Pak. J. Agri. Sci., Vol. 53(4), 759-765; 2016
ISSN (Print) 0552-9034, ISSN (Online) 2076-0906
DOI: 10.21162/PAKJAS/16.1892
http://www.pakjas.com.pk

FORECASTING AREA AND PRODUCTION OF BLACK GRAM PULSE IN BANGLADESH USING ARIMA MODELS

Niaz Md. Farhat Rahman[1] and Md. Azizul Baten[2]

[1]Agricultural Statistics Division, Bangladesh Rice Research Institute (BRRI) Gazipur-1701, Bangladesh; [2]Department of Decision Science, School of Quantitative Sciences, Universiti Utara Malaysia, 01600, Sintok, Kedah Darul Aman, Malaysia;
[*]Corresponding author's e-mails: baten_math@yahoo.com

This study attempted to shed light on the issues such as forecasting of area and production of black gram pulse in Bangladesh. Data on area and production of black gram pulse collected over a period of 47 years (1967-68 to 2013-14) from Bangladesh Bureau of Statistics. Autocorrelation function, partial autocorrelation function and Phillips–Perron unit root test are calculated for obtaining stationarity and identifying tentative autoregressive & moving average orders for the data. Appropriate Box-Jenkins autoregressive integrated moving average model is fitted. The test statistic Ljung-Box Q is used to check the model adequacy, Jarque-Bera test, is used to check the normality of the models while Akaike Information Criterion and Bayesian Information Criterion are used to determine the best forecasting model. ARIMA (0,1,0) model was found suitable for both black gram pulse area and production. The performances of model were validated by comparing with percentage deviation from the actual values and mean absolute percent error (MAPE). Finally, a forecast for the period 2014-15 to 2018-19 was made.

Keywords: *Vigna mungo* L., Box-Jenkins method, non-stationary, autocorrelation function, partial autocorrelation function.

INTRODUCTION

Bangladesh is an agro-based developing country. The growth and stability of the economy of Bangladesh primarily depends largely on the growth of agriculture. The contribution of agriculture sector in GDP is 13.09 percent. In the agriculture sector, the crop sub-sector dominates with 10.74 percent in GDP (BER, 2014).

Black gram (*Vigna mungo* L. Hepper) is one of the highly prized pulses in Bangladesh. It is locally known as 'Mashkalai' belongs to the family Leguminosae and it is extensively grown in both tropical and sub-tropical countries of the world. It contains approximately 25-28% protein, 4.5-5.5% ash, 0.5-1.5% oil, 3.5-4.5% fiber and 62-65% carbohydrate on dry weight basis. It has been reported that the average yield of blackgram is about 1000 kg ha-1 and the protein content is 25-26% (BINA, 2004). In spite of its various uses, its cultivation is decreasing day by day both in acreage and yield (BBS, 2006). Factually other crops in general and black gram pulse in particular provide linkages through which it can stimulate economic growth in other sectors. Black gram pulse cultivation has been suffering from various problems, such as traditional methods of farming, low yields, shortage of key inputs and shortage of irrigation water. Bangladesh has experienced ups and downs in black gram pulse production. On the other hand, farmers do not know future prospect of pulse production and prices while deciding to cultivate this and other crops. Pulses contribute 2.3 per cent value added to agriculture. Total cultivated area in Bangladesh is 8031161 hectares of which pulses constitute 420763 hectares i.e., 5.24 per cent of total cultivated land (BBS, 2004). Most of the people (75 per cent) live in rural areas and 62 per cent of the total labor force of the country is employed in agricultural sectors. Agriculture contributes 15.23 percent to the gross domestic product (GDP) in 2009-10 where crops contribute 11.43 percent to the GDP (BBS, 2010). About a dozen pulse crops are grown in the winter and summer seasons. Among these, Grass pea, lentil, chickpea, black gram, mung bean, field pea, cowpea, and fava bean are grown during the winter season (November–March). Collectively, they occupy 82 percent of the total pulse-cultivation area and contribute 84 percent of the total pulse production. Black gram and mung bean can also be grown in late winter (June–March) in southern areas such as the Bhola, Barisal, and Chittagong districts. These are under minor crops in Bangladesh but it has a great importance in the dietary menu, it is the great source of plant protein for the people who are not able to get regularly animal protein (from meat and fish) because of more expensive food item. In Bangladesh, where 42 per cent people are living under the poverty line (BBS, 2005), and pulses are popular and common food, people take this food almost alternate a day, so, this can play an important role to reduce the malnutrition for the poor people of the country if it becomes available to that type of people.

Forecasting the crop yield or any agricultural production is a formidable challenge. Accurate forecasting is important to both government and industry that needs to predict future production of food grains. The forecast would thus help save much of the precious resources of any country, like Bangladesh which otherwise might be wasted. 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 to adopt appropriate solutions (Muhammed *et al.*, 1992; Shabur and Haque, 1993; Sohail *et al.*, 1994). Rahman (2010) studied the forecasting of Boro rice production in Bangladesh. Awal and Siddique (2011) studied the rice production in Bangladesh using ARIMA model. They examined the best ARIMA model for forecasting Aus, Aman and *Boro* rice production in Bangladesh. Hamjah (2014) studied on Rice Production Forecasting in Bangladesh using Box-Jenkins ARIMA Model. The forecasted results for Aus production showed initially a downward tendency. Raghavender (2010) presented an ARIMA model for the forecasting of rice yield per hectare in Andhra Pradesh based on the data set during 1955 to 2007. ARIMA models were used to forecast pigeon pea production in India. Sarika (2011) studied about modelling and forecasting of pigeonpea (*Cajanus cajan*) production using autoregressive integrated moving average methodology in India. Ravichandran (2012) used data for the period from 1980-1981 to 2009-2010 and showed time series modelling and forecasting India's rice production for 2010-2011. Prabakaran (2014) used the ARIMA model for forecasting of pulses area and production in India. Jambhulkar (2013) estimated rice production in Punjab using ARIMA Model. Pakravan (2011) examines the ARIMA model that is used to make efficient forecast of Iran's Rice Imports trend during 2009-2013. Suleman (2012) forecasted Milled Rice Production in Ghana Using Box-Jenkins Approach. There is a dire need to forecast area and production of black gram pulse in Bangladesh and also a very little work has been done in pulse area and production especially in black gram pulse. Therefore, in this study, we modeled and forecasted the area and production of black gram pulse in Bangladesh using ARIMA methodology. This would help predict future values of area and production of black gram pulse in the country.

MATERIALS AND METHODS

***Data source*:** The time series data of black gram pulse production of Bangladesh for the periods 1967-1968 to 2012-2013, collected from "Year Book of Agricultural Statistics" is published by Bangladesh Bureau of Statistics (BBS), Bangladesh, used in this study. The data pertaining to the agricultural years 1967-1968 to 2010-2011 was used for the model building and forecasting. The data of 2011-12 to 2013-2014 was used for validation of the model. Finally, a forecast for the period 2014-2015 to 2018-2019 was made.

***Time series model*:** Autoregressive Integrated Moving Average (ARIMA**)**

In this study the methodology first refers to use of ARIMA model as propounded by Box-Jenkins (Box and Jenkins, 1976) for forecasting of black gram pulse production in Bangladesh. The Box-Jenkins methodology refers to the set of procedures for identifying, fitting or parameter estimating and diagnostic checking models with time series data and then forecasts follow directly from the fitted model.

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

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

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

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

$\varphi_0, \varphi_1, \varphi_2, ..., \varphi_p$ = 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

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

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

μ = Constant mean of the process

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

$\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$ = 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 which has the general form

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

Y_t is called Auto-Regressive Integrated Moving Average (ARIMA) Model. It is denoted by ARIMA (p, d, q). Where "p" is the order of AR process, "q" is the order of MA process and "d" is the order of differencing Y_t (Anderson, 1971).

Building ARIMA Model:

***Identification stage*:** Identification step involves the use of the techniques to determine the values of p,q and d. The values are determined by using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). For any ARIMA (p, d, q) process, the theoretical PACF has non-zero partial autocorrelations at lags 1, 2,...,p and has zero partial autocorrelations at all lags, while the theoretical ACF has non zero autocorrelation at lags 1, 2,…,q and zero autocorrelations at all lags. The nonzero lags of the sample PACF and ACF are tentatively accepted as the p and q parameters. For a non stationary series the data is differenced to make the series stationary. The number of times the series is differenced determines the order of d. Thus, for a stationary data d = 0 and ARIMA (p, d, q) can be written as ARMA (p, q).

***Estimation stage*:** ARIMA models are fitted and accuracy of the model was tested on the basis of diagnostics statistics.

***Diagnostic checking*:** The best model was selected based on the following diagnostics.

(i*) Akaike Information Criterion (AIC)*: Akaike information criterion, one of leading statisticians, provides guide lines for choosing the best possible model from a set of competing models. It is defined as $AIC = n\log(MSE) + 2k$ where, n is the sample size, MSE is the mean square error and k is the total number of estimable parameters. Akaike mentioned that the model with minimum AIC is closer to the best possible choice.

***Schwartz information criterion (SIC)*:** This is called Bayesian Information criteria. This is defined as $BIC = n\log(MSE) + k\log n$, where, n is the sample size, MSE is the mean square error and k is the total number of estimable parameters. Schwartz shows that BIC is better than AIC. The model with minimum BIC is assumed to describe the data series more adequately.

(*ii) Insignificance of autocorrelations for residuals*: If a model is an adequate representation of a time series, it should capture all the correlation in the series, and the white noise residuals should be independent of each other.

(*iii) Significance of the parameters*: Significance tests for parameter estimates indicate whether some terms in the model might be unnecessary.

***Forecasting stage*:** Future values of the time series are forecasted.

***Model evaluation*:** The Mean Absolute Percent Prediction Error is defined as $MAPPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|\varepsilon t|}{yt}\times 100$. The model with minimum MAPPE is assumed to describe the data series adequately.

R software was used for time series analysis and developing ARIMA models and forecasting.

***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 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.

***Box-Price Q test*:** In order to check the adequacy of the model using a chi-square test, known as the Box-Pierce Q statistic, on the autocorrelations of the residuals, the test statistic Q is defined as

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

Where, rk(e) = 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 < ·), the model is considered inadequate. The analyst should consider a new or modified model and continue the analysis until a satisfactory model has been determined.

Moreover, the properties of the residual can be checked by considering the graph of normality of the residual.

RESULTS AND DISCUSSION

Descriptive statistics for the time series data of black gram area and production for Bangladesh is given in Table 1. The time series data is plotted in Figure 1.

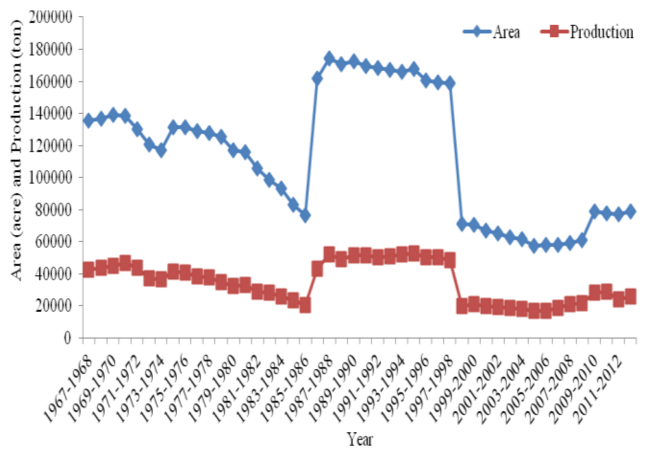

Figure 1. Area and production of black gram pulse in Bangladesh.

The maximum area and production of Black gram pulse were observed 174187 acre in 1988 & 52540 ton in 1995 and minimum area and production were 57360 acre in 2005 & 17000 metric ton in 2006. At present, the area & production of Black gram pulse are 79000 acre & 26000 metric ton in

Table 1. Descriptive statistics for the time series data of black gram area and production in Bangladesh from 1967-1968 to 2013-2014.

Parameters	Mean	Median	Standard deviation	Kurtosis	Skewness
Area	114249.43	118785	40560.19	-1.48	0.003
Production	34928.15	35774	12373.48	-1.53	0.013

2013. The time series plotted in figure 1 for the area and production of Black gram pulse showed a very short term movement.

ARIMA Model:

Identification**:** Last 46 years data of area and production of Black gram pulse in Bangladesh were used for the modeling purpose. In model specification, the plots of autocorrelation function (ACF) and partial auto correlation function (PACF) for Black gram area and production are represented in Figure 3 and Figure 4. Auto correlation function indicated the order of the auto regression compounds "q" of the model while the partial correlation function gave an indication for the parameter "p" i.e. moving average order. From the correlogram it has been clear that both the series are moving average series since the partial autocorrelation is tappers off quickly after first lag but the autocorrelation is dies of gradually and three significant spike of autocorrelation have been found for both cases. So, initially both the series seems to be moving average time series of order 3.

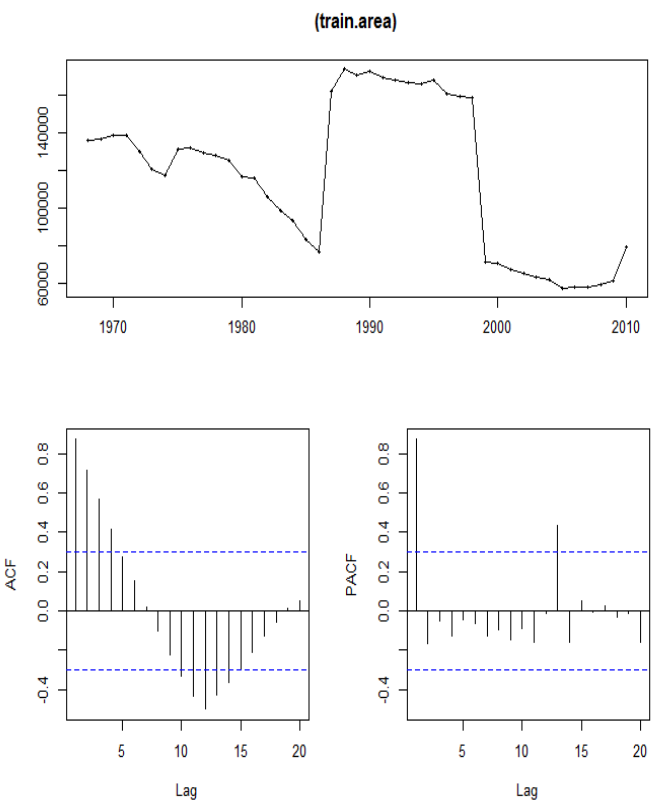

Figure 2. Correlogram for the original series of Black gram pulse area.

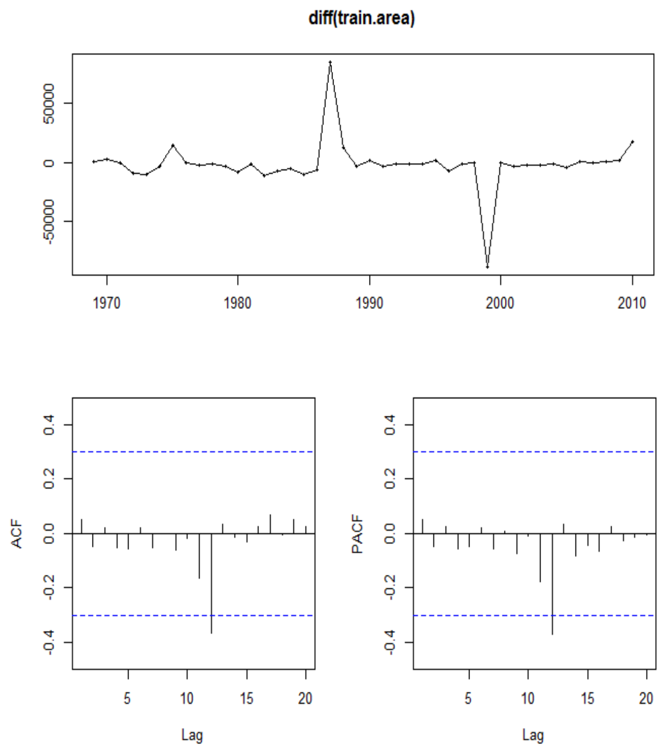

Figure 3. Correlogram for the differenced series of Black gram pulse area.

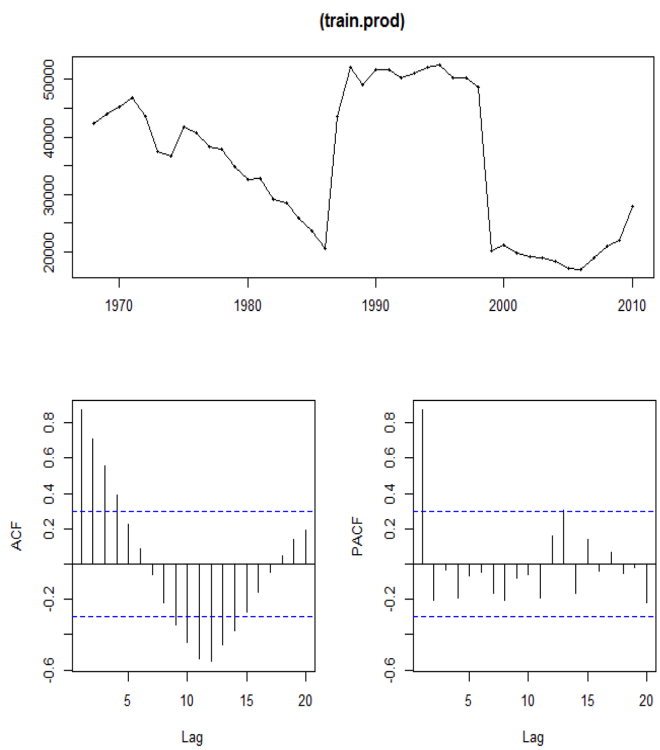

Figure 4. Correlogram for the original series of Black gram pulse production.

***Stationarity test for original series*:** The stationarity of the data has been carried out
by using the Phillips–Perron Unit Root Test.

Phillips–Perron Unit Root Test	Black gram pulse	Z(Alpha) Statistic	Prob.*
	Area	-7.51	0.66
	Production	-8.05	0.63

The original series is not stationary.

Based on the Phillips–Perron Unit Root test is has been found that the series is non-stationary at level. To confirm the stationarity of the series which is the prerequisite for the Box-Jenkings ARIMA model to estimate the parameter, the differences of the original series are considered. From the correlogram of the 1st difference of the original series for Black gram area and production (Figure 3 and Figure 5) shows that the series is stationary. The Phillips–Perron Unit Root test also confirmed the stationarity of the differenced series. This implies that the black gram area and production series are integrated of order one.

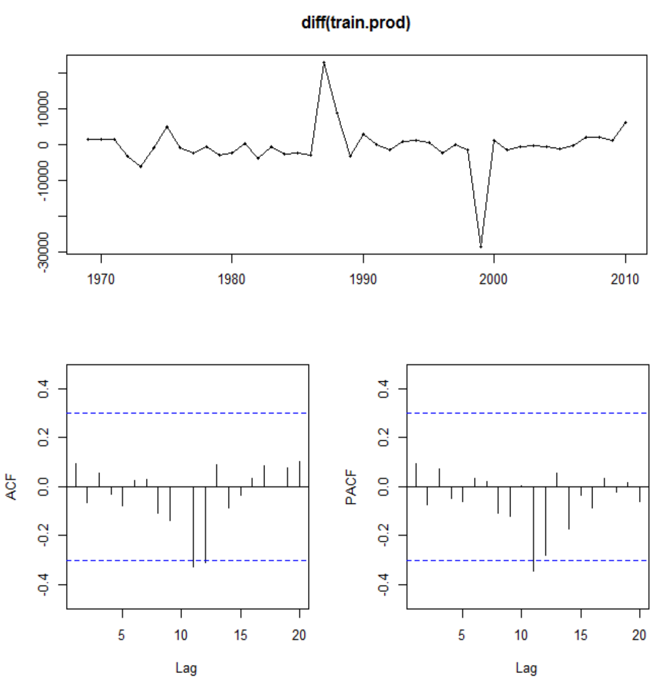

Figure 5. Correlogram for the differenced series of Black gram pulse production.

Stationarity test for 1st difference series.

Phillips–Perron Unit Root Test	Black gram pulse	Z(Alpha) Statistic	Prob.*
	Area	-38.15	0.01
	Production	-36.19	0.01

The first differenced series is stationary.

***Diagnostic checking and model estimation*:** Usually the different models for area and production of Black gram pulse were fitted using different significant values of p and q. ARIMA (0,1,0) were selected as the best model for area and production of Black gram pulse based on the minimum values of AIC and BIC.

Table 2. Autoregressive integrated moving average (ARIMA) models fitted for time series data on black gram area and production and corresponding selection criterion, that is, Akaike information criteria (AIC) and Schwarz-Bayesian information criteria (SBC).

Parameters	**ARIMA models**	**AIC**	**BIC**
Area	ARIMA (0,1,0)	799.36	800.79
	ARIMA (1,1,0)	801.49	804.24
	ARIMA (0,1,1)	801.50	804.23
Production	ARIMA (0,1,0)	718.15	719.58
	ARIMA (1,1,0)	720.10	722.83
	ARIMA (0,1,1)	720.04	722.77

***Residual analysis*:** Figures 6 & 7 are shows the relevant statistics for the residuals of the estimated models. From Figure 6 and Figure 7, the graph of ACF and PACF of the residual in terms of both the ACF and PACF of area and production of Black gram pulse are observed within the boundary of the permissible limit for up to 11 lag orders, which implies the residuals are free from autocorrelation.

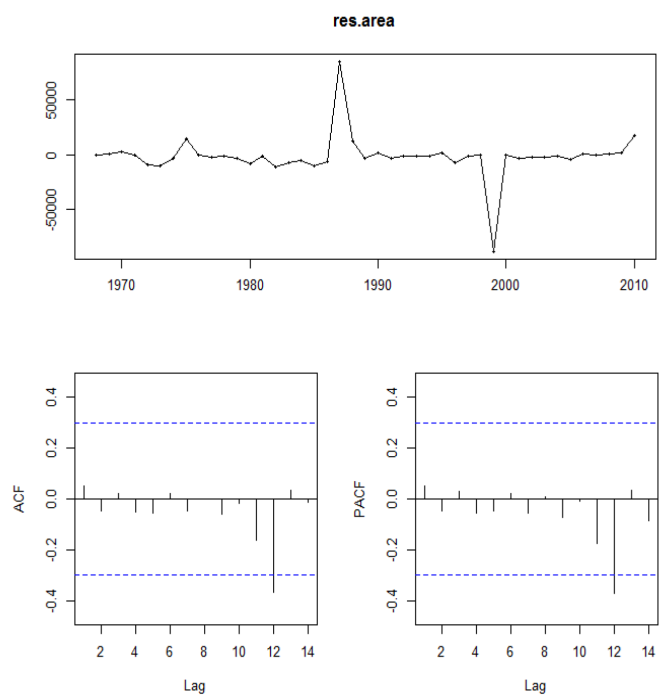

Figure 6. Normality of the residuals of Black gram pulse area.

Table 3. Performance of autoregressive integrated moving average (ARIMA) models for black gram pulse area and production for Bangladesh.

Parameters	Years of prediction	Actual values	Forecasted values	% error in prediction (±)	MAPE in prediction (%)
Area	2011-12	78000	76237	2.26	7.91
	2012-13	77000	73409	4.66	
	2013-14	79000	70514	10.74	
Production	2011-12	29000	27604	4.81	10.17
	2012-13	24000	27206	-13.36	
	2013-14	26000	26806	-3.10	

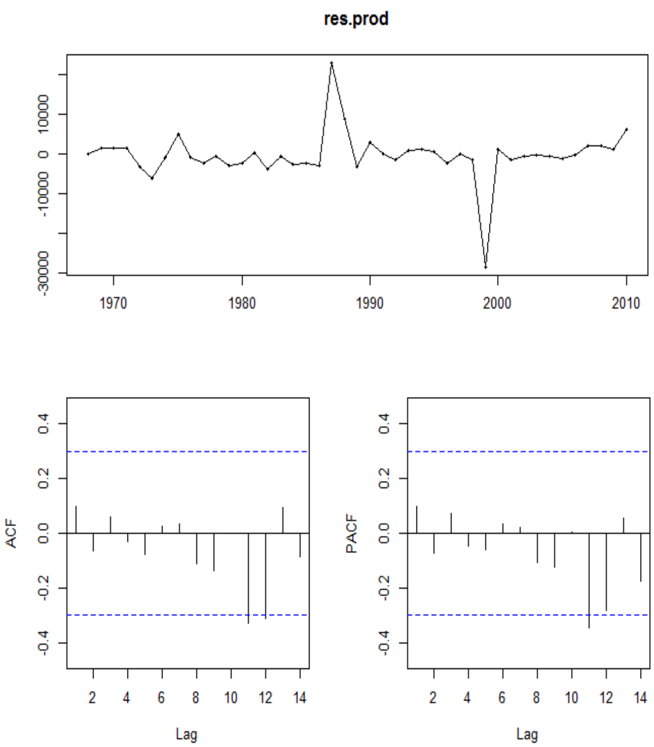

Figure 7. Normality of the residuals of Black gram pulse production.

To check the Autocorrelation assumption, "Box-Ljung" test is also used for both cases. This test strongly suggests that there is no autocorrelation among the residuals of the fitted ARIMA (0,1,0) models for area and production of Black gram pulse at 5% level of significance. The normality if the residuals have been checked by Jarque-Bera normality test. The test shows that the residuals of both the series are normally distributed.

Box-Ljung Test	Black gram pulse	X-squared Statistic	P value
	Area	0.13	0.722
	Production	0.42	0.517

There is no autocorrelation among the residuals.

Jarque Bera Test	Black gram pulse	X-squared Statistic	P value
	Area	409.11	0.00
	Production	268.86	0.00

The residual series are normally distributed.

***Forecast using ARIMA model*: **The actual and forecasted values for black gram pulse area and production along with percentage of deviation are presented in Table 4. The forecasted values of cultivable area of black gram pulse, for the years 2011- 12, 2012-13, and 2013-14 were 76237, 73409 and 70514 acre with the deviation of 2.26, 4.66, and 10.74% respectively. Similarly, the forecasted values for black gram production for the years 2011- 12, 2012-13, and 2013-14 were 27604, 27206 and 26806 ton with the deviation of 4.81, -13.36, and -3.10%, respectively. Negative value in % deviation showed that predicted values were higher than the actual values. From the mean absolute percentage error (MAPE) it has been clear that about 92.09% of area and 89.83% of production are accurately predicted by the fitted models.

The principal objective of developing an ARIMA model for a variable is to generate post sample period forecasts for that variable. ARIMA (0, 1, 0) model has been select as the candidate model for the purpose of forecast the Black gram area and production. The forecasts for Black gram pulse area and production during 2014-15 to 2018-19 are presented in Table 4 where it has the decreasing trend and it has decreased for the year 2014-15 to 2018-19 compared to 2013-14. Table 4 showed the value of the forecasted Black gram pulse area and production, respectively.

Table 4. Forecasted values of black gram pulse area and production in Bangladesh.

Year of prediction	Black gram pulse Area (acre)	Black gram pulse Production (ton)
2014-15	76778	25531
2015-16	74514	25057
2016-17	72209	24578
2017-18	69861	24095
2018-19	67472	23607

***Conclusion*:** The Box-Jenkins approach as an ARIMA model was used the patterns of the past movement of a variable to forecast the future values. In this study, a short run forecasting of Black gram area and production model were formulated. The forecasted Black gram area and production showed a decreasing trend. Based on the forecasting and validation results, it may be concluded that ARIMA model could be successfully used for forecasting black gram area and production of Bangladesh for the immediate subsequent years. 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

Anderson, T.W. 1971. The statistical analysis of time series. John Wiley, New York.

Awal, M.A. and M.A.B. Siddique. 2011. Rice production in Bangladesh employing by ARIMA model. Bang. J. Agric. Res. 1:51-62.

Bangladesh Bureau of Statistics. 2004. Yearbook of Agricultural Statistics of Bangladesh, Ministry of Planning, Govt. of Bangladesh, Dhaka, Bangladesh.

Bangladesh Bureau of Statistics. 2005. Yearbook of Agricultural Statistics of Bangladesh, Ministry of Planning, Govt. of Bangladesh, Dhaka, Bangladesh.

Bangladesh Bureau of Statistics. 2006. Yearbook of Agricultural Statistics of Bangladesh, Ministry of Planning, Govt. of Bangladesh, Dhaka, Bangladesh.

Bangladesh Bureau of Statistics. 2010. Yearbook of Agricultural Statistics of Bangladesh, Ministry of Planning, Govt. of Bangladesh, Dhaka, Bangladesh.

Box, G.E.P. and G.M. Jenkin. 1976. Time series of analysis, forecasting and control. San Franscico, Holden-Day, California, USA.

Bangladesh Institute of Nuclear Agriculture. 2004. Annual Report 2003-04. Ministry of Agriculture, Government of Bangladesh, Dhaka, Bangladesh.

Bangladesh Rice Research Institute. 2012. Annual Report 2011-12. Ministry of Agriculture, Government of Bangladesh, Dhaka, Bangladesh.

Government of Bangladesh. 2014. Bangladesh Economic Review. Ministry of Planning. Government of Bangladesh, Dhaka, Bangladesh.

Hamjah, M.A. 2014. Rice production forecasting in Bangladesh: An application of Box-Jenkins ARIMA model. Math. Theo. Model. 4:1-11.

Jambhulkar, N.N. 2013. Modeling of rice production in Punjab using ARIMA model. Int. J. Sci. Res. 2:1444-1451.

Jarque, C.M. and A.K. Bera. 1987. A test for normality of observations and regression residuals. Int. Stat. Rev. 55:163-172.

Muhammed, F., M. Siddique, M. Bashir and S. Ahamed. 1992. Forecasting rice production in Pakistan using ARIMA models. J. Anim. 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. Int. J. Agric. Manage. Develop. 1:39-44.

Prabakaran, K., P. Nadhiya, S. Bharathi, and M. Isaivani. 2014. Forecasting of pulses area and production in India- An ARIMA approach. Ind. Stream Res. J. 4:1-7.

Raghavender, M. 2010. Forecasting paddy production in Andhra Pradesh with ARIMA model. Int. J. Agric. Stat. Sci. 6:251-258.

Rahman, N.M.F. 2010. Forecasting of boro rice production in Bangladesh: An ARIMA approach. J. Bang. 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. Agric. Stat. Sci. 8:305-311.

Sarika, M.A.I and C. Chattopadhyay. 2011. Modeling and forecasting of pigeonpea (*Cajanus Cajan*) production using autoregressive integrated moving average methodology. Ind. J. AgriC. Sci. 81:520-523.

Shabur, S.A. and M.E. Haque. 1993. Analysis of rice in Mymensingh town market pattern and forecasting. Bang. J. Agric. Econ. 16:130-133.

Sohail, A., A. Sarwar and M. Kamran. 1994. Forecasting total food grains in Pakistan. J. Eng. Appl. Sci. 13:140-146.

Suleman, N. and S. Sarpong. 2012. Forecasted milled rice production in Ghana using Box-Jenkins approach. Int. J. Agric. Manage. Develop. 2:79-84.