Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.4, 2014

Rice Production Forecasting in Bangladesh: An Application Of Box-Jenkins ARIMA Model

Mohammed Amir Hamjah[1]

1) MS (Thesis) in Statistics, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh.

Email: arif.stat.sust@hotmail.com.

ABSTRACT

The study was undertaken to fit the best Auto-Regressive Integrated Moving Average (ARIMA) model that could be used to forecast the rice productions of Bangladesh such as in Aus, Boro, Aman season covering the whole country. This data for the present study is available in the Bangladesh Agricultural Ministry's websites www.moa.gov.bd. The best selected ARIMA model for Aus productions is ARIMA (2,1,2), for Aman it is ARIMA (2,1,2) and, for Boro it is ARIMA (1,1,3). In this study, it was tried to make a comparison between the original series and forecasted series which also shows the same manner indicating fitted model are statistically well behaved to forecast rice productions in Bangladesh. It is found from the analysis that ARIMA model gives good forecasting for short term analysis.

KEYWORDS: Rice production, ARIMA, Forecasting, Bangladesh.

1. INTRODUCTION

The dominant food crop of Bangladesh is rice, accounting for about 75 percent of agricultural land use and it contributes 28 percent of GDP. In the mid-1980s, Bangladesh was the fourth largest rice producer in the world, but its productivity was low compared with other Asian countries, such as Malaysia and Indonesia. It is currently the world's sixth-largest rice producer. High yield varieties of seed, application of fertilizer, and irrigation have increased yields, although these inputs also raise the cost of production and chiefly benefit the richer cultivators. The cultivation of rice in Bangladesh varies according to seasonal changes in the water supply. The largest harvest is Aman, occurring in November and December and accounting for more than half of annual production. Some rice for the Aman harvest is sown in the spring through the broadcast method, matures during the summer rains, and is harvested in the fall. The higher yielding method involves starting the seeds in special beds and transplanting during the summer monsoon. The second harvest is Aus, involving traditional strains but more often including high-yielding, dwarf varieties. Rice for the Aus harvest is sown in March or April, benefits from April and May rains, matures during in the summer rain, and is harvested during the summer. With the increasing use of irrigation, there has been a growing focus on another rice-growing season extending during the dry season from October to March. The production of this Boro rice, including high-yield varieties, expanded rapidly until the mid-1980s, when production leveled off at just below 4 million tons (Wikipedia).

2. REVIEW OF LITERATURE FOR ARIMA MODEL

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 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. 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 production 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.

3. OBJECTIVE OF THE STUDY

The main objective of this study is to develop an ARIMA model for forecasting the rice productions in the Bangladesh. The specific objective of the study is to develop an Autoregressive Integrated Moving Average (ARIMA) model for different seasonal rice productions such as Aus, Aman and Boro season in Bangladesh.

4. 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 (exception: monthly, quarterly, yearly, etc.) with a consistency in the activity and the method of measurement.

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 (ARMIA) model with external regressor, that is, ARIMAX model. This model is the generalized model of the non-stationary ARMA model denoted by ARMA(p,q) can be written as

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

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

And it consists of the combination of Auto-Regressive series, AR(p) and Moving Average series, MA(q) , where AR(p) can be defined as $Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + e_t$; and MA(q) can be defined as $Y_t = e_t - \theta_1 e_{t-1} \dots - \theta_q e_{t-q}$

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

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

Box and Jenkins procedure's steps

1. **Preliminary analysis:** create conditions such that the data at hand can be considered as the realization of a stationary stochastic process.
2. **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.
3. **Estimate:** efficient, consistent, sufficient estimate of the parameters of the ARIMA model (maximum likelihood estimator).
4. **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.
5. **Usage of the model:** if the model passes the diagnostics step, then it can be used to interpret a phenomenon, forecast.

6.1. Diagnostic Tests of Residuals

6.1.1. 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.

6.1.2. 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 $H_0 : \rho_1(e) = \rho_2(e) = \dots\dots = \rho_k(e) = 0$

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).

7. DATA SOURCE AND USED SOFTWARE

The crop data sets are available from Bangladesh Agricultural Ministry's websites named as www.moa.gov.bd. These data set are available from the year1972 to 2006. 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**.

8. RESULTS AND DISCUSSION OF ARIMA MODEL

8.1. ARIMA Modeling for Aus Production

Dickey-Fuller unit root test is used to check whether time sequence Aus production data series is stationary or not. It is found that stationarity condition satisfied at the difference order one with the Pr(|t| ≥ - 46.1394) < 0.01, which suggests that there is no unit root at the first order difference at 5% level of significance. The graphical stationarity test using ACF and PACF is shown in the Figure-1.

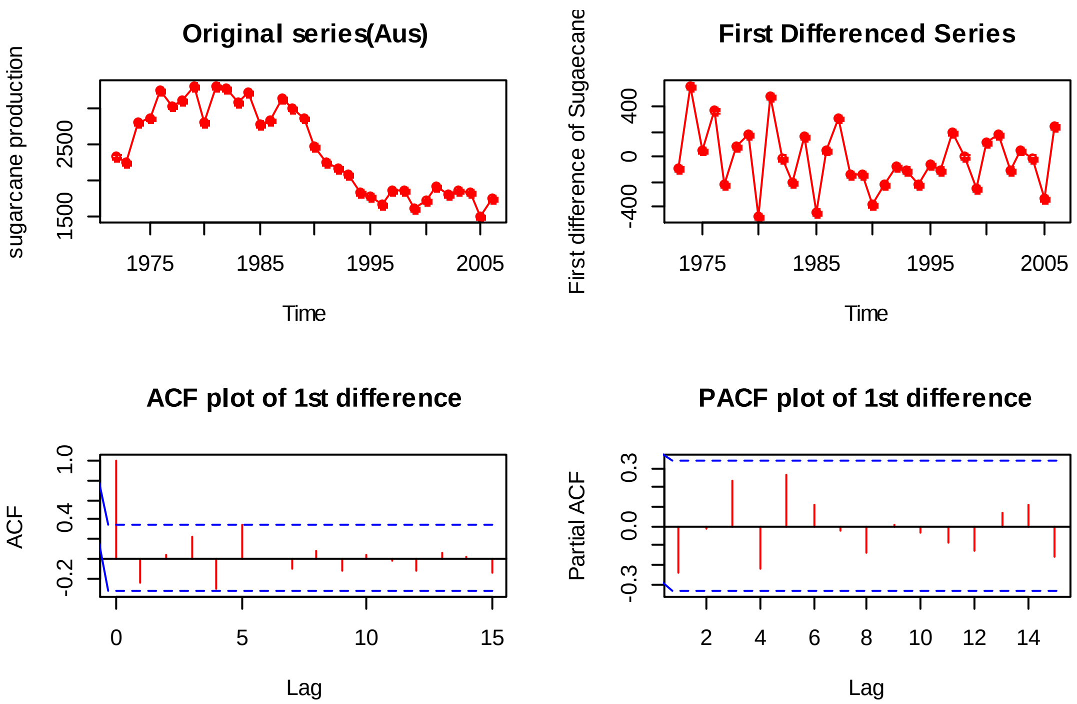

Figure-1: Graphically Stationarity Checking for Aus Production

From the Figure-1, it is clear that at the first difference of Aus production series shows more stable variance than the original series, whereas in the original series is not stationary. That is, our difference order is one to make the Aus production series as stationary. Again, from the ACF and PACF, it is clear that there is no significant spike in the first order differenced series which also indicate that there is no significant effects of Auto-Regressive and Moving Average in the first order difference, that is, the Aus production series is stationary at the first order difference.

From the tentative order analysis, the best selected ARIMA model for Aus production forecasting is ARIMA(2,1,2) with AIC = 474.42 and BIC = 482.05. The parameter estimates of the fitted ARIMA(2,1,2) is given in the Table 1.

From the summary statistics of the Table 1, which shows that first and second order Auto-regressive Lag and first and second order Moving Average Lag have statistically significant effects on Aus production at 10% level of significance.

Table 1: Summary Statistics of ARIMA (2,1,2) for Aus Production

Coefficients	Estimates	Std. Error	t-value	p-value
ar1	1.1393	0.1803	6.3197	0.05
ar2	-0.5489	0.1723	-3.1856	0.0968
ma1	-1.5025	0.1281	-11.7276	0.0271
ma2	1	0.1342	7.4494	0.0425

To check the Autocorrelation assumption, “Box-Ljung” test is used. From the test, it is found that the $\Pr(|\chi_1^2| \geq 0.0724)= 0.7879$, which strongly suggests that we may accepted that there is no autocrrelation among the residuals of the fitted ARIMA(2,1,2) models at 5% level of significance. Again, to check the normality assumptions, “Jarque-Bera” test is used. From the test, it is found that the $\Pr(|\chi_2^2| \geq 0.0097) = 0.9951$, which strongly suggests to accept the norality assumpytion that the residuals of the fitted ARIMA(2,1,2) model for Aus productions follow normal distributins.

Finally, considering all graphical and formal test, it is clear that our fitted ARIMA (2, 1,2) model is the best selected model for forecasting Aus productions in the Bangladesh.

8.2. ARIMA Modeling for Aman Production

Dickey-Fuller unit root test is used to check whether time sequence Aman production series is stationary or not. It is found that stationarity condition satisfied at the difference order one with the $\Pr(|t| > -36.4425) < 0.01$, which suggests that there is no unit root in the first order difference at 5% level of significance. The graphical stationarity test using ACF and PACF is shown in the *Figure 2.*

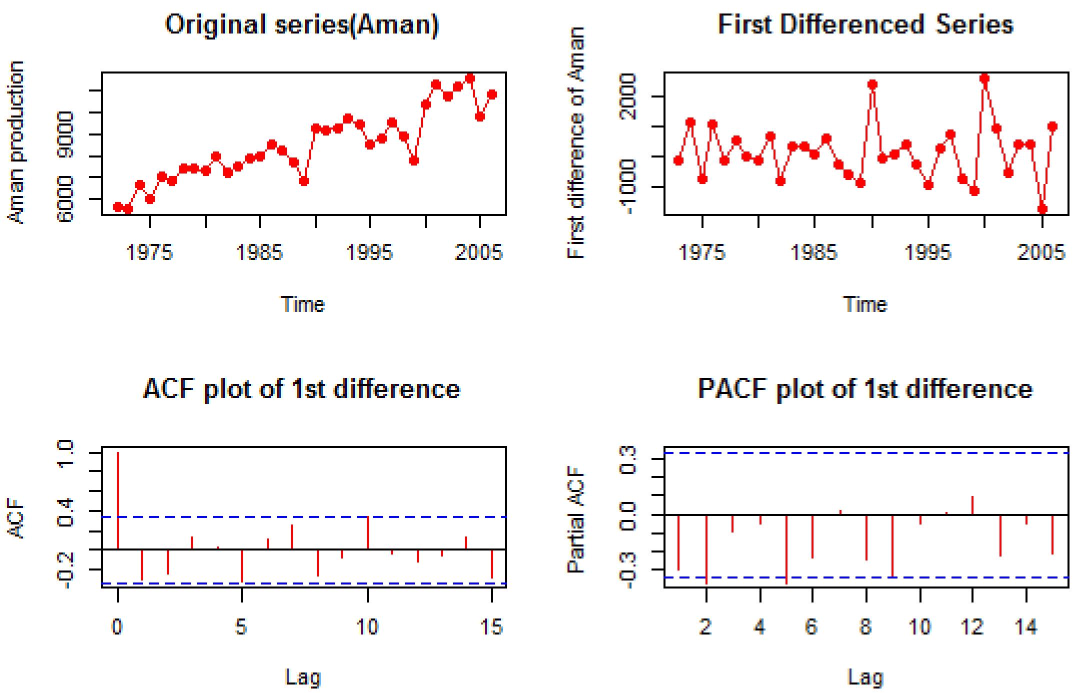

Figure 2: Graphically Stationarity Checking for Aman Production

From the *Figure 2*, it is clear that at the first difference of Aman production series shows constant variance, but the original series shows an upward trend, whereas in the original series is not stationary. That is, our difference order is one to make the Aman production series as stationary by removing the trend effects. Again, from the ACF and PACF, it is clear that there is no significant spike in the first order differenced series which also indicate that there is no significant effects of Auto-Regressive and Moving Average in the first order difference, that is, the Aman production series is stationary at the first order difference.

Finally, the best selected ARIMA model for Aman production forecasting in the tentative order analysis is ARIMA (2, 1, 2) with AIC = 558.86 and BIC = 566.49. The parameter estimates of the fitted model for Aman production is shown in the Table 2

Table 2: Summary statistics of ARIMA model for Aman Production

Coefficients	**Estimates**	**Std. Error**	**t-value**	**p-value**
ar1	-0.6189	0.1062	-5.8278	0.0541
ar2	-0.9603	0.0735	-13.0589	0.0243
ma1	0.5104	0.2458	2.0766	0.1429
ma2	0.7924	0.167	4.7454	0.0661

From the Table 2, it is clear that two year Auto-regressive Lag and two years Moving Average Lag have significant effects on Aman production. Here considerable thing is that the first orders Moving Average term have also significance effects on 14% level of significanc. Here we consider 14% level of significance because only these order of the fitted ARIMA model gives the best results and fulfill the forecasting.

To check the Autocorrelation assumption, "Box-Ljung" test is used. From the test, we get the Pr($|\chi_1^2| \geq 1.3537$) = 0.2446, which suggests to accepth the assuptions that there is no autocrrelation among the residuals of the fitted ARIMA(2,1,2) models at 5% level of significance. Again, to check normality assumptions, "Jarque-Bera" test is used. From the test, we get the Pr($|\chi_2^2| \geq 3.7184$) = 0.1558 which suggests to accept the norality assumpytions such that the residuals of the fitted ARIMA(2,1,2) model for Aman productions follows normal distributins at 5% level of significance.

Finally, considering all graphical and formal test, it is clear that our fitted model ARIMA (2, 1, 2) for Aman production is the best selected model which serve us forecasting criteria.

8.3. ARIMA Modeling for Boro Production

To check the stationarity, "Dickey-Fuller" unit root test is ued on the Log-transformed data set of Boro production because of avoiding the non-normality assumptions for the fitted model. From the "Dickey-Fuller" unit root test, it is found that Pr(|t| > -3.9135) = 0.02457, which suggests that for the first difference the Log-transformed Boro production series satisfies the stationarity condition at 5%level of significance. The graphical stationarity test using ACF and PACF is shown in the *Figure3*

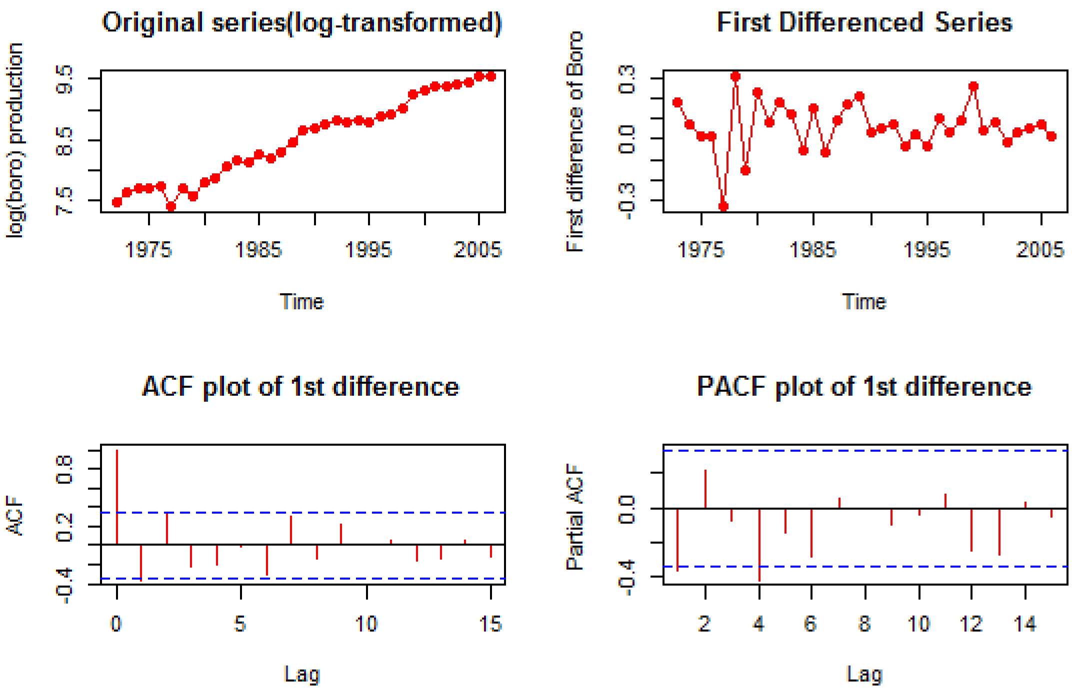

Figure 3: Graphically Stationarity Checking for Boro Production

From the *Figure 3*, it is clear that at the first difference of Aman production series shows constant variance, but the original series shows an upward trend, and it is not stationary. That is, our difference order is one to make the Boro production series as stationary by removing the trend effects. Again, from the ACF, it is clear that there is no significant spike in the first order differenced series which also indicate that there is no significant effects of

Auto-Regressive and Moving Average in the first order difference, that is, the Boro production series is stationary at the first order difference. Again, there is only one significance spike in the PACF at Lag order four, but other test imply that the series is stationary, so we can take difference order one to make the Boro production series stationary, otherwise, we face over differencing problem.

Therefore, from the tentative order analysis, the best selected ARIMA model for Boro production forecasting is ARIMA (1, 1, 3) with AIC= -43.76 and BIC= -36.13. The parameter estimates of the fitted ARIMA model for Boro production is shown in the Table 3.

Table 3: Summary Statistics of ARIMA Model for Amna Production

Coefficients	**Estimates**	**Std. Error**	**t-values**	**p-values**
ar1	0.9997	0.0023	427.4574	0.0007
ma1	-1.1527	0.1442	-7.9956	0.0396
ma2	0.6773	0.2892	2.3419	0.1285
ma3	-0.5117	0.2315	-2.21	0.1353

From the summary statistics of Table 3, it is obvious that shows that one year Auto-regressive Lag and three year Moving Average Lag have significant effects on Boro production at 13% level of significance.
To check the Autocorrelation assumptions, “Box-Ljung” test is used. From the test, we get the $\Pr(|\chi_1^2| \geq 0.2014) = 0.6536$, which suggests that we may accept the assuptions that there is no autocrrelation among the residuals of the fitted ARIMA(1,1,3) model at 5% level of significance. Again, to check the normality assumption,“Jarque-Bera” is used. It gives the $\Pr(|\chi_2^2| \geq 4.7834) = 0.0915$ which suggests to accept the norality assumpytions such that the residuals of the fitted ARIMA(1,1,3) model for Boro productions follow normal distributin at 5% level of significance.

Finally, considering all graphical and formal test, it is clear that our fitted ARIMA (1, 1, 3) model is the best selected model for forecasting Boro production in the Bangladesh.

9. FORECASTING RICE PRODUCTIONS USING THE FITTED MODEL

After selecting the best model, now we are going to use this model to forecast rice productions. To forecast the following “Forecasting Criteria” are considered which are shown in the *Table 4.*

Rice productions	**Selected Model**	**Forecasting Criterion**			
		ME	**RMSE**	**MAPE**	**MAE**
Aus	ARIMA(2,1,2)	-15.50667	209.4102	6.77762	162.9278
Aman	ARIMA(2,1,2)	168.6118	747.4399	6.642766	559.192
Boro	ARIMA(1,1,3)	0.0117	0.1047	0.9429	0.7713

Table-4: Forecasting Criteria for the Best Selected Model

10. COMPARISON BETWEEN ORIGINAL AND FORECASTING SERIES

Based on the sample data, we forecast eight year forward for Aus, Aman and Boro production in Bangladesh. It is possible to make a comparison between Original and forecasted series, this comparison are shown in Figure-4.

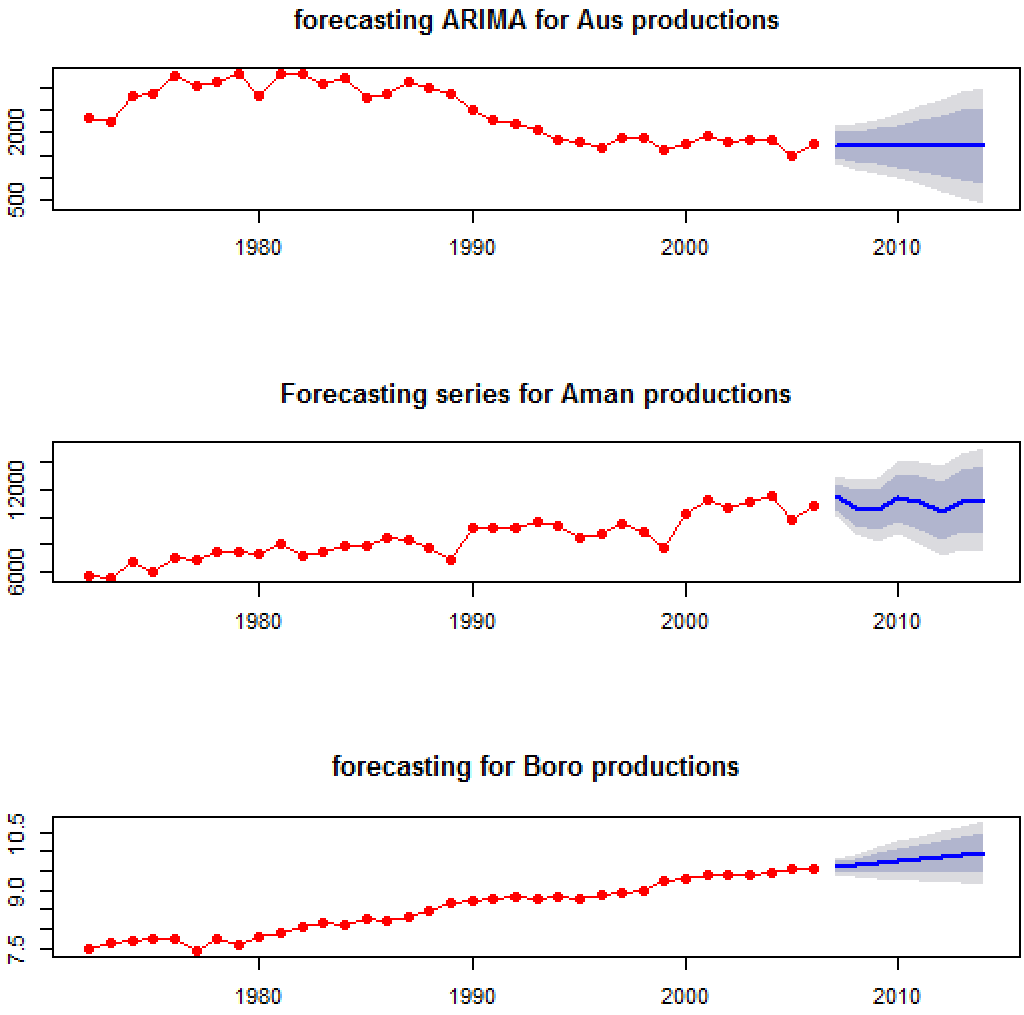

Figure 4: Graphical Comparison between Original and Forecasted Series

From the *Figure 4*,

- It is clear (in the top plot) that the original series of the Aus production (red color), which show initially a downward tendency, after sometimes it shows a slightly equal productions and the forecasting series (blue color) also shows the same manner. In the forecasting plot, in sample forecasting part shows a downward trend and similarly the out sample forecasting part also shows a downward trend. That is, forecasting Aus productions may be good.
- It is clear (in the Middle plot) that the original series of the Aman productions (red color), which shows an upward production tendency and the forecasting series also shows an upward production tendency (blue color). 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 Aman productions may be good.
- It is clear (in the Bottom plot) that the original series of the Log-transformed Boro production (red color), which shows an upward production tendency and the forecasting series also shows an upward

production (blue color) tendency. 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 Boro productions may be good.

Finally, all of the fitted model clearly explain the real situations which implies that these fitted models are statistically good fitted model for rice productions covering the Bangladesh area.

11. 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 seasonal rice productions named as Aus, Aman and Boro in the Bangladesh. To select the best model for forecasting different seasonal productions the latest available model selection criteria such as 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 was considered more than 5% level of significane. The best selected Box-Jenkins ARIMA model for forecasting Aus productions is ARIMA (2, 1, 2), for Aman it is ARIMA (2, 1, 2) , and for Boro it is ARIMA (1,1,3) on the Log-transformed data. These three model is able to explain the practical situations and that's why these model are best model. These model could be used to take a decisions to a researchers, policymakers, rice producer and Businessmen covering the whole Bangladesh.

12. ACKNOWLEDGEMENT

I want to thanks Professor Md. Ahmed Kabir Chowdhury, Department of Statistics, Shah-Jalal University of Science and Technology, Sylhet-3114, Bangladesh to help and advise me to make this article.

13. 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.
- 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 INBANGLADESH EMPLOYING
- BY ARIMA MODEL, *Bangladesh J. Agril. Res*. 36(1) : 51-62