ISSN 1022-8594
JUJS

Vol. 38, No. 2, pp. 49-59

A Time Series Analysis for the Pineapple Production in Bangladesh

Md. Moyazzem Hossian[a]* and Faruq Abdulla[b]
[a]Dept. of Statistics, Jahangirnagar University, Savar, Dhaka.
[b]Dept. of Statistics, Islamic University, Kushtia-7003

Abstract
Pineapple is one of the most important commercial fruit crops in the world. It is the third most important tropical fruit in the world and in Bangladesh, pineapple ranks 4th in terms of total cropping area and production. The production of fruits including pineapple is increasing day by day in Bangladesh. Also, the world pineapple demand has been expanding rapidly. Moreover, a large number of people were involved in the production and marketing of Pineapple in Bangladesh which contribute our economy as well as GDP. Thus, it is necessary to estimate the Pineapple production in Bangladesh. The main purpose of this paper is to identify the Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the production of Pineapple in Bangladesh. This paper considered the published secondary data of yearly Pineapple production in Bangladesh over the period 1972 to 2013. The best selected ARIMA model to forecast the Pineapple productions in Bangladesh is ARIMA (0,2,1). The comparison between the original series and forecasted series shows the same manner indicating the fitted model behaved statistically well and suitable to forecast the Pineapple productions in Bangladesh i.e., the models forecast well during and beyond the estimation period.

Keywords: Pineapple, ARIMA Model, Forecasting, Bangladesh.

Introduction

Pineapple [Ananas comosus (L.) Merr. Family: Bromeliaceae] is one of the most important commercial fruit crops in the world. It is known as the queen of fruits due to its excellent flavour and taste (Baruwa, O. I., 2013 [1]). Pineapple is the third most important tropical fruit in the world after Banana and Citrus (Bartholomew, D. P., et al., 2003 [2]).

***Correspondence E-mail:** mmhmm.justat@gmail.com

In Bangladesh, pineapple ranks 4th in terms of total cropping area and production. The prospect of pineapple farming is bright in this country, because, though about 90 countries produce the fruit around the globe, Bangladeshi pineapples are more juicy and tasty than the others. If marketed properly, these pineapples are sure to fetch huge amounts of foreign currency through exports (Golam S., 2014 [3]). Agriculture is the main source of livelihood for the people in Bangladesh. Fruits such as, mango and pineapple are produced here in large quantities. The tropical climate is better for pineapple cultivation. Generally, it is grown almost all over Bangladesh especially in hilly and high land where there is no water stagnation. Long time drought is harmful for the production of pineapple. Drought affects its quality, quantity and size. Although Bangladesh is not a tropical country, the climate and the soils of many parts of Bangladesh are much more suitable for pineapple production. It is widely cultivated in the districts of Sylhet, Moulvibazar, Chittagong, Bnandarban, Dhaka and Tangail (Hasan, S. S., et al., 2010 [4]).

Pineapple fruits exhibit high moisture, high sugars, soluble solid content ascorbic acid and low crude fiber. Thus pineapple can be used as supplementary nutritional fruit for good personal health (Hemalatha, R. and Anbuselvi, S., 2013 [5]). The pineapple fruits are normally consumed fresh or as fresh pineapple juice. Field ripe fruits are best for eating fresh, and it is only necessary to remove the crown, rind, eyes and core. Pineapple may be consumed fresh, canned, juiced, and are found in a wide array of food stuffs - dessert, fruit salad, jam, yogurt, ice cream, candy, and as a complement to meat dishes (Debnath, P., et al., 2012 [6]). Pineapple juice is largely consumed around the world, mostly as a canning industry byproduct, in the form of single strength, reconstituted or concentrated and in the blend composition to obtain new flavors in beverages and other products (Carvalho et al., 2008 [7]). Pineapple contains considerable amount of calcium, potassium, vitamin C, carbohydrates, crude fiber, water and different minerals that is good for the digestive system and helps in maintaining ideal weight and balanced nutrition. Pineapple is a common fruit in Bangladesh and it has minimal fat and sodium (Sabahelkhier, K. M., et al. 2010 [8]). Md. Farid Hossain, et al., 2015 [9] discusses nutritional values and importance of pineapple in the health aspects.

Pineapple creates low blood pressure, cure inflammation disease, used for weight loss, control the death rate and prevent diabetes & radical damage. It cures the damaged teeth and makes them strong and healthy. Also help to cure sinusitis and throat problem. Cure different diseases like asthma, obesity, swollen in the body, problems of digestion and heart problem. Pineapples are rich of manganese which creates strong bones and muscular body. Atherosclerosis and immune disease can be also cured due to high antioxidant. It does not let damage the cells of body, it is so hot so it is used to ignore cold weather, also used for perfect powerful unbreakable body, prevent cancer, heart attack, nausea and gives the long natural hairs. Use to solve acne, wrinkles, age problem and create strong nails, soft lips and thick hair (Idrees, 2014 [10]). The *Garo* tribal community of Netrakona district in Bangladesh uses fruit juice for fever and leaf juice for helminthiasis and jaundice (Rahmatullah, M., et al., 2009 [11]). The production of fruits including pineapple is increasing day by day in Bangladesh. Among all the fruits produced in the country, pineapple ranks 4th in terms of total cropping area and production. Also, the world pineapple demand has been expanding rapidly. Moreover, a large number of people were involved in the production and marketing of Pineapple. Thus, it is necessary to estimate the Pineapple production in Bangladesh. The main purpose of this paper is to identify the Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the Pineapple production in Bangladesh.

Materials and Methods

Data Source

This study considered the published secondary data of yearly Pineapple production in Bangladesh which was collected over the period 1972 to 2013 from the Food and Agricultural Organization (FAO) website (http://faostat3.fao.org).

ARIMA Model

Suppose that $\{\zeta_t\}$ is a white noise with mean zero variance σ^2, then $\{Y_t\}$ is defined by $Y_t = \zeta_t + \beta_1 \zeta_{t-1} + \beta_2 \zeta_{t-2} + \ ... + \beta_q \zeta_{t-q}$ is called a moving average process of order q and is denoted by $MA(q)$. If the process $\{Y_t\}$ is given by $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ ... \ + \alpha_p Y_{t-p} + \zeta_t$ 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 $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ ... \ + \alpha_p Y_{t-p} + \zeta_t + \beta_1 \zeta_{t-1} + \beta_2 \zeta_{t-2} + \ ... \ + \beta_q \zeta_{t-q}$, where, Y_t is the original series, for every t, we assume that ζ_t is independent of $Y_{t-1}, Y_{t-2}, ..., Y_{t-p}$. 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. An $ARIMA(p,1,q)$ process is defined as, $W_t = \alpha_1 W_{t-1} + \ ... \ + \alpha_p W_{t-p} + \zeta_t + \beta_1 \zeta_{t-1} + \ ... \ + \beta_q \zeta_{t-q}$, where, $W_t = Y_t - Y_{t-1}$.

Box-Jenkins Method

The influential work of Box-Jenkins (Box and Jenkins, 1970 [12]) 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.

Ljung-Box Test

Ljung-Box (Ljung and Box, 1978 [13]) 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 number of observations used to estimate the model. This statistic Q^* approximately follows the chi-square distribution with $(k-q)$ df, 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 estimated 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. There are many summary statistics available in literature for evaluating the forecast errors of any Time Series or Econometric model. Thus, here an attempt is made to identify the best models for Pineapple 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.

Mean Percent Forecast Error (MPFE): Mean Percent Forecast Error (MPFE) is defined as, $MPFE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{Y_t^a - Y_t^f}{Y_t^a}\right)$, 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 (Theil, 1966 [14]) 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.

Results and Discussion

In order to make forecasting a time series it is necessary to check the time series is stationary or not first. 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 $\Pr\left(|\tau| \geq \text{-}5.4682\right) < 0.01$, Phillips-Perron (PP) test with $\Pr\left(|\tau| \geq \text{-}38.8115\right) < 0.01$ and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) with $\Pr\left(|\tau| \geq 0.0277\right) > 0.1$ at 5% level of significance adequately declared that the data series is stationary and suggest that there is no unit root The graphical representations of the original and second differenced series are presented in Figure 1(a), (b).

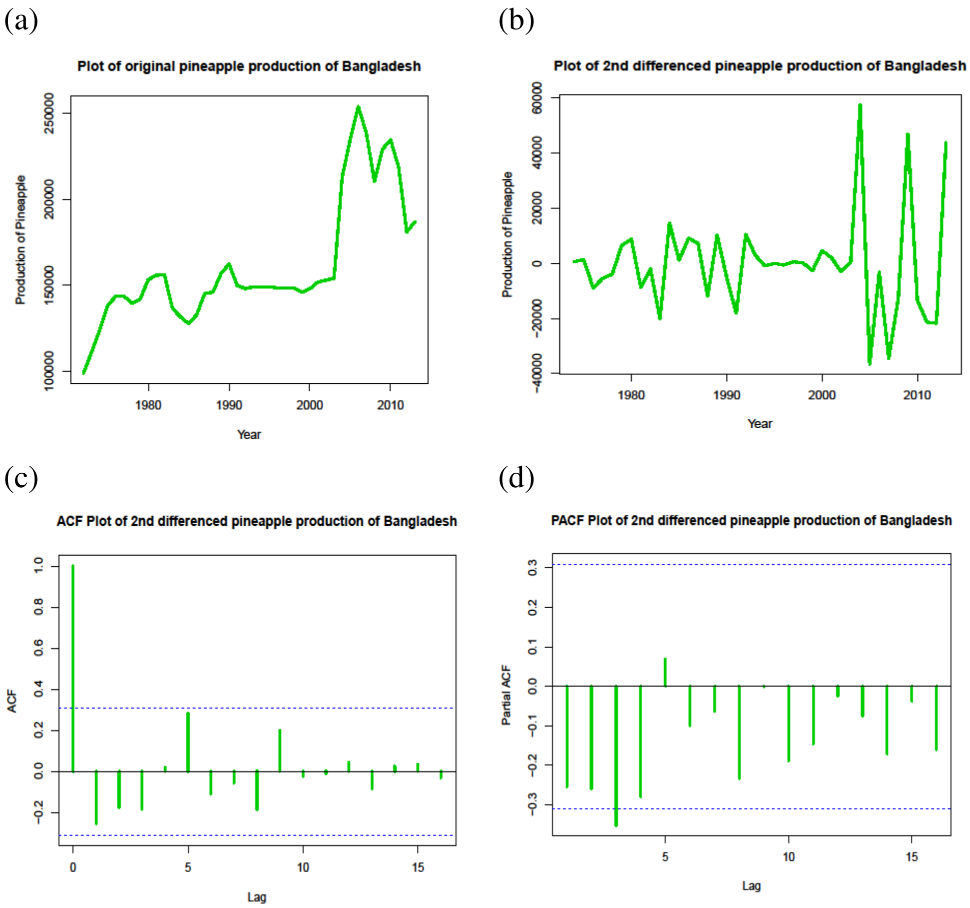

Figure 1. (a) Time series (original series) plot, (b) Time series (2nd differenced) plot (c) ACF and (d) PACF of 2nd differenced Pineapple production in Bangladesh.

It is clear that the yearly Pineapple production in Bangladesh initially shows an increasing trend. However, after the year 1982 the production is fluctuated up to the year 1990 and over the periods 1991 to 2003 there was almost equal production. After the year 2003 there was a dramatic increase and reached a peak at 2006 and then again there was a gradually decreasing trend i.e., the variance is unstable which leads the Pineapple production data series is not stationary (Figure 1(a)). However, it is clear that the second differenced

Pineapple production data series shows stable variance which indicates 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 1(c)) and exponentially decay PACF (Figure 1(d)) indicates an autoregressive moving average process. The PACF with significant spike at lag 1 and ACF with no significant spike suggest that only first order autoregressive model may be appropriate for the pineapple production in Bangladesh. However, using the tentative procedure, it is clear that ARIMA(0,2,1) model with $\mathrm{AIC} = 889.82$, $\mathrm{AIC}_{\mathrm{C}} = 890.14$ and $\mathrm{BIC} = 893.19$ is the best selected model for forecasting the Pineapple production in Bangladesh. The estimates of the parameters of the fitted ARIMA(0,2,1) model are shown in Table 1. The value of the most useful "forecasting criteria" of the fitted ARIMA(0,2,1) model are $RMSPE = 0.07671859$, $MPFE = -0.01642267$, and $TIC = 0.04350895$.

Table 1. Summary statistics and forecasting criteria of the fitted ARIMA (0,2,1) model

Coefficients	Estimates	Std.Error	t-value	p-value
ma1	-1.000	0.0944	-10.59322	0.02995967

Several graphical representations of the residuals for the fitted ARIMA(0,2,1) model are presented in Figure 2, suggest that there is no significant pattern and hence there is no autocorrelation among the residuals. Also, the "Box-Pierce" test with $\Pr\left(\left|\chi_1^2\right| \geq 2.5731\right) = 0.1087$ and the "Box-Ljung" test with $\Pr\left(\left|\chi_1^2\right| \geq 2.7614\right) = 0.09657$ at 5% level of significance suggest 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 and suggests that the residuals of the fitted ARIMA(0,2,1) model are normally distributed. The Histogram with Normal Curve of the residuals of the fitted ARIMA(0,2,1) model is given in Figure 2. Therefore, it is clear that the fitted ARIMA(0,2,1) model is the best fitted model and adequately used to forecast the Pineapple production in Bangladesh.

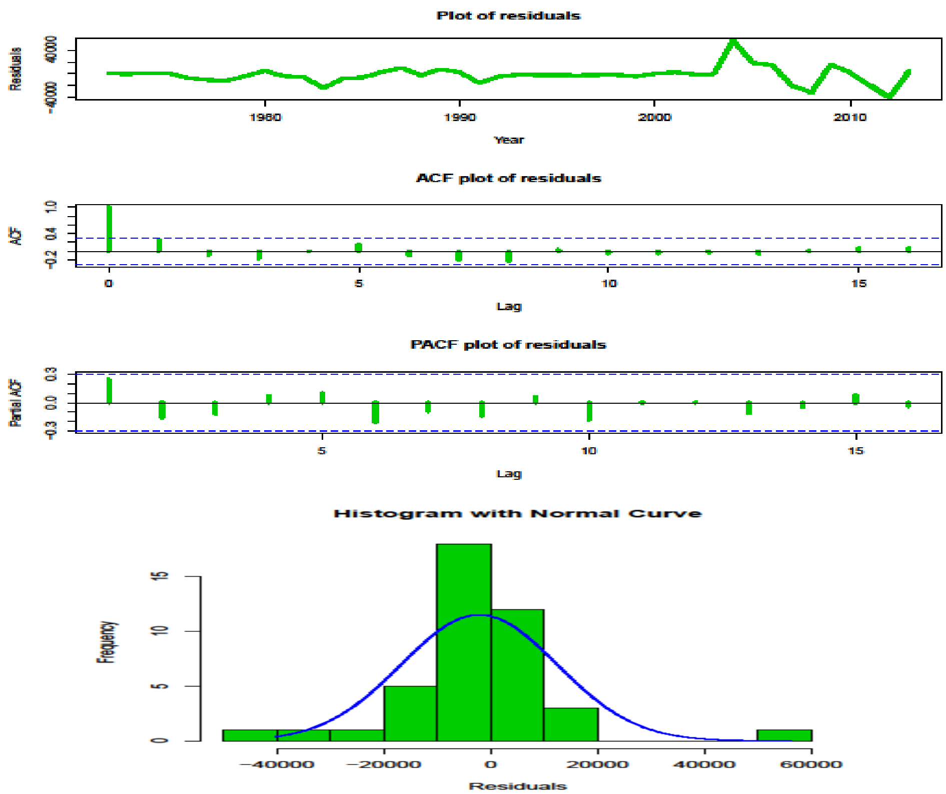

Figure 2. Several plots of residual plots and Histogram with Normal Curve.

By using the best fitted model ARIMA(0,2,1), the forecasted Pineapple production and 95% confidence level for twelve years are shown in Table 2. The graphical comparison of the original series and the forecast series is shown in Figure 3. It is observed that the forecast series (blue-color) fluctuated from the original series (dark-green-color) with a very small amount that is it shows the production in same manner of the original series (Figure 3). Therefore, the forecasted series is really better representation of the original Pineapple production series in Bangladesh. Also, in the forecasting plot, in sample forecasting part shows an upward trend and similarly the out sample forecasting part also shows an upward trend. That is, forecasting pineapple productions may be good.

Table 2. Forecasted Pineapple production (tonnes) in Bangladesh

Year	Forecasted	LCL	UCL
2014	189142.2	159637.30	218647.0
2015	191284.3	149064.32	233504.3
2016	193426.5	141119.94	245733.0
2017	195568.6	134487.76	256649.5
2018	197710.8	128665.68	266755.8
2019	199852.9	123400.12	276305.7
2020	201995.1	118542.74	285447.4
2021	204137.2	113998.41	294276.0
2022	206279.4	109702.07	302856.7
2023	208421.5	105607.14	311235.9
2024	210563.7	101678.99	319448.4
2025	212705.8	97891.17	327520.5

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

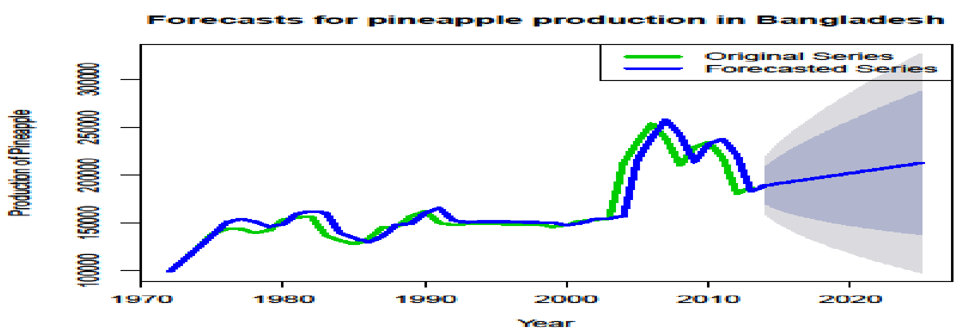

Figure 3. Comparison between the original and forecasted Pineapple production in Bangladesh.

Conclusion

The best selected Box-Jenkins ARIMA model for forecasting the Pineapple 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 the Pineapple 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 Pineapple production in Bangladesh.

Reference

[1] Baruwa, O. I., Profitability and constraints of pineapple production in Osun State, Negeria, *Journal of Horticultural Research*, 2013, **21(2),** 59. DOI: 10.2478/johr-2013-0022.

[2] Bartholomew, D. P., Paul, R. E. and Rorbach, K. G., The pineapple 'Botany, Production and Uses', University of Hawaii Manoa Honolulu, USA, 2003. Available: http://bookshop.cabi.org/ Uploads/Books/ PDF/ 978085995038/

[3] Golam Samdani, *Pineapple farming: Prospect high*, The Independent (Daily News Paper), Sunday, 03 August, 2014.

[4] Hasan, S. S., Ali, M. A. and Khalil, M. I., Impact of Pineapple Cultivation on the Increased Income of Pineapple Growers, *The Agriculturists,* 2010, **8(2)**, 50.

[5] Hemalatha, R. and Anbuselvi, S., Physicochemical constituents of pineapple pulp and waste, *J. Chem. Pharm. Res.*, 2013, **5(2)**, 240.

[6] Debnath, P., Dey. P., Chanda, A. and Bhakta, T., A Survey on Pineapple and its medicinal value, *Scholars Academic Journal of Pharmacy*, 2012, **1(1),** 24.

[7] De Carvalho, L. M. J., I. M. de Castro and C. A. B. da Silva, A study of retention of sugars in the process of clarification of pineapple juice (*Ananas comosus*, L. Merril) by micro- and ultra-filtration, *J. Food Eng.*, 2008, **87**, 447.

[8] Sabahelkhier, K. M., Hussain, A. S. and Ishag, K. E. A., Effect of maturity stage on protein fractionation, in vitro protein digestibility and anti-nutrition factors in pineapple (Ananas comosus) fruit grown in Southern Sudan, *Afr. J. Food Sci.*, 2010, **4(8),** 550.

[9] Md. Farid Hossain, Shaheen Akhtar, Mustafa Anwar, Nutritional Value and Medicinal Benefits of Pineapple, *International Journal of Nutrition and Food Sciences,* 2015, **4(1)**, 84. doi: 10. 11648/j.ijnfs. 20150401.22

[10] Idrees, *Top 10 Pineapple Producing Countries*, 15 August, 2014. Accessed on 07 March, 2015. Available at http://worldknowing. com/ top-10-pineapple-producing-countries/

[11] Rahmatullah, M., Mukti, I. J., Haque, A. K. M. F., Mollik, M. A. H., Parvin, K., Jahan, R., Chowdhury, M. H. and Rahman, T., An ethno botanical survey and pharmacological evaluation of medicinal plants used by the Garo tribal community living in Netrakona district, Bangladesh, *Adv. Nat. Appl. Sci.*, 2009, **3(3),** 402.

[12] Box George and Jenkins Gwilym, *Time series analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

[13] Ljung, G. M. and Box, G. E. P., On a Measure of a Lack of Fit in Time Series Models, *Biometrika*, 1978, **65(2)**, 297, doi:10. 1093/ biomet /65.2.297.

[14] Theil, H., *Applied Economic Forecasting*, North-Holland Publishing Company, Amsterdam, 1966.