

Forecasting Aus Rice Area and Production in Bangladesh using Box-Jenkins Approach

N M F Rahman^{1*}, M M Hasan², M I Hossain², M A Baten³, S Hosen⁴, M A Ali⁵ and M S Kabir⁶

ABSTRACT

Forecasting of rice area and production is an essential procedure in supporting policy decision regarding food security and environmental issues. The main aim of this paper is to forecast Aus rice area and production in Bangladesh. Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) time-series methodology is considered for modeling and forecasting country's Aus rice area and production data commencing from 1971-72 to 2013-14. It was observed that ARIMA (1, 1, 5) and ARIMA (1, 1, 4) model were performed better than the other ARIMA models for forecasting Aus area and production respectively. The Aus rice areas were forecasted 1041.12 and 499.31 thousand hectares and for production it was 2059.642 and 1781.545 thousand metric tons in 2014-15 and 2023-24 respectively. The forecasted Aus rice area and production both were showed a decreasing trend. Government should create an enabling environment to develop HYV rice varieties for Aus rice crop to ensure ongoing food security by increasing cultivated area and production.

Key words: Aus rice, Box-Jenkins method, non-stationary, ACF and PACF

INTRODUCTION

Agro-based developing country like Bangladesh is striving hard for rapid development of its economy. The economic development of the country is mainly based on agriculture. The total cultivable area in Bangladesh is about 8.52 Mha and net cultivated area is 7.45 Mha and 0.47 Mha cultivable areas are unplanted. The contribution of agriculture sector in GDP is 19.29 percent. The crop sub-sector dominates with 13.44% from which rice contributes 46%. Almost 47.5% of manpower is directly involved in agriculture. In Bangladesh almost 66% of the labour force depends on agriculture for employment (BBS, 2012). Of all crops, rice plays the leading role by contributing 91% of total food grain production. More than 99% of the people eat rice as their main food @416 gm/person/day. Rice occupies about 76% of total cropped area from where modern varieties cover 84.67%

and local varieties cover 15.33%. The present status of area and production of rice is 11.42 million ha and 34.43 million MT (BBS, 2012). The projected population is expected to be 189.85 M by 2030 and it would require 28 million MT of clean rice or 42.50 million MT of paddy (BBS, 2012).

Three major rice crops namely, Aus, Aman and Boro constitute 100% of total rice production and grow in three different seasons. Aus is typically planted in March-April and harvested in June-July and practically harmonized with the climatic season hot summer (March-May). Aus rice occupies about 8.97% of total cropped area from where modern varieties cover 6.65% and local varieties cover 2.32%. And 6.88% of total production comes from Aus rice where modern varieties cover 5.91% and local varieties cover 0.98%. The present status of total area and production of Aus rice is 1.02 million ha and 2.37 million MT (BBS, 2012). The Aus rice area and production is

¹ and ²Agricultural Statistics Division, BRRI; ³Department of Statistics, Shahjalal University of Science and Technology (SUST), Sylhet; ⁴Grain Quality and Nutrition Division, BRRI; ⁵Director (Research), BRRI; ⁶Director (Administration and Common Service), BRRI, Gazipur, Bangladesh. *Corresponding author's E-mail: niaz.sust@gmail.com

decreasing continuously comparing to Boro, which is dominated rice crop in Bangladesh. Boro rice cultivation fully depends on irrigation and the pressure of ground water is increasing day by day and ground water level is going down but Aus rice requires only 5% supplement irrigation and the pressure of ground water is required to be low for Aus than Boro. It is necessary to transfer Boro cultivated area to Aus and also make sure the food security of the country. So if we do forecast the Aus rice area and production then it would help the policy makers to formulate decision in this regard.

The autoregressive integrated moving average (ARIMA) methodology developed by Box and Jenkins (1976) have been used by a number of researchers to forecast demands in terms of internal production, consumption, imports and exports (Muhammed *et al.*, 1992; Shabur and Haque, 1993; Sohail *et al.*, 1994). The forecasting studies on rice using ARIMA model are available by (Raghavender, 2009, 2010; Ravichandran, 2012; Biswas and Bhattacharyya, 2013; Singh *et al.*, 2013; Jambhulkar, 2013) in India; (Pakravan, 2011) in Iran; (Suleman and Sarpong, 2012) in Ghana; (Rahman, 2010; Awal and Siddique, 2011; Hamjah, 2014) in Bangladesh. Keeping the above requirement in view, the present study was carried out to develop an empirical model for forecasting from the observed Aus rice area and production data based on Box-Jenkins ARIMA method.

MATERIALS AND METHODS

Data source

The time series data of Aus rice area and production of Bangladesh for the periods 1971-1972 to 2013-2014, collected from 'Year Book of Agricultural Statistics' is published by Bangladesh Bureau of Statistics (BBS), Bangladesh, used in this study.

Time series model. In this study the methodology first refers to use of ARIMA model as propounded by Box-Jenkins for forecasting of Aus rice area and production in Bangladesh. The Box-Jenkins methodology refers to the set of procedures for identifying, fitting, and checking models with time series data.

(1) A pth-order autoregressive model: AR(p), which has the general form

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

where, Y_t = Response (dependent) variable at time t

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ = Response variable at time lags t-1, t-2, . . . , t-p, respectively.

$\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p$ = Coefficients to be estimated

ε_t = Error term at time t

(2) A qth-order moving average model: MA(q), which has the general form

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

where, Y_t = Response (dependent) variable at time t

μ = Constant mean of the process

$\theta_1, \theta_2, \dots, \theta_q$ = Coefficients to be estimated

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ = Errors in previous time periods that are incorporated in the response Y_t

ε_t = Error term at time t

(3) Autoregressive Integrated Moving Average Model: ARIMA (p,d,q), which has the general form

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

where, p, q and d denote the autoregressive, moving average and differenced order parameter of the process, respectively.

Diagnostic tests of residuals

Jarque-Bera test. The normality assumption is defined here using Jarque-Bera (1987) test,

which 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

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

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 two degrees of freedom, and can be used to test the hypothesis of skewness being zero and excess kurtosis being zero.

Box-Pierce Q test. In order to check the adequacy of the model using a chi-square test, known as the Box-Pierce Q statistic, (Box and Pierce, 1970) on the autocorrelations of the residuals, the test statistic Q is defined as

$$Q_m = n(n+2) \sum_{k=1}^m (n-k)^{-1} r_k^2 \approx \chi^2_{m-r}$$

where, r_k = the residual autocorrelation at lag k,
n= the number of residuals,
m= the number of time lags includes in the test.

If the p-value associated with the Q statistic is small (p-value $< \alpha$), the model is considered inadequate.

RESULTS AND DISCUSSIONS

Original series

Figure 1 presents the yearly cultivated area and yearly production of rice in Bangladesh from 1971-72 to 2013-14 along with the best (adjudged through R^2 value) fitted trend. Both the curve shows decreasing pattern and also both the trend curve shows cubic trend ($R^2=0.98$ for area and $R^2=0.89$ for production and significant at 95% confidence limit).

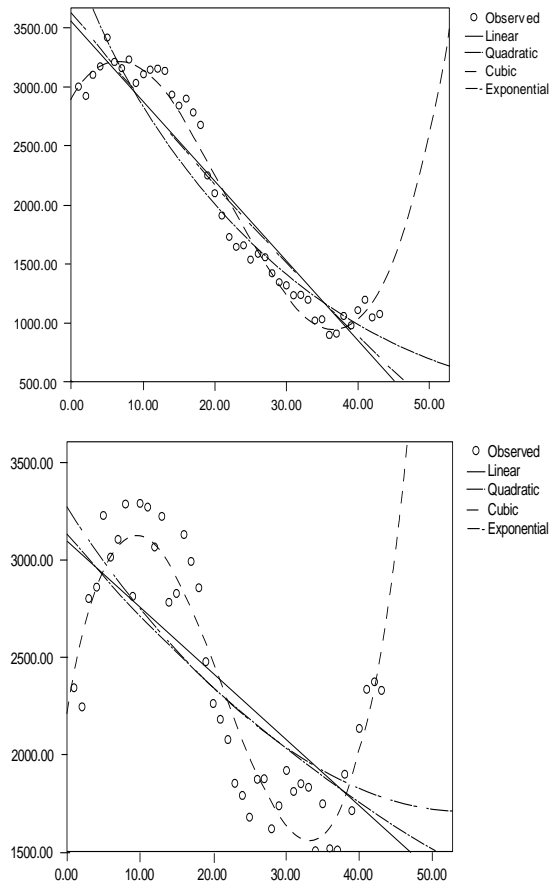


Fig. 1. Trend curve for yearly area and production of Aus rice in Bangladesh.

The normality of the original series has been tested by Jarque-Bera test for normality (Fig. 2). This test is based on the hypothesis that the underlined series is normal. P-value for Aus area and production is 0.072 and 0.147 respectively, indicates that both the series are normally distributed.

Test for stationarity

The assumption of Box-Jenkins ARIMA model is that the underlined time series should be stationary. To test the stationarity of the data, both informal and formal tests have been done.

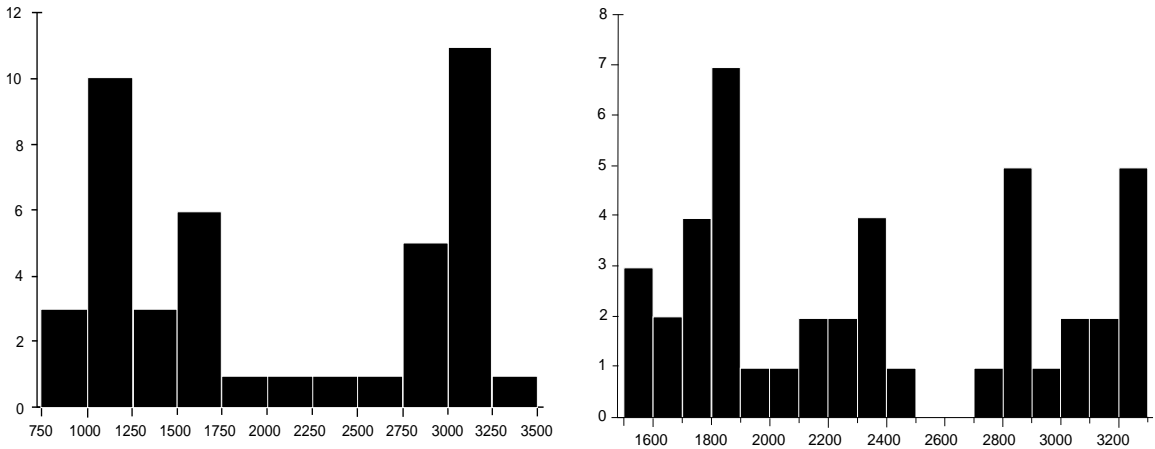


Fig. 2. Normality test of the original series for Aus area and production.

Informal approach (Correlogram). It is clear that from the constructed correlogram (Figs. 3-6) for the both series area and production of Aus rice that the autocorrelation function decaying gradually that indicates the series consists of higher order of moving average term and the fall off of the spike of partial autocorrelation function indicates that

there may be one autoregressive coefficient exists in the series. So, it is revealed from the correlogram that the original series for both area and production of Aus rice are not stationary but stationary after taking the first difference. These results support with work of Hamjah (2014).

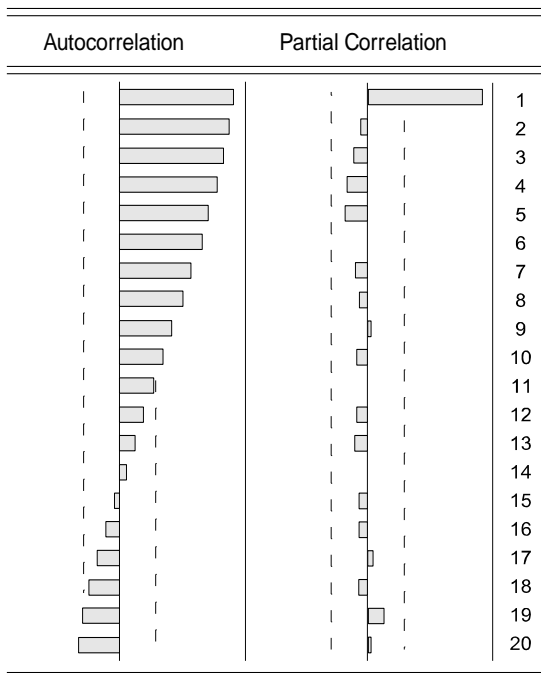


Fig. 3. Correlogram for Aus area (Original series).

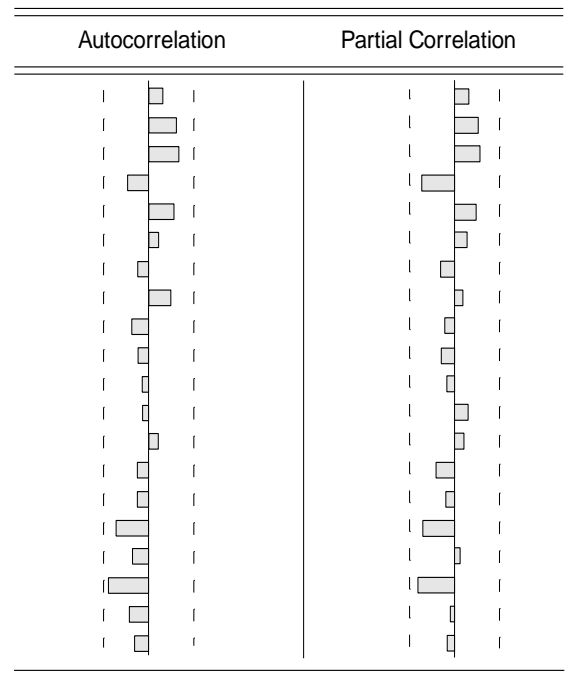


Fig. 4. Correlogram for Aus area (Differenced series).

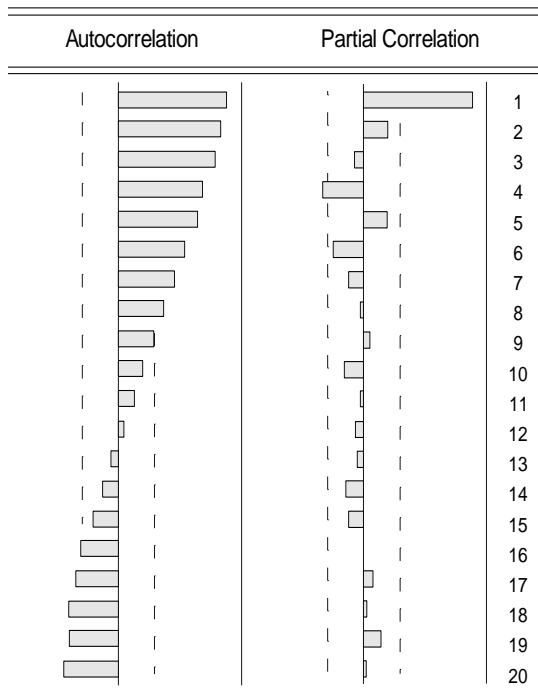


Fig. 5. Correlogram for Aus production (Original series).

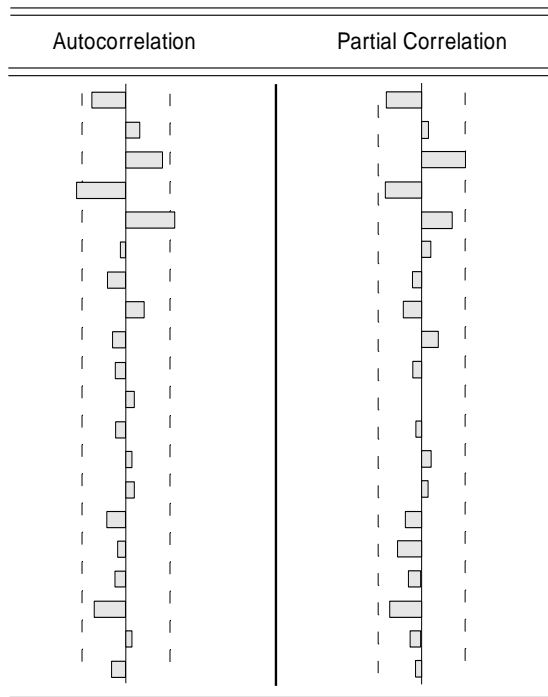


Fig. 6. Correlogram for Aus production (differenced series).

Formal approach (ADF test). The hypothesis of the Augmented Dekay Fuller test is:

H_0 : the series has a unit root, i.e. **non-stationary**

H_1 : the series has no unit root, i.e. **stationary**

Aus area	Original series	Differenced series
Augmented Dickey-Fuller test (P-value)	-0.35 (0.90)	-5.71 (0.000)

The non-stationary (P-value=0.90) original series become stationary after taking the difference of order 1.

Aus production	Original series	Differenced series
Augmented Dickey-Fuller test (P-value)	-1.32 (0.61)	-7.95 (0.000)

Under this hypothesis ADF depicted that at the level the series has a unit root (P-value=0.61) that means the series is not stationary. So, to make the series stationary we take the difference order one that makes the series stationary (P-value=0.000).

Model selection criteria

Some tentative ARIMA models for Aus area and production were considered and

the best fitted model is accepted on the basis of minimum RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), AIC (Akaike's Information Criterion), SBC (Schwarz's Bayesian Criterion), HQ criterion and maximum R^2 and adjusted R^2 value (Tables 1, 2 and 3).

Table 1. Comparison of different ARIMA models with model fit statistics for area of Aus rice.

Model	RMSE	MAE	MAPE	AIC	SBC	HQ criterion	R ²	Adj. R ²
ARIMA (1,1,4)	94.97	70.38	4.04	12.23	12.49	12.33	0.42	0.34
ARIMA (1,1,5)	84.93	69.54	4.02	12.06	12.36	12.17	0.54	0.46
ARIMA (1,1,6)	86.31	72.30	4.24	12.14	12.48	12.27	0.52	0.42

Table 2. Estimated parameters of ARIMA (1,1,5).

Variable	Coefficient	Std. error	t-statistic	Prob.
C	-60.20	32.51	-1.85	0.072
AR(1)	0.41	0.18	2.21	0.034
MA(1)	-0.20	0.11	-1.88	0.048
MA(3)	0.20	0.07	3.01	0.005
MA(5)	0.27	0.06	4.23	0.000
MA(6)	0.66	0.06	10.67	0.000
MA(7)	-0.62	0.10	-5.99	0.000
F-statistic	6.65	Durbin-Watson stat		1.87
Prob (F-statistic)	0.000102			

Table 3. Comparison of different ARIMA models with model fit statistics for production of Aus rice.

Model	RMSE	MAE	MAPE	AIC	SBC	HQ criterion	R ²	Adj. R ²
ARIMA (1,1,3)	175.04	149.72	6.67	13.46	13.71	13.55	0.49	0.42
ARIMA (1,1,4)	155.41	127.76	5.94	13.23	13.48	13.33	0.60	0.53
ARIMA (1,1,5)	156.69	128.10	5.98	13.27	13.56	13.38	0.60	0.53

For the series of production of Aus rice it is observed from the results that ARIMA (1,1,4) provides the best fitted model.

For the series of area of Aus rice it is observed from the results that ARIMA (1,1,5) provides the best fitted model.

From the above tables we have seen that entire coefficient of estimated model is significant at 5% level of significance. The R² of the estimated models are 0.54 and 0.53 that implies that about 54 and 53% variation the Aus area and production can be explained by the estimated coefficients and the rest are unexplained, may be due to

other factors that are not encountered in this model. Test of autocorrelation was done by Durbin-Watson test and the value of D-W statistics suggested that the estimated coefficients are free from autocorrelation problem. The minimum value of RMSE, MAE, MAPE, AIC, SBC and H-Q criterion are also confirmed. The existing results are supported by Hamjah (2014) and Biswas and Bhattacharyya (2013).

Results of empirical models. Estimation of parameters for area (Table 2) and production (Table 4) suggests the models may be represented mathematically as:

Aus rice area	$\hat{Y}_t = -60.20 + 0.40Y_{t-1} + 0.20\varepsilon_{t-1} - 0.20\varepsilon_{t-3} - 0.27\varepsilon_{t-5} - 0.66\varepsilon_{t-6} + 0.622\varepsilon_{t-7}$
Aus rice production	$\hat{Y}_t = -30.89 - 0.587Y_{t-1} - 0.63\varepsilon_{t-1} - 0.52\varepsilon_{t-3} - 0.55\varepsilon_{t-5} - 0.89\varepsilon_{t-6}$

The time series models those are tested and found satisfactory in all stages of model fitting process being used for estimation of the time series coefficient. These models would be

used for forecast purpose, which is the ultimate goal of univariate time series analysis. Moreover, the residual analysis of the estimated models confirmed the stability

Table 4. Estimated parameters of ARIMA (1,1,4).

Variable	Coefficient	Std. error	t-statistic	Prob.
C	-30.89	59.52	-0.51	0.607
AR(1)	-0.581	0.13	-4.22	0.001
MA(1)	0.626	0.042	14.93	0.000
MA(3)	0.516	0.064	8.00	0.000
MA(5)	0.545	0.043	12.60	0.000
MA(6)	0.891	0.029	30.03	0.000
F-statistic	10.11	Durbin-Watson stat		1.82
Prob (F-statistic)	0.000175			

to forecast and others relevant diagnosis of the fitted models. The works of Awal and Siddique (2011), Biswas and Bhattacharyya (2013) and Hamjah (2014) were also observed as the satisfactory and adequate model for rice production forecasting.

Results of residual analysis. The residuals correlogram is a good measure to have the idea about the adequacy of the fitted model (Fig. 7). The constructed correlogram by using the residuals from the fitted models indicate that all the series are free from autocorrelation problem since all the spikes

are laying by the limit of permissible lines and that has also been tested by the Q-stat.

Normality test of residuals. Normality of the residuals is a vital issue since it is expected that produced residuals from the model used for estimation process would be dispersed consistently. By assuming the hypothesis that the series under study are normally distributed after being tested and found that all the models produced normal residuals implying the adequacy of the models (Figs. 8 and 9).

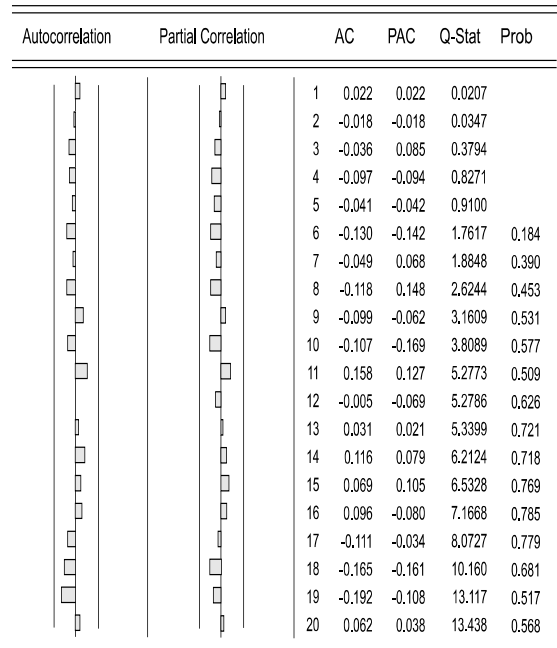
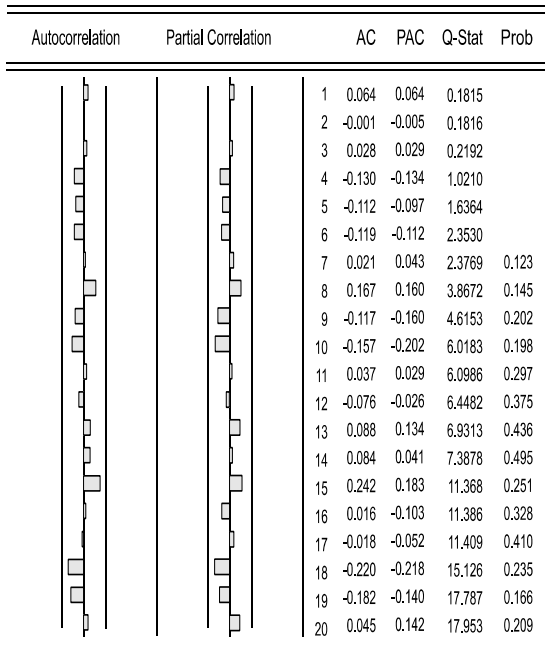


Fig. 7. Correlogram for the residuals of Aus area and production.

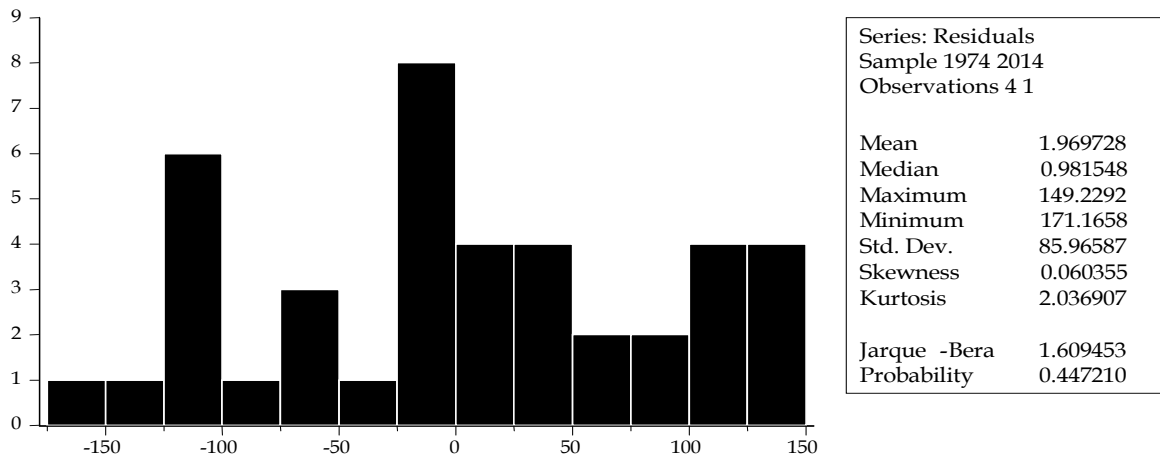


Fig. 8. Normality of residuals for Aus area.

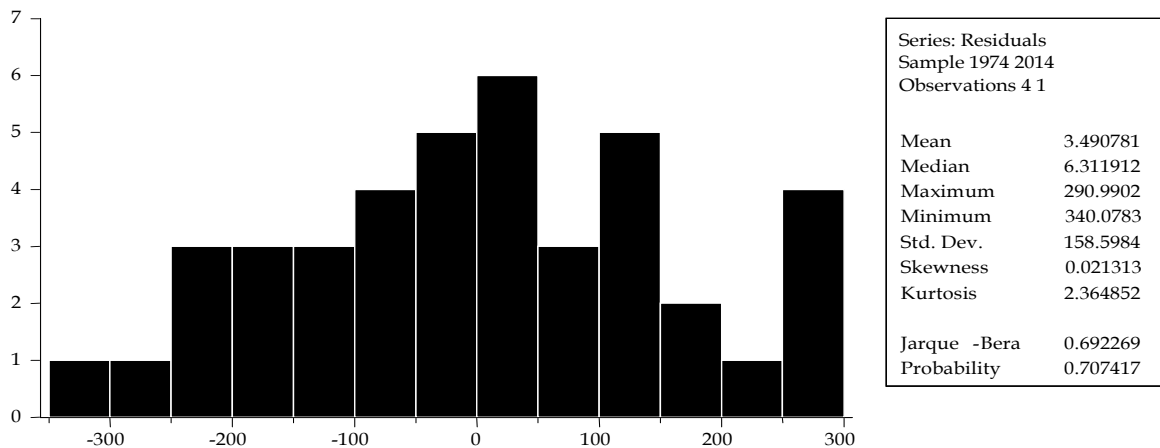


Fig. 9. Normality of residuals for Aus production.

Breusch-Godfrey serial correlation LM test. Breusch-Godfrey serial correlation LM test is based on the hypothesis that the residuals are not serially correlated. The

estimated p-value for Aus area and production (p-value=0.65 and 0.94) confirms that the residuals are free from serial correlation.

Aus area	F-statistic	0.428913	Prob. F(2,32)	0.6549
Aus production		0.061797	Prob. F(2,33)	0.9402

Forecasting

Projections have been made for the Aus area and production on the basis of their actual performance during 1971-72 to 2013-14. Table 5 shows these projections. Since the time series model building is an iterative process, until and unless the forecast value found by maintain the proper statistical

rigor it is iterated from the identification process. The same things followed here and used the models those strictly avail the criteria for having better forecast. Forecast of Aus rice area and production showed a decreasing trend. The same result for Aus rice production is found in Hamjah (2014) and Awal and Siddique (2011).

Table 5. Forecasted value of area (000' ha) and production (000' ton) of Aus rice.

Year	ARIMA (1,1,5)			ARIMA (1,1,4)		
	LPL	Forecast	UPL	LPL	Forecast	UPL
2015	855.12	1041.12	1227.12	1773.64	2059.64	2345.64
2016	794.92	980.92	1166.92	1742.74	2028.74	2314.74
2017	734.72	920.72	1106.72	1711.84	1997.84	2283.84
2018	674.52	860.52	1046.52	1680.94	1966.94	2252.94
2019	614.32	800.32	986.32	1650.04	1936.04	2222.04
2020	554.11	740.11	926.11	1619.14	1905.14	2191.14
2021	493.91	679.91	865.91	1588.24	1874.24	2160.24
2022	433.71	619.71	805.71	1557.34	1843.34	2129.34
2023	373.51	559.51	745.51	1526.45	1812.45	2098.45
2024	313.31	499.31	685.31	1495.55	1781.55	2067.55

LPL: Lower Predictive Value; UPL: Upper predictive value

CONCLUSION

The Box-Jenkins approach as an ARIMA model was used for the patterns of the past movement of a variable to forecast the future values. In this study, a short run forecasting of Aus rice area and production model were formulated. To select the best model for forecasting the latest available model selection criteria such as MSE, RMSE, MAE, MAPE, AIC, SBC, H-Q criterion, R square and Adjusted R square were used. The best selected ARIMA model for forecasting Aus area and production were observed ARIMA (1,1,5) and ARIMA (1,1,4) respectively. The Aus rice area and production of Bangladesh were observed 1041.12 thousand hectare and 2059.64 thousand metric tons in 2014-15 and it were forecasted 499.31 thousand hectare and 1781.55 thousand metric tons for 2023-2024. The forecasted Aus rice area and production showed a decreasing trend. These projections help the government to make policies with regard to relative price structure, production and consumption and also to establish relations with other countries of the world.

REFERENCES

- Awal, M A and M A B Siddique. 2011. Rice production in Bangladesh employing by ARIMA model. *Bangladesh Journal of Agricultural Research*, 36:51-62.
- BBS (Bangladesh Bureau of Statistics). 2012. *Yearbook of Agricultural Statistics of Bangladesh*. Government of Bangladesh.
- Biswas, R and B Bhattacharyya. 2014. ARIMA Modeling to Forecast Area and Production of Rice in West Bengal. *Journal of Crop and Weed*, 9:26-31.
- Box, G E P and G M Jenkin. 1976. *Time Series of Analysis, Forecasting and Control*, Sam Franscico, Holden-Day, California. USA.
- Box, G E P and D A Pierce. 1970. Distribution of residual autocorrelations in Autoregressive-integrated moving average Time Series Models, *Journal of the American Statistical Association*, 65: 1509-1526.
- BRRI (Bangladesh Rice Research Institute). 2012. *Annual Report 2011-12*. BRRI, Gazipur, Bangladesh.
- Hamjah, M A. 2014. Rice Production Forecasting in Bangladesh: An application of Box-Jenkins ARIMA model. *Mathematical theory and modeling*, 4:1-11.
- Jambhulkar, N N. 2013. Modeling of rice production in Punjab using ARIMA model. *IJSR-International Journal of Scientific Research*, 2:1-2
- Jarque, C M and A K Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55: 163-172.
- Muhammed, F, M Siddique, M Bashir and S Ahamed. 1992. Forecasting rice production in Pakistan using ARIMA models. *J. Animal Plant Sci.*, 2: 27-31.
- Pakravan, M R, M K Kelashemi and H R Alipour. 2011. Forecasted Iran's rice imports trend during 2009-2013. *International Journal of Agricultural Management and Development*, 1: 39-44.

- Raghavender, M. 2009. Forecasting paddy yield in Andhra Pradesh using seasonal time series model, *Bulletin of pure and applied sciences*.
- Raghavender, M. 2010. Forecasting paddy production in Andhra Pradesh with ARIMA Model. *Int. J. Agricult. Stat. Sci.*, 6: 251-258.
- Rahman, N M F. 2010. Forecasting of Boro rice production in Bangladesh: An ARIMA approach, *J. Bangladesh Agril. Univ.*, 8: 103-112.
- Ravichandran, S, P Muthuraman and P R Rao. 2012. Time - Series modelling and forecasting India's rice production - ARIMA Vs STM modelling approaches. *Int. J. Agricult. Stat. Sci.*, 8:305-311.
- Shabur, S A and M E Haque. 1993. Analysis of rice in Mymensing town market pattern and forecasting. *Bangladesh J. Agric. Econ.*, 16: 130-133.
- Singh, D P, Kumar, P and K Prabakaran. 2013. Application of ARIMA model for forecasting Paddy production in Bastar division of Chhattisgarh. *American International Journal of Research in Science, Technology, Engineering & Mathematics*, 5: 82-87.
- 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, 13: 140-146.
- Suleman, N and S Sarpong. 2012. Forecasted milled rice production in Ghana using Box-Jenkins approach. *International Journal of Agricultural Management and Development*, 2: 79-84.