Forecasting Major Fruit Crops Productions in Bangladesh using Box-Jenkins ARIMA Model

Mohammed Amir Hamjah
B.Sc. (Honors), MS (Thesis) in Statistics
Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh
Email: arif.stat.sust@hotmail.com.

Abstract
Bangladesh has a large agrarian base country where 77% of total population is living in the rural areas and 90 % of the rural population directly related with agriculture. Banana, Guava, Papaya, Jackfruit, Pineapple, Mango etc. are the major fruits crops in Bangladesh. The main objective of this study is to fit the Box-Jenkins Auto-Regressive Integrated Moving Average, that is, ARIMA Model to forecast the different types of major fruits productions in Bangladesh. From the study, it is 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. From the comparison between the original series and forecasted series which are also shown the same manner indicating fitted model are statistically well behaved to forecast fruits productions in Bangladesh.
Keywords: Fruits Productions, Forecasting, ARIMA Model, Bangladesh.

1. Introduction

Bangladesh has a large agrarian base country with 77% of total population is living in the rural areas and 90 percent of the rural population directly related with agriculture. *Agriculture* is the single largest producing sector of the economy since it comprises about 18.6% (data released on November, 2010) of the country's *GDP* and employs around 45% of the total labor force. Different types of fruits are produced in Bangladesh. Banana, Guava, Papaya, Jackfruit, Pineapple, Mango etc. are the major fruits crop in Bangladesh.
Mango is the king of fruits. It is very delicious and attractive fruit. The mango grows in almost all parts of Bangladesh. But high percentage of mangos grow well in Rajshahi and many south-east other parts of the country. The position of mango is 1st in terms of area and 2nd in production among the fruits grown in Bangladesh. Mango shares 31.22% of the area and 24.38% production fruit crops in Bangladesh.
Banana is the year round fruit of Bangladesh. Banana is found in any season. The origin of banana is not still clearly identified by the scientists. Its cultivation is distributed through the warmer countries. Some important countries are Brazil, India, Thailand, Uganda, Colombia, Nigeria and Srilanka. The banana occupies 2nd position in terms of area and production. Banana is most important fruits in Bangladesh. It grows well in Noakhali, Dhaka, Kustia, Jessore, Barishal and Rajshahi.
Guava is called the apple of the tropics, is one of the most common fruit in Bangladesh. It is native to tropical America. In Bangladesh, it is grown all over the country but commercially grown in Barishal, Pirozpur, Jhalkathi, Comilla and Chittagong. Guava is a rich source of vitamin C, a protein. Firm ripe fruit is richer in vitamin C than those of fully ripe or over ripe. It is mainly used in many countries as a desert. However the most commercial use of guava is for jelly preparation.

2. Objective of the Study

The main objective of this study is to develop an ARIMA model for forecasting different types of Fruit crops productions in the Bangladesh. The specific objective of the study is to develop an Autoregressive Integrated Moving Average (ARIMA) model for different seasonal fruits productions such as Banana, Guava, and Mango in Bangladesh.

3. Review of Literature

There are a lot of study have been done by the researcher to fit an ARIMA model in the agriculture sector in all over the world for different types of agricultural crops. ARIMA model is used in different agriculture sector to forecast agricultural productions. The relevant work for forecasting by using Box-Jenkins (1970) ARMA model, from which we get the idea about forecasting techniques for different types of agricultural productions forecasting such as Goodwin and Ker (1998) added new dimensions to the evolution of this literature. They introduced a univariate filtering model, an ARIMA (0, 1,2) to best represent crop yield series. Mohammed Amir Hamjah (2014) has used Box-Jenkins ARIMA model to forecast different types of Seasonal rice productions in Bangladesh. From his study, it was found that the best selected ARIMA model for Aus productions is ARIMA (2,1,2), for Aman productions is ARIMA (2,1,2) and, for Boro productions is ARIMA (1,1,3). Rachana *et al.* (2010), used ARIMA models to forecast pigeon pea production in India. Badmus and Ariyo ARIMA (1,1,1) and ARIMA (2,1,2) for cultivation area and production resrespectively. Falak and Eatzaz (2008), analyzed future prospects of wheat production in Pakistan. Applying ARIMA model. Hossian et. al. (2006) forecasted three different varieties of pulse prices namely motor, mash and mung in Bangladesh with monthly data from Jan 1998 to Dec 2000; Wankhade et al. (2010) forecasted pigeon pea production in India with annual data from 1950-1951 to 2007-2008; Mandal (2005) forecasted sugarcane production. Rahman (2010) fitted an ARIMA model for forecasting Boro rice production in Bangladesh. M. A. Awal and M.A.B. Siddique's study was carried out to estimate growth pattern and also examine the best ARIMA model to efficiently forecasting Aus, Aman and Boro rice productions in Bangladesh. Nasiru Suleman and Solomon Sarpong (2011) made a paper with the title "Forecasting Milled Rice Production in Ghana Using Box-Jenkins Approach". The analysis is revealed that ARIMA (2, 1, 0) is the best model for forecasting milled rice production.

4. Data Source and Used Software

The crop data-sets are available from Bangladesh Agricultural Ministry's website named as *www.moa.gov.bd.* This analysis has completely done by statistical programming based open source Software named as **R** with the version **2.15.1.** The additional library packages used for analysis are **forecas**t and **tseries**.

5. Methodology

A time series is a set of numbers that measures the status of some activity over time. It is the historical record of some activity, with measurements taken at equally spaced intervals with a consistency in the activity and the method of measurement.

5.1. Moving Average Processes:

Moving average models were first considered by Slutsky (1927) and Wold (1938). The Moving Average Series can be written as

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \ \dots - \theta_q e_{t-q} \qquad (1)$$

We call such a series a moving average of order **q** and abbreviate the name to **MA(q)**. where, Y_t is the original series and e_t is the series of errors.

5.2. Auto-Regressive Process:

Yule (1926) carried out the original work on autoregressive processes. Autoregressive processes are as their name suggests regressions on themselves. Specifically, a p^{th-} order autoregressive process {Yt} satisfies the equation

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \cdots + \Phi_p Y_{t-p} + e_t \qquad (2)$$

The current value of the series Y_t is a linear combination of the p most recent past values of itself plus an "innovation" term e_t that incorporates everything new in the series at time t that is not explained by the past values. Thus, for every t, we assume that e_t is independent of $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots Y_{t-q}$.

5.3. Autoregressive Integrated Moving Average (ARMIA) model

The Box and Jenkins (1970) procedure is the milestone of the modern approach to time series analysis. Given an observed time series, the aim of the Box and Jenkins procedure is to build an ARIMA model. In particular, passing by opportune preliminary transformations of the data, the procedure focuses on Stationary processes.

In this study, it is tried to fit the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model. This model is the generalized model of the non-stationary ARMA model denoted by ARMA(p,q) can be written as

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \dots \dots \dots . -\theta_q e_{t-q} \qquad (3)$$

Where, Y_t is the original series, for every t, we assume that e_t is independent of $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$.

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

$$W_t = \Phi_1 W_{t-1} + \Phi_2 W_{t-2} + \cdots + \Phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \dots \dots \dots . \theta_q e_{t-q} \qquad (4)$$

Box and Jenkins procedure's steps

i. **Preliminary analysis:** create conditions such that the data at hand can be considered as the realization of a stationary stochastic process.
ii. **Identification**: specify the orders p, d, q of the ARIMA model so that it is clear the number of parameters to estimate. Recognizing the behavior of empirical autocorrelation functions plays an extremely important role.
iii. **Estimate:** efficient, consistent, sufficient estimate of the parameters of the ARIMA model (maximum likelihood estimator).
iv. **Diagnostics:** check if the model is a good one using tests on the parameters and residuals of the model. Note that also when the model is rejected, still this is a very useful step to obtain information to improve the model.
v. **Usage of the model:** if the model passes the diagnostics step, then it can be used to interpret a phenomenon, forecast.

5.4. Procedure of Maximum Likelihood Estimation (MLE) Method

The advantage of the method of maximum likelihood is that all of the information in the data is used rather than just the first and second moments, as is the case with least squares. Another advantage is that many large-sample results are known under very general conditions. For any set of observations, $Y_1, Y_2, \ldots, Y_n$ time series or not, the likelihood function L is defined to be the joint probability density of obtaining the data actually observed. However, it is considered as a function of the unknown parameters in the model with the observed data held fixed. For ARIMA models, L will be a function of the ϕ's, θ's, μ, and σ_e^2 given the observations $Y_1, Y_2, \ldots, Y_n$. The maximum likelihood estimators are then defined as those values of the parameters for which the data actually observed are most likely, that is, the values that maximize the likelihood function.

We begin by looking in detail at the AR (1) model. The most common assumption is that the white noise terms are independent, normally distributed random variables with zero means and common variance, σ_e^2. The probability density function (pdf) for each e_i is then

$$(2\pi\sigma_e^2)^{-(1/2)} exp^{\left(-\left[\frac{e_2^2}{\sigma_e^2}\right]\right)}, for -\infty < e_t < \infty$$

and, by independence, the joint pdf for $e_1, \ldots, e_n$ is

$$(2\pi\sigma_e^2)^{-(\{n-1\}/2)} exp^{\left(-\left[\Sigma\frac{e_2^2}{\sigma_e^2}\right]\right)}, for -\infty < e_t < \infty \qquad (5)$$

Now consider

$$\begin{aligned} Y_2 - \mu &= \phi(Y_1 - \mu) + e_2 \\ Y_3 - \mu &= \phi(Y_2 - \mu) + e_3, .. \qquad (6) \\ \ldots\ldots\ldots\ Y_n - \mu &= \phi(Y_{n-1} - \mu) + e_n \end{aligned}$$

If we condition on $Y_1 = y_1$ Equation (6) defines a linear transformation between $e_2, e_3, \ldots, e_n$ and $Y_2, \ldots, Y_n$ (with Jacobian equal to 1). Thus the joint pdf of $Y_2, \ldots, Y_n$ given $Y_1 = y_1$ can be obtained by using Equation (6) to substitute for the e's in terms of the Y's in Equation (5). Thus we get

$$f(Y_2, \ldots, Y_n | Y_1 = (2\pi\sigma_e^2)^{-(\{n-1\}/2)} exp\left[\frac{-1}{2\sigma_e^2}\sum_2^n\{(Y_t - \mu) - \phi(Y_{t-1} - \mu)\}^2\right] \qquad (7)$$

Now consider the (marginal) distribution of Y_1. It follows from the linear process representation of the AR(1) process that Y_1 will have a normal distribution with mean μ and variance $\sigma_e^2/(1 - \phi^2)$. Multiplying the conditional pdf in equation (7) by the marginal pdf of Y_1 gives us the joint pdf of $Y_1, Y_2, \ldots, Y_n$ that we require. Interpreted as a function of the parameters φ, μ and σ_e^2, the likelihood function for an AR(1) model is given by

$$L(\phi, \mu, \sigma^2) = (2\pi\sigma_e^2)^{-\left(\frac{\{n-1\}}{2}\right)}(1 - \phi^2) exp\left[\frac{-1}{2\sigma_e^2}S(\phi, \mu)\right] \qquad (8)$$

where, $S(\phi, \mu) = \sum_2^n\{(Y_t - \mu) - \phi(Y_{t-1} - \mu)\}^2 + (1 - \phi^2)(Y_1 - \mu)$

The function S(φ,μ) is called the unconditional sum-of-squares function. As a general rule, the logarithm of the likelihood function is more convenient to work with than the likelihood itself. For the AR (1) case, the log-likelihood function, denoted by (ϕ, μ, σ^2), is given by

$$l(\phi, \mu, \sigma^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma_e^2) + \frac{1}{2}log(1 - \phi^2) - \frac{1}{2\sigma_e^2}S(\phi, \mu) \qquad (9)$$

For given values of φ and μ, $l(\phi, \mu, \sigma^2)$ can be maximized analytically with respect to σ_e^2 in terms of the yet to be determined estimators of φand μ. We obtain

$$\widehat{\sigma_e^2} = \frac{S(\phi, \mu)}{n} \qquad (10)$$

As in many other similar contexts, we usually divide by n −2 rather than n (since we are estimating two parameters, ϕ and μ) to obtain an estimator with less bias. For typical time series sample sizes, there will be very little difference.

Consider now the estimation of φ and μ. A comparison of the unconditional Sum of Squares function S(ϕ,μ) with the conditional Sum of Squares function Sc(ϕ, μ) $= \sum_2^n(Y_t - \mu)\ \{-\phi(Y_{t-1} - \mu)\}^2$ of AR process reveals one simple difference. Since $S_c(\phi,\mu)$ involves a sum of n −1 components, whereas $(1 - \phi^2)(Y_1 - \mu)^2$ does not involve n, we shall have $S(\phi, \mu) \approx Sc(\phi, \mu)$. Thus the values of φ and μ that minimize S(ϕ,μ) or $S_c(\phi,\mu)$ should be very similar, at least for larger sample sizes. The effect of the rightmost term in Equation (10) will be more substantial when the minimum for ϕ occurs near the stationarity boundary of ±1.

5.5. Diagnostic Tests of Residuals

5.5.1. Jarque-bera test

We can check the normality assumption using Jarque-Bera (1978) test, which is a goodness of fit measure of departure from normality, based on the sample kurtosis(k) and skewness(s). The test statistics Jarque-Bera (JB) is defined as

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right) \sim \chi^2_{(2)}$$

Where n is the number of observations and k is the number of estimated parameters. The statistic JB has an asymptotic chi-square distribution with 2 degrees of freedom, and can be used to test the hypothesis of skewness being zero and excess kurtosis being zero, since sample from a normal distribution have expected skewness of zero and expected excess kurtosis of zero.

5.5.2. Ljung-Box test

Ljung-Box 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 $H_0 : \rho_1(e) = \rho_2(e) = \ldots\ldots = \rho_k(e) = 0$ is tested with the Box-Ljung statistic $Q^* = N(N+1)\sum_{i=1}^{k}(N-k)\,\rho^2_k(e)$

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.

The most useful forecast evaluation criteria are Mean Square Error (MSE) proposed by Ou and Wang (2010), Root Mean Square Error (RMSE) proposed by Ou and Wang (2010), Mean Absolute Error(MAE), Root Mean Square Error Percentage(RMSPE), Mean Absolute Percentage Error (MAPE) proposed by Sutheebnjard and Premchaiswadi (2010).

6. Results and Discussions

6.1. ARIMA Modeling of Mango Production

At first, it is very essential to find out for which order of difference of the time sequence Mango production data satisfies the sationarity conditions. We used Log-transformation on the Mango production data-set to avoid the violation of Normality assumption. It can be theoretically and graphically checked by using Dickey-Fuller unit root test and plotting time series of Log-transformed Mango production respectively. From Dickey-Fuller unit root test is , it is found that stationarity condition satisfied at the difference order one with the p-value = 0.01, which strongly suggests that there is no unit root at the first order difference of Log-transformed Mango production at 1% level of significance. The graphical stationarity test using ACF and PACF are shown in the Figure-1.

From the Figure-1, it is transparent that the original Log-transformed Mango production series does not show constant variance but first order differenced series shows a more stable variance than the original series. Again, from the ACF and PACF, it is clear that there are some significant spikes in the ACF plot indicating existence of Moving Average effects on the original Mango production series, that is, the series is not stationary. At the same time, from the ACF and PACF of first order differenced series, it is obvious that there is no significant spike which also tell us that the series is stationary with first order difference; and implies that there is no significant effects of Autoregressive and Moving Average order at first order difference series, indicating stable variance.

Now, from the formal test (Dickey-Fuller Unit Root Test) and graphical representation of Mango production, it is obvious that at the difference order one the series become stationary.

From the tentative order analysis, the best selected ARIMA model to forecast the Mango productions in Bangladesh is ARIMA (2,1,3) with the AIC = - 60.42 and BIC = - 51.63. The parameter estimates of the fitted ARIMA (2, 1, 3) are given in the Table-1.

From the Table-1, it is obvious that first and second order Auto-regressive Lag and first, second and third order Moving Average Lag have statistically significant effects on Log-transformed Mango production at 20% level of significance.

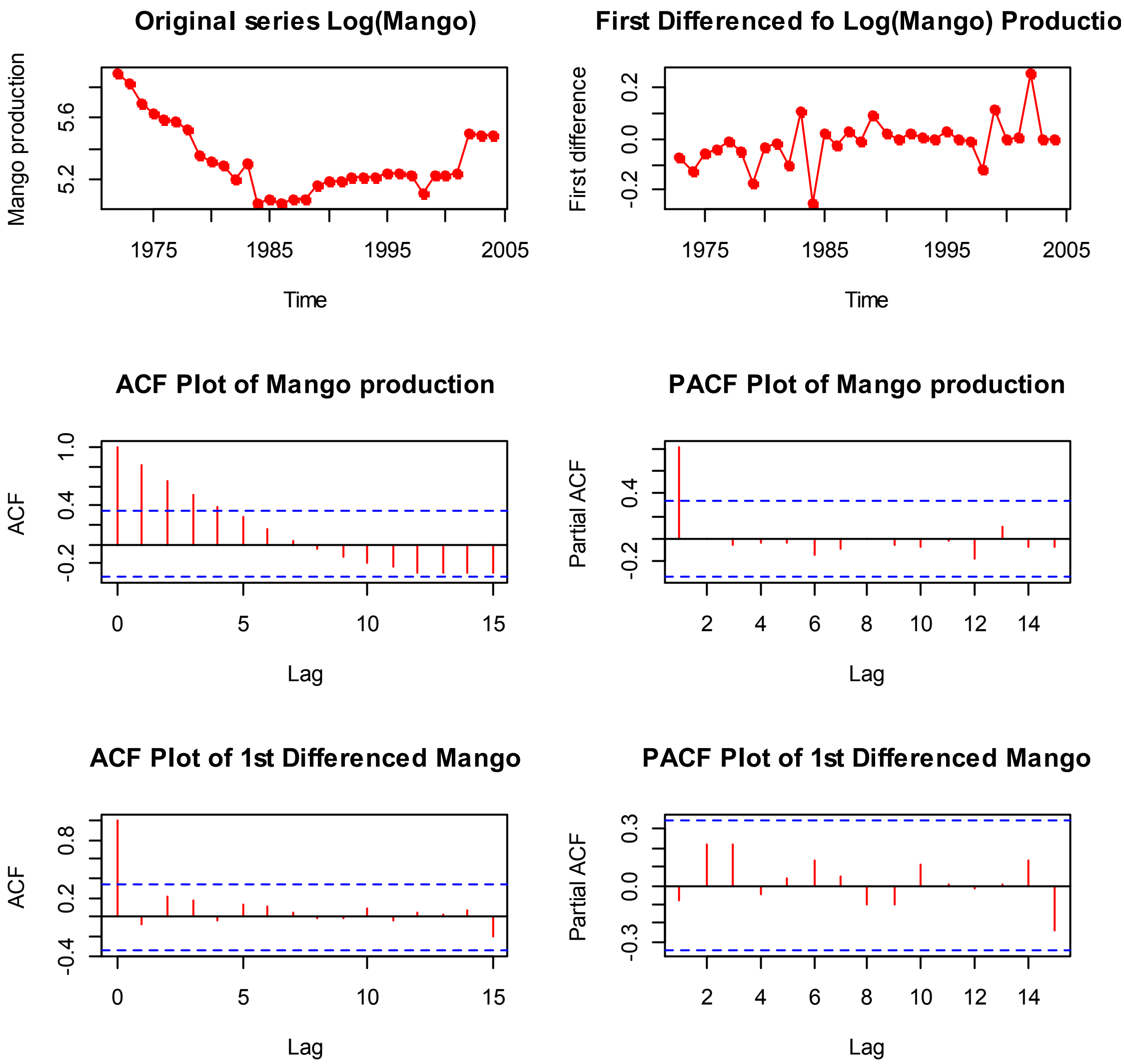

Figure-1: Graphically Stationarity Checking for Mango Productions

Table-1: Summary Statistics of the ARIMA Model for Mango Production				
Coefficients	**Estimates**	**Std. Error**	**t-value**	**p-value**
ar1	-0.6818	0.2392	-2.8507	0.1074
ar2	-0.6323	0.178	-3.5523	0.0873
ma1	0.6194	0.2591	2.3909	0.1261
ma2	1.0928	0.1559	7.0081	0.0451
ma3	0.3658	0.2598	1.4077	0.1966

To check Autocorrelation assumption, “Box-Ljung” test is used. From the test, it is found that the $\Pr(|\chi^2_{(1)}| \geq 0) = 0.997$, which strongly suggests that we may accept the assuption that there is no autocrrelation among the residuals of the fitted ARIMA model at 5% level of significance. Again, to check the normality assumption, “Jarque-Bera” test is used, from which, we get the $\Pr(|\chi^2_{(2)}| \geq 1.3353) = 0.5129$, which refers to accept the norality assumpytion that the residuals are from normal distribution. Graphical Residuals Diagnostics are shown in the Figure-2.

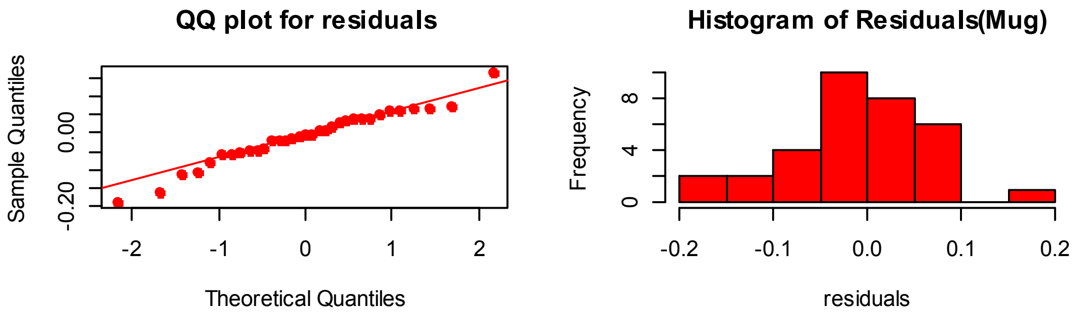

Figure-2: Graphical Residuals Diagnostics Checking for ARIMA Model of Mango Production

From the Figure-2, it is clear that almost all of the points are closed to the Q-Q line or on the Q-Q line, which indicates that residuals are normally distributed of the Mango production model. At the time, from the Histogram of the residuals of Mango production model, it is obvious that residuals are normally distributed.

Finally, considering all of the Graphical and Formal test, it is obvious that our fitted model ARIMA (2,1,3) is the best fitted model to forecast Mango production in the Bangladesh

6.2. ARIMA Modeling of Banana Production

To check stationarity of time sequence Banana Production series, Dickey-Fuller Unit Rot test is used. From Dickey-Fuller unit root test is , it is found that stationarity condition satisfied at the difference order one with the p-value = 0.01, which strongly suggests that there is no unit root at the first order difference of Banana production at 1% level of significance. The graphical stationarity test using ACF and PACF are shown in the Figure-1.

From the Figure-3, it is transparent that the original Banana production series does not show constant variance but first order differenced series shows a more stable variance than the original series. Again, from the ACF and PACF, it is clear that there are significant spikes in the ACF plot indicating existence of Moving Average effects on the original Banana production series, that is, the series is not stationary. At the same time, from the ACF and PACF of first order differenced series, it is obvious that there is no significant spike which also tell us that the series is stationary with first order difference; and implies that there is no significant effects of Autoregressive and Moving Average order at first order differenced series, indicating stable variance.

From the formal test (Dickey-Fuller Unit Root Test) and graphical representation (ACF and PACF) of Banana production, it is obvious that at the difference order one the series become stationary.

From the tentative order analysis, the best selected ARIMA model to forecast the Banana productions in Bangladesh is ARIMA (3,1,2) with the AIC = 311.39 and BIC = 320.18. The parameter estimates of the fitted ARIMA (3, 1, 2) for Banana productions are given in the Table-2.

From the Table-2, it is obvious that first, second and third order Auto-regressive Lag and first and second order Moving Average Lag have statistically significant effects on Banana production. Here, third order Auto-Regressive Lag have statistically significance effects on Banana productions at 28% level of significance and we accept this, because ARIMA(3,1,2) is the best model to forecast Banana productions in Bangladesh and satisfy the all other assumptions of a fitted ARIMA model for the purpose of forecasting Banana productions in Bangladesh.

From the "Box-Ljung" test of Autocorrelation assumption checking, it is found that the $\Pr(|\chi^2_{(1)}| \geq 0.0004)$ = 0.9848, which strongly suggests that we may accept the assumption that there is no autocorrelation among the residuals of the fitted ARIMA model at 5% level of significance. Again, from the "Jarque-Bera" normality test, we get the $\Pr(|\chi^2_{(2)}| \geq 3.4432)$ = 0.1788, which refers to accept the normality assumption that the residuals are from normal distribution. Graphical Residuals Diagnostics are shown in the Figure-4

From the Figure-4, it is clear that almost all of the points are closed to the Q-Q line or on the Q-Q line, but some of the points are slightly scattered from the QQ line. At the same time, from the formal test, we observed that residuals are normally distributed. So, it can be said that residuals are normally distributed of the Banana production model. Again, from the Histogram of the residuals of Banana production model, it is obvious that residuals are slightly negatively skewed normally distributed.

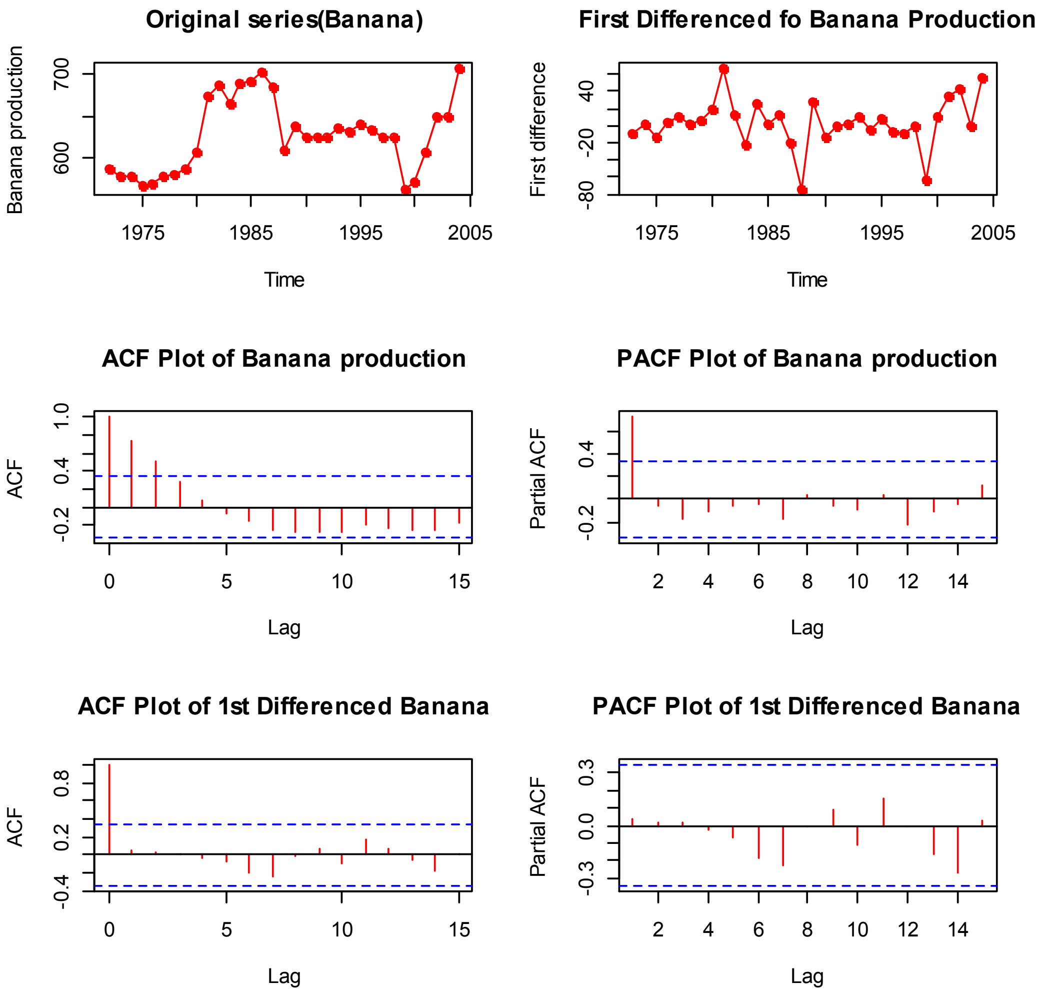

Figure-3: Graphically Stationarity Checking of Banana Production Series

Table-2: Summary Statistics of the ARIMA Model for Banana Production

Coefficients	Estimates	Std. Error	t-value	p-value
ar1	1.166	0.2063	5.6515	0.0557
ar2	-0.8332	0.2443	-3.41	0.0908
ar3	0.1868	0.2263	0.8253	0.2804
ma1	-1.1847	0.1563	-7.5819	0.0417
ma2	1	0.1338	7.4723	0.0423

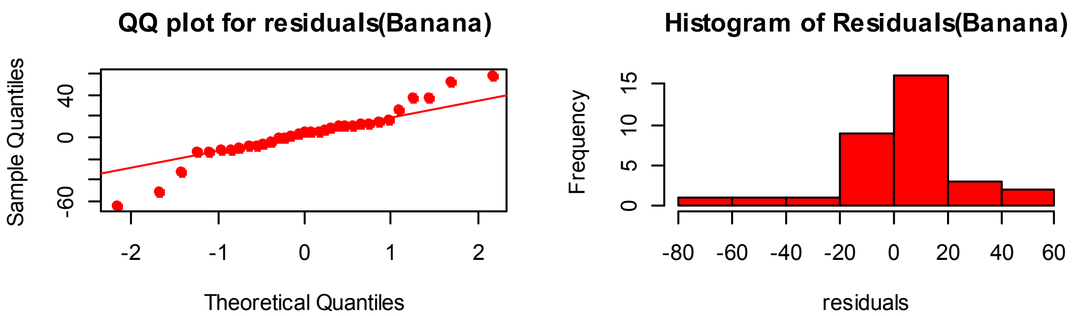

Figure-4: Graphically Residuals Diagnostic Checking for ARIMA Model of Banana Production

Finally, considering all of the Graphical and Formal test, it is obvious that our fitted model ARIMA (2, 1,3) is the best fitted model for forecasting Banana production in the Bangladesh

6.3. ARIMA Modeling of Guava Production

At first, it is very essential to find out for which order of difference of the time sequence Guava production data satisfies the sationarity conditions. We take Log-transformation on Guava production to avoid the violation of Normality assumption. It can be theoretically and graphically checked by using Dickey-Fuller unit root test and plotting time series of Log-transformed Guava production respectively. From Dickey-Fuller unit root test is, it is found that stationarity condition satisfied at the difference order one with the p-value < 0.01, which strongly suggests that there is no unit root at the first order difference of Mango production at 1% level of significance. The graphical stationarity test using ACF and PACF are shown in the Figure-1.

From the Figure-5, it is obvious that the original Log-transformed Guava production series does not show constant variance and it shows an upward trend but first order differenced series shows stable variance by removing the trend effects on Guava productions. Again, from the ACF and PACF, it is clear that there are significant spikes in the ACF plot indicating existence of Moving Average effects on the original Guava production series, that is, the series is not stationary. At the same time, from the ACF and PACF of first order differenced series, it is obvious that there is no significant spike which also tell us that the series is stationary with first order difference; and implies that there is no significant effects of Autoregressive and Moving Average order at first order difference series, indicating stable variance.

Now, from the formal test (Dickey-Fuller Unit Root Test) and graphical representation (ACF and PACF) of Guava production, it is obvious that at the difference order one the series become stationary.

From the tentative order analysis, the best selected ARIMA model to forecast the Guava productions in Bangladesh is ARIMA (1,1,2) with the AIC = -13.27 and BIC = -7.16. The parameter estimates of the fitted ARIMA (1,1,2) are given in the Table-3.

From the Table-3, it is obvious that first order Auto-regressive Lag has statistically significant effects on Guava productions; and first and second order Moving Average Lag have statistically significant effects on Guava production at 6% level of significance.

From the "Box-Ljung" test of Autocorrelation assumption checking, it is found that the $\Pr(|\chi^2_{(1)}| \geq 0.00005) = 0.9816$, which strongly suggests that we may accept the assumption that there is no autocorrelation among the residuals of the fitted ARIMA model at 5% level of significance. Again, from the "Jarque-Bera" normality test, we get the $\Pr(|\chi^2_{(2)}| \geq 0.4747) = 0.1788$, which refers to accept the normality assumption that the residuals are normally distributed. Graphical Residuals Diagnostics are shown in the Figure-6.

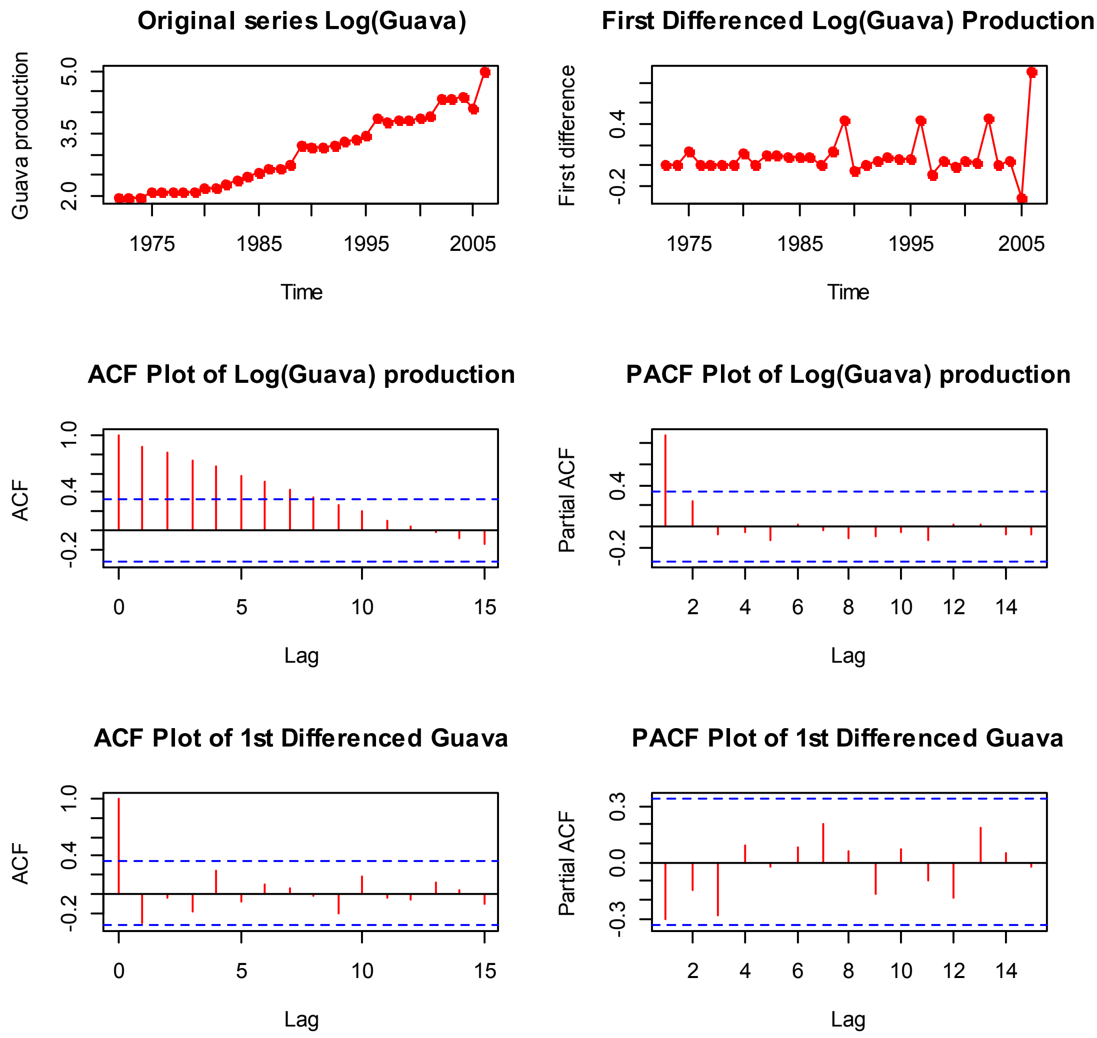

Figure-5: Graphically Stationarity Checking of Guava Production Series

Table-3: Summary Statistics of the ARIMA Model for Guava Production				
Coefficients	**Estimates**	**Std. Error**	**t-value**	**p-value**
ar1	0.9433	0.0578	16.3097	0.0195
ma1	-1.7967	0.1883	-9.5435	0.0332
ma2	1	0.2043	4.894	0.0642

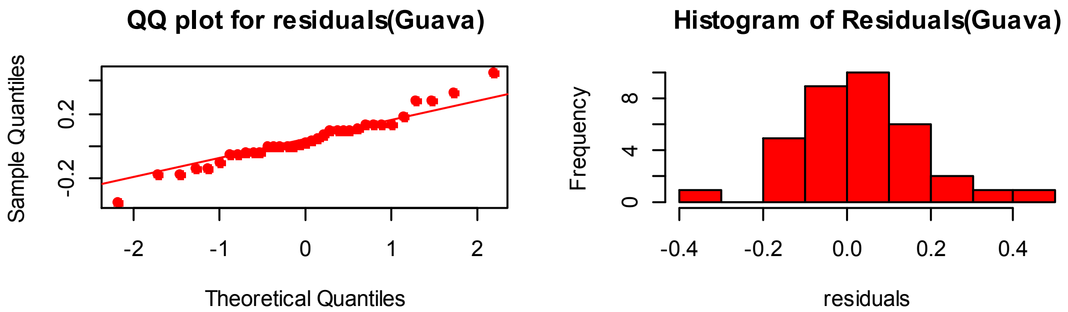

Figure-6: Graphically Residuals Diagnostic Checking for ARIMA Model of Guava Production

From the Figure-6, it is clear that almost all of the points are closed to the Q-Q line or on the Q-Q line, which implies that residuals are normally distributed. At the time, from the Histogram of the residuals of Guava production model, it is obvious that residuals are normally distributed.

Finally, considering all of the Graphical and Formal test, it is obvious that our fitted model ARIMA (1, 1,2) is the best fitted model for forecasting Guava production in the Bangladesh

7. Forecasting Fruits Productions Using the Fitted Model

After selecting the best model, now we are going to use these models to forecast different types of fruits productions in Bangladesh. To forecast the following "Forecasting Criteria" are considered which are shown in the Table 4.

Table-4: Forecasting Criteria for the Best Selected Model

Fruits Productions	Selected Model	Forecasting Criterion			
		ME	RMSE	MAPE	MAE
Mango	ARIMA(2,1,2)	-0.008311165	0.07149998	1.029789	.05462055
Banana	ARIMA(3,1,2)	3.304066	24.2942	2.73271	17.17811
Guava	ARIMA(1,1,2)	0.03865199	0.1589102	1.00591	0.1161164

8. Comparison between Original and Forecasted Series

Based on the sample data, we forecast ten years forward for Mango, Banana and Guava production in Bangladesh. It is possible to make a comparison between Original and forecasted series, this comparison are shown in Figure-7.

From the Figure-7,

- It is clear (in the Top Plot) that the original series of the Log- transformed Mango productions (red color), which show initially a downward tendency, after sometimes it shows a slightly equal productions; and finally, at the last stage it is tried to increase productions. At the same time, the forecasted series (blue color) also shows the same manner. In the forecasted plot, in-sample forecasting and out-sample forecasting part are shown same manner. That is, forecasting Mango productions may be good.
- It is transparent (in the Middle Plot) that the original series of the Banana productions (red color), which shows constant production tendency and at the end of the series, it is tried to increase productions; and the forecasting series also shows the same pattern production tendency (blue color). In the forecasting plot, in-sample forecasting part and out-sample forecasting part also shows similar trend. That is, forecasted Banana productions may be good.
- It is obvious (in the Bottom plot) that the original series of the Log-transformed Guava production (red color), which shows an upward production tendency and the forecasting series also shows an upward production (blue color) tendency. In the forecasted plot, in-sample forecasting part and out-sample forecasting part are shown the same pattern indicating this model may be good forecasting model for Guava productions in Bangladesh.

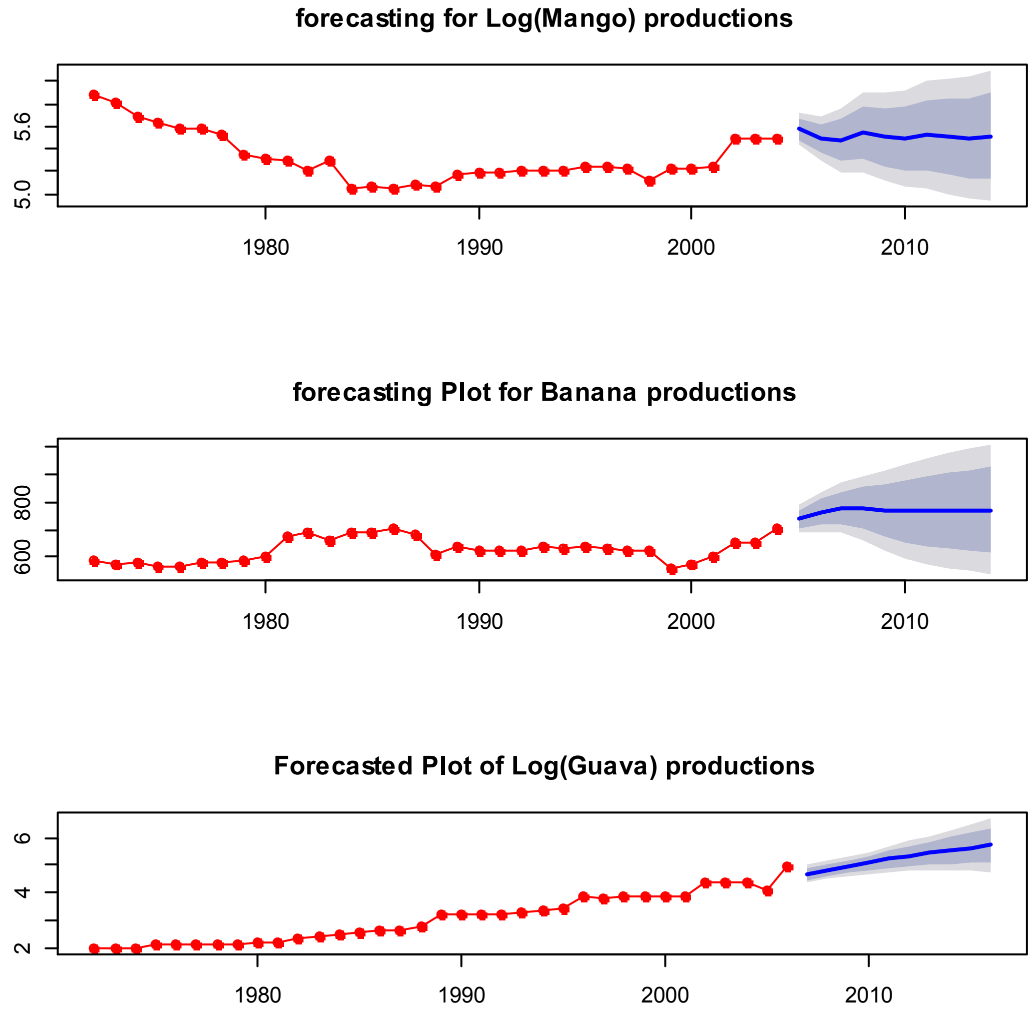

Figure-7: Graphical Comparison between Original and Forecasted Series

Finally, all of the fitted model clearly explain the practical situations very well which implies that these fitted models are statistically good fitted model for forecasting fruits productions in Bangladesh.

Conclusion and Recommendations

A time series model is used for patterns in the past movement of a variable and uses that information to forecast the future values. In this analysis, it is tried to fit the best model to forecast the different types of major fruits productions named as Mango, Banana and Guava in the Bangladesh. To select the best model for forecasting the different types of major fruits productions, the latest available model selection criteria such as AIC, BIC, AIC, BIC, etc. are used. Again, to select the fitted model, it is tried to fit a best simple model because the model contains less parameters give the good model for forecasting. To satisfy this conditions, sometimes it is considered more than 5% level of significance. The best selected Box-Jenkins ARIMA model for forecasting Mango productions is ARIMA(2,1,3) on Log-transformed data; for Banana productions, it is ARIMA (3, 1, 2) , and for Guava productions, it is ARIMA (1,1,2) on the Log-transformed data. These three model is able to explain the practical situations very well and that's why these model are best model to forecast. These model could be used to take a decisions to a researchers, policymakers, fruits producers and Businessmen covering the whole Bangladesh. At the same time, Box-Jenkins ARIMA model give the good representation of short time forecasting.

References

Jonathan D. Cryer and Kung-Sik Chan (2008), Time Series Analysis with applications in R, Springer, 2nd edition.
Gujarati D N. (2003). Basic Econometrics, 4th edn. pp 465-7. McGraw-Hill Companies Inc., New York
Yule, G. U. (1926). "Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series." *Journal of the Royal Statistical Society,* 89, 1, 1–63.

Hamjah, M. A. (2014), Rice Production Forecasting in Bangladesh: An Application of Box-Jenkins ARIMA Model, *The International Institute for Science, Technology and Education, Mathematical Theory and Modeling,* Vol-4, No-4, P: 1-11.

Wold, H. O. A. (1938). A Study of the Analysis of Stationary Time Serie.(2nd ed. 1954). Uppsala: Almqvist and Wiksells.

Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control. San Francisco, Holden- Day, California, USA.

Nasiru Suleman and Solomon Sarpong (2011), "Forecasting Milled Rice Production in Ghana Using Box-Jenkins Approach", *International Journal of Agricultural Management & Development (IJAMAD).*

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):344-53*

Sohail, A., A. Sarwar and M. Kamran (1994), "Forecasting total Food grains in Pakistan". J. Eng. Appl. Sci., Department of Mathematics and Statistics, *University of Agriculture*, Faisalabad.

Akter, M and W. M. H. Jaim (2002), "Changes in the major foodgrains production in Bangladesh and their sources during the period from 1979/80 to 1998/99". *The Bangladesh Journal of Agricultural Economics,* 25 (1):1- 16

N. M. F. Rahman (2010), "Forecasting of boro rice production in Bangladesh: An ARIMA approach", *Journal of Bangladesh Agricultural University*. 8(1): 103–112, 2010.

M. A. AWAL and M.A.B. Siddique (March, 2011), Rice Production in Bangladesh Employing By ARIMA Model, *Bangladesh J. Agril. Res*. 36(1) : 51-62