International Journal of Sciences & Applied Research

International Journal of Sciences & Applied Research

www.ijsar.in

An application of Box-Jenkins methodology for forecasting rice production in Bangladesh

Mashfiqul Huq Chowdhury, Juthi Saha and Gowranga Kumar Paul*

Department of Statistics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902.

Corresponding author: *Gowranga Kumar Paul, Department of Statistics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902.

Abstract

Rice is the main agricultural crop of Bangladesh, accounting for about 75 percent of agricultural land use (and 28 percent of GDP). For country planning, forecasting is the main tool for predicting the production to determine the situation what would be the value of production coming year. In this research, our main objective is to find a suitable model for forecasting rice production of Bangladesh. A time series modeling approach, Box-Jenkins methodology was used for this purpose. This study utilizes yearly data of rice production of Bangladesh during 1972-2015 from Bangladesh Rice Research Institute (BRRI). Initially data was non-stationary but it was transformed to stationary after taking suitable differences. Four basic chronological steps of Box-Jenkins methodology namely: identification, estimation, diagnostic checking and forecasting were fitted out in developing the ARIMA model. Validity of the model was tested using standard graphical explanation of residuals given by Box and Jenkins. As a second step of validation, forecasted values of rice production were checked using actual data series. Various time series models are fitted on this data. We select the ARIMA model, which exhibited the least AIC values. After completion of diagnostic checking, ARIMA (5,2,3) model has been identified as an appropriate model for forecasting rice production of Bangladesh for the next fifteen years (2016-2030). We hope that the findings of this study would be more useful for policy makers, researchers as well as producers in order to forecast future national production of rice more accurately in the short run.

Keywords: ARIMA, Box-Jenkins, GDP, Rice

Introduction

Rice is interwoven with Bengali culture. It is the symbol of wealth (Murshed, 2012). Rice is the staple food of about 160 million people of Bangladesh. It provides nearly 48% of rural employment, about two-third of total calorie supply and about one-half of the total protein intakes of an average person in the country. Rice sector contributes one-half of the agricultural GDP and one-sixth of the national income in Bangladesh. Almost all of the 13 million farm families of the country grow rice. Rice is grown on about 10.5 million hectares which has remained almost stable over the past three decades (ICCB, 2017).

Rice production in Bangladesh is a crucial part of the national economy (Bangladesh, 2009). Rice is cultivated in Bangladesh throughout the year as Aush, Aman or Boro. Aman (broadcast and transplanted) is generally cultivated in December-January, Boro in March-May and Aus in July-August cropping seasons. Among these transplanted cropping Aman is most important and occupied about 46% of the rice cultivated land in 2009-10. The rest 41, 9 and 4 percent of the land is occupied by Boro, Aus and Sown Aman respectively (BRRI, 2017).

Rice is considered as the one of the most important cereal in human consumption. World population particularly the population of developing world is increasing at an alarming rate. To feed this ever increasing human population always remains a challenging task to the planners of the individual countries and also the world bodies. Thus, forecasting production behaviors of the major crops like rice play vital role towards the planners for food and nutritional security.

Bangladesh has significant strides in its economic sector performance since independence in 1971. Bangladesh is among the top 12 developing countries with a population of over 20 million, who achieved 6 plus percent growth in 2016. By any standards, Bangladeshi economy has done well. The economy of Bangladesh is the 44^{th} largest in the world in nominal terms, and 32^{nd} largest by purchasing power parity; it is classified among the next eleven emerging market economies and a Frontier market (BDU, 2016).

According to the estimates of International Rice Research institute (2009), China tops the list of top 10 rice producing nations with a production of 193 million metric tons. The latest data puts India in the second spot as the country produces 148 million metric tons. Bangladesh is the fourth-largest rice producer. In spite of the decline in the country’s arable land since its independence in 1971, the rice area harvested increased from almost 10 million hector in 1995 to nearly 12 million hector in 2010. Rice yield also improved in the last decade, from a low of 2.7 ton/ha in 1995 to almost 4.3 ton/ha in 2010. These increases in rice yield and area harvested contributed to growth in rice production, which nearly doubled from over 26 million ton in 1995 to 50 million ton in 2010 (RDRS,2012),.

Background of the study

To maintain an acceptable level of food security, Bangladesh needs to continuously increase domestic rice production. Bangladesh is adding about 2 million people to its population every year and will continue to do so for the foreseeable future. Thus, Bangladesh will require about 40 million tons of rice for the year 2020. During this time total rice area will also shrink to 10.28 million hectares. Rice yield therefore, needs to be increased from the present 2.74 to 4.50 ton/ha. Therefore, rice production needs to increase by a minimum of 40 million tons/year (BRRI , 2017).

A few studies on crop production have been conducted so far. Still many researchers are trying to identify different characteristics of such a data for various time periods by utilizing newer techniques available in Statistics. Amin et. al. (2014) has worked on the time series modeling for forecasting wheat production of Pakistan. They propose time series model to forecast the wheat production of Pakistan.

Yaziz et. al. (2011) has worked on a comparative study on Box-Jenkins and Garch models in forecasting crude oil price. Maity and Chatterjee (2012) have worked on the forecasting GDP growth rate of India. This study has attempted to shed light on the issues such as forecasting growth rates of GDP of India. Data on GDP have been collected over a period of 60 years from various publications of Reserve Bank of India. A very simple tentative ARIMA (1, 2, 2) model has been fitted on data to estimate

the parameters of autoregressive and moving average components of this model.
The objectives of this study are to forecast the rice production in Bangladesh during 2016-2030 by applying ARIMA model.

Data and methodology

This study used secondary data from Bangladesh Rice Research Institute (BRRI) and Bangladesh Bureau of Statistics (BBS) for yearly production of rice. To forecast the rice production the study used the amount of rice produced during 1972 to 2015. To identify the perfect model for forecasting the rice production we need to collect the previous information very carefully. The study used Autoregressive integrated moving average (ARIMA) model because this model forecasts future values of a time as a linear combination of its own past value and series of errors. A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where,

- p is the of autoregressive terms,
- d is the number of non-seasonal differences needed to stationarity, and
- q is the of lagged forecast error in the prediction equation.

In terms of y, the general forecasting equation is:

$$y_t = \mu + \phi_1 y_{t-1} + \dots \dots \dots + \phi_p y_{t-p} - \dots \dots \dots - \theta_1 e_{t-1} - \dots \dots \dots - \theta_q e_{t-q} \qquad \dots \dots \dots (i)$$

Here the moving average parameters (θ's) are defined so that their signs are negative in the equation, following the convention introduced by Box and Jenkins. To identify the appropriate ARIMA model for Y, we begin by determining the order of differencing (d) needing to stationarize the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating.

Steps to Conduct Box-Jenkins Modeling

To identity a perfect ARIMA model for a particular data series, Box and Jenkins developed a method known as Box-Jenkins methodology or in short, B-J methodology. The Box-Jenkins method requires that the discrete time series data be equally spaced over time and that there be no missing values in the series. Box-Jenkins forecasting models are based on statistical concepts and principles and are able to model a wide spectrum of time series behavior. The series also needs to be at least weakly stationary.

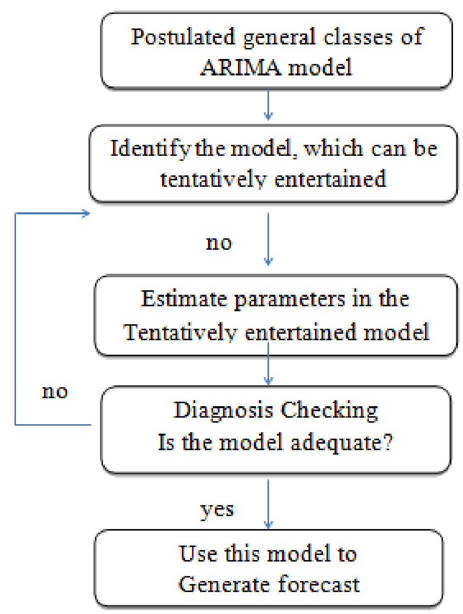


Figure 1: The Box-Jenkins Methodology for Optimal Model Selection.

Results and discussion

Checking Stationarity:

To check stationarity at first we plot the amount rice production against time. Figure 2 shows that a non-stationary is observed so it is not possible to fit ARIMA model because ARIMA model is defined for stationary data.

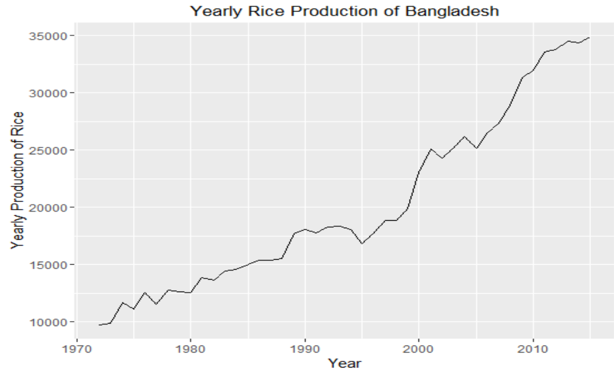
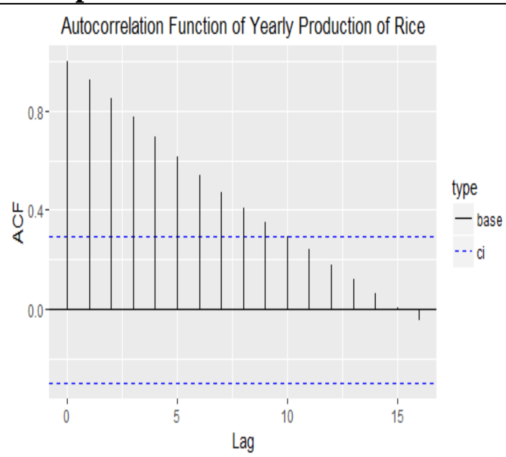
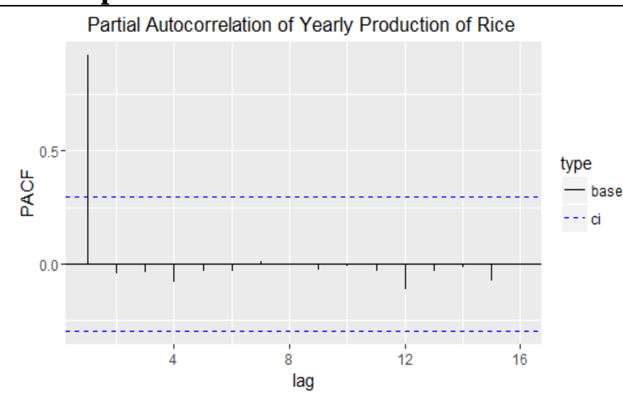
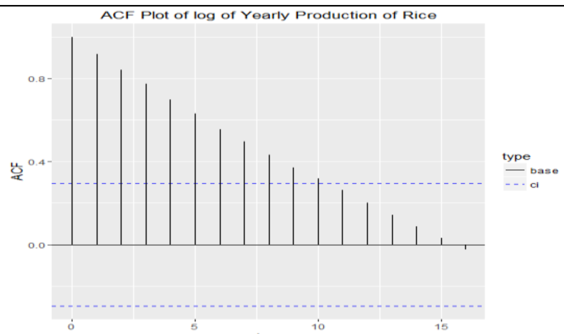
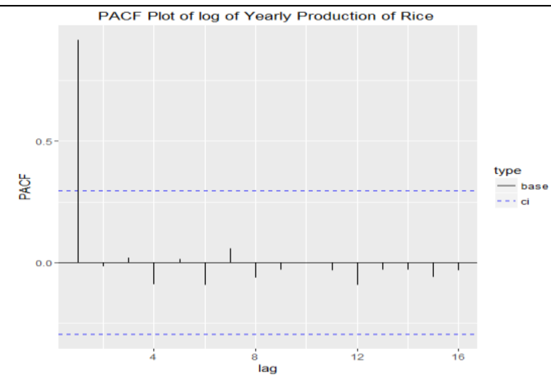


Figure 2: Rice Production during 1972-2015.

ACF and PACF Plots

By taking main data set, 1st difference, log transformed and 2nd differences, it is convenient to examine stationarity by using autocorrelation function (ACF) and partial auto correlation function (PACF) graphs (called correlogram) or by using different tests like Dickey-Fuller or Ljung-Box test. Figure 3 - Figure 5 illustrates ACF and PACF plots (no difference, 1st difference and log transformed) show decreasing and for several lags,the spikes fall outside the confidence limit. Thus the reasult shown by the correlogram is that the data are non stationary.

Index	ACF plots	PACF plots
Figure 3: (No Difference)		
Figure 4: (Log Transformed)		

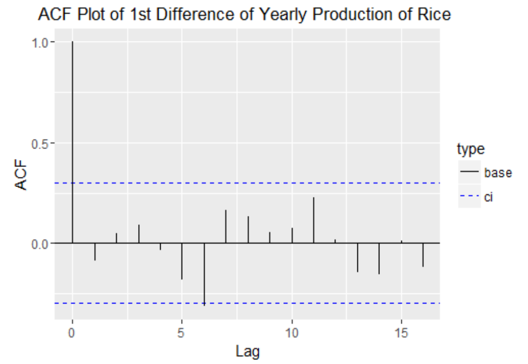
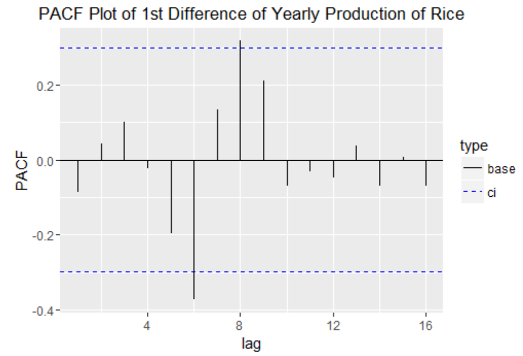
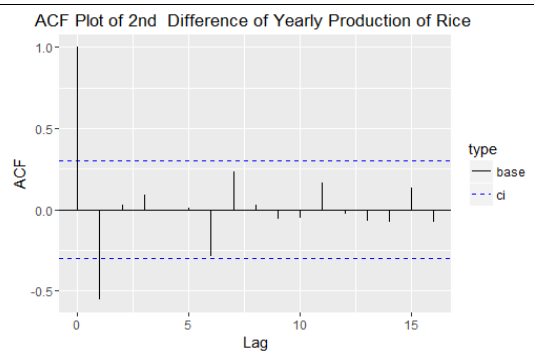
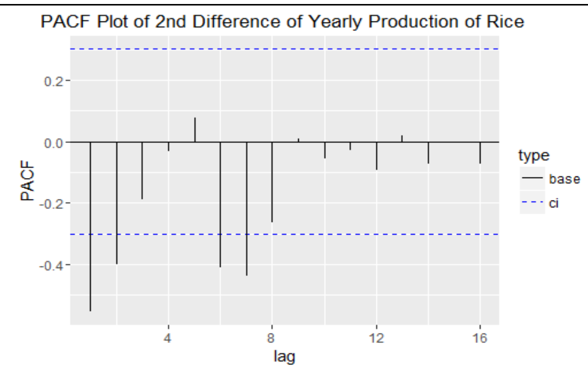
Index	ACF plots	PACF plots
Figure 5: (First Difference)		
Figure 6: (Second Difference)		

Table 1: Augmented Dickey-Fuller test results about stationarity.

Variable	Lag Order	Difference	ADF Statistic	$p-value$
Rice Production	3	0	-1.627	0.7214
Log Rice Production	3	0	-2.762	0.2718
Rice Production	3	1	-3.219	0.0969
Rice Production	3	2	-4.085	0.0158
Rice Production	3	3	-6.997	0.0100

For lag order 3, the calculated Dickey-Fuller test statistic for second difference is -4.085 and the p -value is 0.01. For this p -value, the null hypothesis can be rejected at 5 percent level of significance. Thus, the test concludes that the series is stationary for second difference. Since the data are concluded as stationary from the test, so we choose our difference parameter $d = 2$.

According to Figure 6, the ACF and PACF plots (2nd difference) show that there is no lags and the spikes fall outside the confidence limit. Thus the reasult shown by the correlogram is that the data are stationary when we take 2nd difference of time series data set.

Augmented Dickey-Fuller Test about Stationarity (Taking Differences)

A statistical test to check stationarity of a time series data set is ADF test. The following table shows the calculated Dickey-Fuller test statistic, Lag order and $p-value$ for different differences.

Model Selection

To determine an appropriate ARIMA model it is necessary to choose other parameters such as order of the autoregressive parameter (p) and order of moving average parameter (q) of ARIMA model. Among the different plausible models, the best model can be selected by using the Akaike Information Criterion (AIC) and Bayesian information criterion (BIC). Following table shows the AIC values for different combinations of p and q, that is, for different ARIMA $(p,2,q)$ models.

Table 2: Values for Different ARIMA $(p,2,q)$ Model.

Model	*AIC*	AIC_C	*BIC*
ARIMA(0,2,0)	730.1682	730.2682	731.9059
ARIMA(0,2,1)	703.3683	703.6760	706.8437
ARIMA(0,2,2)	705.1168	705.7483	710.3298
ARIMA(0,2,3)	706.9018	707.9829	713.8525
ARIMA(0,2,4)	703.5712	705.2379	712.2596
ARIMA(0,2,5)	703.7866	706.1866	714.2126
ARIMA(1,2,0)	716.3019	716.6096	719.7772
ARIMA(1,2,1)	705.0886	705.7202	710.3016
ARIMA(1,2,2)	707.0685	708.1496	714.0192
ARIMA(1,2,3)	708.2810	709.9477	716.9693
ARIMA(1,2,4)	703.8329	706.2329	714.2589
ARIMA(1,2,5)	705.4383	708.7324	717.6020
ARIMA(2,2,0)	711.4522	712.0838	716.6652
ARIMA(2,2,1)	707.0086	708.0897	713.9593
ARIMA(2,2,2)	708.8964	710.5630	717.5847
ARIMA(2,2,3)	704.5468	706.9468	714.9728
ARIMA(2,2,4)	705.7516	709.0457	717.9153
ARIMA(2,2,5)	704.8704	709.2341	718.7718
ARIMA(3,2,0)	711.9960	713.0771	718.9466
ARIMA(3,2,1)	708.4524	710.1190	717.1407
ARIMA(3,2,2)	710.4521	712.8521	720.8782
ARIMA(3,2,3)	705.5529	708.8470	717.7166
ARIMA(3,2,4)	705.0776	709.4412	718.9789
ARIMA(3,2,5)	707.0129	712.6379	722.6520
ARIMA(4,2,0)	713.9928	715.6594	722.6811
ARIMA(4,2,1)	710.4515	712.8515	720.8775
ARIMA(4,2,2)	712.4432	715.7373	724.6069
ARIMA(4,2,3)	705.9972	710.3609	719.8986
ARIMA(4,2,4)	706.9486	712.5736	722.5876
ARIMA(4,2,5)	703.4239	710.5206	720.8006
ARIMA(5,2,0)	715.6416	718.0416	726.0676
ARIMA(5,2,1)	710.8098	714.1039	722.9735
ARIMA(5,2,2)	707.1905	711.5542	721.0919
ARIMA(5,2,3)	**702.6755**	**708.3005**	**718.3145**
ARIMA(5,2,4)	707.4669	714.5636	724.8436
ARIMA(5,2,5)	703.9314	712.7314	723.0458

From the above table, we have found that the AIC value for the model ARIMA (5,2,3) is minimum. Hence, our required appropriate order of ARIMA model is ARIMA (5,2,3).

Estimation and diagnostic of the residuals

The objective of this study is to develop a general model that will help to predict the rice production in near future. Finally, we found our best model which is ARIMA (5,2,3). Here, the model includes five (5) autoregressive, three (3) moving average coefficients and the degree of difference is 2. Thus, the model can be written as

$$\hat{Y}_t = (2+\phi_1)Y_{t-1} + (\phi_2 - 1)Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3}$$

Where,

$Y_t =$ Rice production of Bangladesh

$\theta =$ Moving average coefficients

$e_t =$ Error term

$\phi =$ Autoregressive coefficients

Now, the parameters have to be estimated using the maximum likelihood method. Following table gives the maximum likelihood estimates of the parameters and significance of the parameters.

Diagnostic Checking:

To check the randomness of the residual, we use Run test and after conducting the run test the $p-$value founded is 0.5317. Hence the residuals are random. Ljung-Box statistic was used for checking the independence of the residuals. For 10 degrees of freedom, the test shows the calculated Chi-squared statistic is 5.0015 and the $p-$value is 0.8911. Hence, it can be said that the residuals are independent of each other. Following plot (Figure: 7) shows the standardized residuals, ACF of residuals and $p-$values for Ljung-Box statistic taking different lag values that is the diagnostic plot for the model.

Table 3: Estimates and Significance Levels of the Parameters of ARIMA (5,2,3) Model.

Parameters	Estimates	Standard errors	p-Value
ϕ_1	0.8014	0.1885	0.000
ϕ_2	-0.4340	0.1971	0.006
ϕ_3	-0.0787	0.2048	0.006
ϕ_4	-0.1771	0.1973	0.006
ϕ_5	-0.2382	0.1818	0.000
θ_1	-2.1362	0.2822	0.000
θ_2	2.0619	0.5854	0.004
θ_3	-0.7898	0.3183	0.000

From the table it is seen that, $p-$ value of all coefficients are less than 0.05. Therefore, it can be said that all the parameters are highly significant. To ensure that the chosen ARIMA model fits the data well, we have to check the residuals estimated from this model whether or not they are white noise.

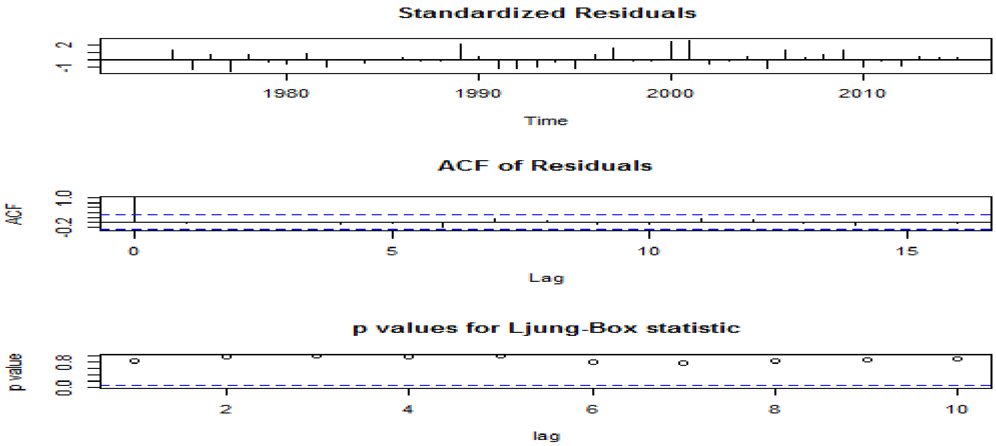


Figure 7: Diagnostic Plot for ARIMA Model

The figure illustrates that the independence assumption of residuals are also satisfied since none of the ACF values of lag 1 or more is significant. The plot also shows the large $p-$values for Ljung–Box statistics for different lag values which is also the evidence of independence of residuals. In order to check the residuals are normally distributed or not, Shapiro-Wilk test is conducted and after conducting the test, the $p-$value founded is 0.1367. So that residuals are normally distributed. Thus, all the diagnostic checks support that the selected model has not only the smallest AIC value but also satisfy all of the assumptions of a regression model.

Actual and Fitted Plot

Before forecasting, it is convenient to plot the actual and fitted values for the sample points using the selected model to check how the model fit for the actual values. A plot of the actual values and the fitted values using the model is given in figure 8. In the figure 8 the red line denotes the actual values and the blue line denotes the fitted values. From the plot, it is seen that the model has a much close fit. We can conclude that the actual and fitted values are very close to each other.

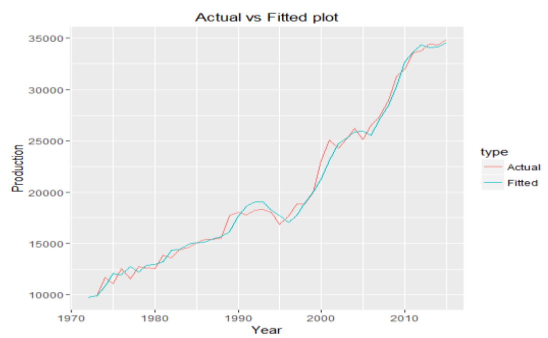


Figure 8: Actual vs. Fitted Values for ARIMA (5,2,3) Model

Forecasting

The point forecast with 95% confidence interval of yearly rice production of Bangladesh for the years 2016 to 2030 by using the selected model is given in Table 4.

Table 4: Forecasts of per capita GDP for next 16 years (2016-2030).

Time Period	Points Forecasts	95% Confidence Interval
2016	35706.88	(34193.60, 37220.16)
2017	37265.63	(35424.50, 39106.76)
2018	39013.29	(36720.40, 41306.17)
2019	40664.34	(37683.89, 43644.80)
2020	41894.44	(38209.91, 45578.97)
2021	42606.84	(38394.89, 46818.80)
2022	42891.40	(38211.68, 47571.12)
2023	43063.07	(37932.53, 48193.60)
2024	43468.31	(37889.42, 49047.20)
2025	44335.38	(38264.06, 50406.70)
2026	45679.15	(39015.50, 52342.80)
2027	47307.99	(39948.87, 54667.11)
2028	48907.50	(40789.79, 57025.22)
2029	50184.82	(41295.47, 59074.17)
2030	50999.77	(41367.83, 60631.70)

On the basis of our selected model, our analysis reveals that rice production of Bangladesh would become 41894.44 thousand tons in 2020 and 50999.77 thousand tons (50.9 million metric tons) in 2030 under the assumption that there is no irregular movement of variation is occurred.

Actual and Fitted Forecasting Plot

The point forecasts for rice production of Bangladesh with the 95% confidence interval for fifteen years along with the data has been shown in above table and The forecast values with the 95% confidence intervals are shown in figure 9, where the two lines indicate the forecast values for next fifteen years and the red and green lines indicate the 95% confidence intervals for those forecasts. It is found that the upward trend in rice production will still continue for this decade.

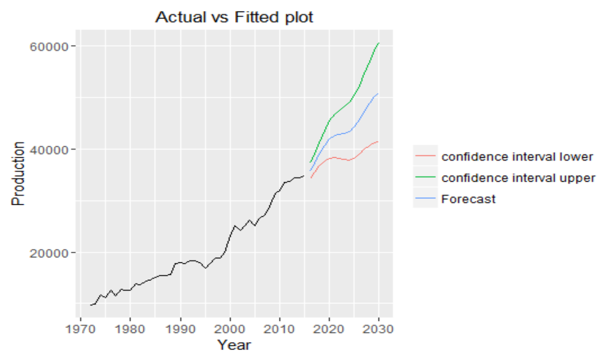


Figure 9: Forecasts with 95% Confidence Intervals.

Conclusions

The dominant food crop of Bangladesh is rice, accounting for about 75 percent of agricultural land use (and 28 percent of GDP). Rice production increased every year. Total rice production in Bangladesh was about 9.74 million tons in the year 1972 when the country's population was only about 70.88 millions. The country is now producing about 44 million tons to feed her around 160 million people. This indicates that the growth of rice production was much faster than the growth of population. The continuous growth in agriculture and rice had been mainly possible because of scientific research and innovation, development of new rice varieties (high-yielding, short duration, stress-tolerant rice) better farm management (with seeds from high-yielding varieties, irrigation, fertilizer, and pest management).

Bangladesh is largely self-sufficient in rice production. However, challenges still lie ahead as Bangladesh becomes more densely populated, and there are still millions who cannot eat adequately in a day. In some years and in some seasons, the level of food security and hunger rises due to crop loss, low rice yield caused by natural disasters, market failure, socio-economic factors, and institutional weakness. The 2015 projections by the United Nations estimated that the population of Bangladesh in 2021 and 2050 would be about 170 million and 202 million respectively (UN, 2015). This is a big challenge for Bangladesh Government to meet the food needs of its increasing population.

Therefore, it is for government to impose policies to encourage and support domestic producers more to supply consumption needs and stop this product importing. Also, government and investors should try to decrease the production costs. We hope that the findings of this study can help the policy makers to set up a fruitful policy on rice agricultural production.

References

Amin, M. M. Amanullah & Akbar A. (2014), Time Series Modeling for Forecasting Wheat Production of Pakistan, *The Journal of Animal & Plant Sciences*, pp. 1444-1451.

Bangladesh (2009), “Bangladesh: A Country Study:Rice", *Library of Congress*, Washington, D.C. September 1988. Retrieved March 21, 2009.

BDU (2016), “Bangladesh Development Update: Economy Requires Focus on Sustainable and Inclusive Growth”, *The World Bank*, Feature Story published in April, 2016.

BRRI (2017), Bangladesh Rice Research Institute “Rice in Bangladesh”, *Bangladesh Rice Research Institute,* http://www.knowledgebank-brri.org/riceinban.php.

ICCB (2017) International Convention City Bashundhara, “The most influential platform for rice & grain industry”, *Rice & Grain Tech Expo 2017*

Maity Bipasha & Chatterjee Bani (2012), Forecasting GDP Growth Rates of India: An Empirical Study, *International Journal of Economics and Management Research*, Vol. 1, No. 9, 2012, pp. 52-58.

Murshed, S.M (2012). "Asiatic Society of Bangladesh". In Sirajul Islam and Ahmed A. Jamal. *Banglapedia: National Encyclopedia of Bangladesh* (Second ed.). Asiatic Society of Bangladesh.

RDRS (2012), “Recent Developments in the Rice Sector”, *Ricepedia*, Source: FAOSTAT, AQUASTAT database of the FAO, and national statistical sources as of September 2012.

Yaziz, S.R., Ahmad M. H., Nian L. C. and Muhammad N.,(2011), A Comparative Study on Box-Jenkins and Garch Models in Forecasting Crude Oil Prices, *Journal of Applied Sciences*, Vol-11, pp.1129-1135.