. Probability, 4: 337-342. DOI: 10.12785/jsap/040218

Hossain, M.F., 2014. A study of banana production in Bangladesh: Area, yield and major constraints. ARPN J. Agric. Biol. Sci., 9: 206-210.

Hoque, M.A., 2006. Effect of planting time on yield and quality of Sabri banana. Bangladesh J. Agric. Res., 31: 323-330.

Islam, S.M. and M.A. Hoque, 2005. Status of banana production in Bangladesh. Proceedings of the International Conference on Mechanical Engineering, (CME' 05), Dhaka, Bangladesh pp: 33-41.

Khanum, F., M.S. Swamy, K.K.R. Sudarshana, K. Santhanam and K.R. Viswanathan, 2000. Dietary fiber content of commonly fresh and cooked vegetables consumed in India. Plant Foods Human Nutrit., 55: 207-218.
DOI: 10.1023/A:1008155732404

Kumar, K.P.S., D. Bhowmik, S. Duraivel and M. Umadevi, 2012. Traditional and medicinal uses of banana. J. Pharmacognosy Phytochem., 1: 51-63.

Ljung, G.M. and G.E.P. Box, 1978. On a measure of lack of fit in time series models. Biometrika, 65: 297-303. DOI: 10.1093/biomet/65.2.297

Mohapatra, D., S. Mishra and N. Sutar, 2010. Banana and its by-product utilization: An overview. J. Sci. Indus. Res., 69: 323-329.

Mohiuddin, A.K.M., M.K. Saha, M.S. Hossian and A. Ferdoushi, 2014. Usefulness of banana (*Musa paradisiaca*) wastes in manufacturing of bio-products: A review. Agriculture, 12: 148-158. DOI: 10.3329/agric.v12i1.19870

Mukul, A.Z.A. and M.A. Rahman, 2013. Production and profitability of banana in Bangladesh-an economic analysis. Int. J. Econom. Finance Manage. Sci., 1: 159-165. DOI: 10.11648/j.ijefm.20130103.15

Roy, S., M. Asaduzzaman, M.H.R. Pramanik and A.K.M.A. Prodhan, 2006. Effect of banana plant extracts on germination and seedling growth of some vegetable crops. Bangladesh J. Crop Sci., 17: 235-242.

Sharrock, S. and C. Lustry, 2000. Nutritive value of banana, in INIBAP. Annual Report, INIBAP, Montpellier, France.

Siti Hawa, J., 1998. Commercial exploitation of the banana diversity in Malaysia. Proceedings of the 1st National Banana Seminar in Malaysia, (BSM' 98), pp: 45-51.

Theil, H., 1966. Applied Economic Forecasting. 1st Edn., North-Holland Publishing Company, Amsterdam, pp: 474.