Scholars Journal of Agriculture and Veterinary Sciences
OPEN ACCESS
Abbreviated Key Title: Sch J Agric Vet Sci
ISSN 2348–8883 (Print) | ISSN 2348–1854 (Online)
Journal homepage: https://saspublishers.com

Forecasting of Potato Production in Bangladesh using ARIMA and Mixed Model Approach

Mohammad Mukhlesur Rahman[1*], Mohammad Amirul Islam[2], Md. Golam Mahboob[3], Nur Mohammad[1], Istiak Ahmed[1]

[1]Scientific Officer, Agricultural Statistics & ICT Division, Bangladesh Agricultural ResearchInstitute (BARI), 1701 Gazipur, Dhaka, Bangladesh
[2]Professor, Department of Agricultural and Applied Statistics, Bangladesh Agricultural University, 2202 Mymensingh, Bangladesh
[3]Principal Scientific Officer, Department of Natural Resource Management, Bangladesh Agricultural Research Council, 1217 Dhaka, Bangladesh

DOI: 10.36347/sjavs.2022.v09i10.001 | **Received:** 17.08.2022 | **Accepted:** 29.09.2022 | **Published:** 07.10.2022

***Corresponding author:** Mohammad Mukhlesur Rahman
Scientific Officer, Agricultural Statistics & ICT Division, Bangladesh Agricultural ResearchInstitute (BARI), 1701 Gazipur, Dhaka, Bangladesh

Abstract **Original Research Article**

A time series model is used to forecast future values by identifying patterns of historical movement of a variable. This study attempted to develop the best potato predicting model in Bangladesh using BBS provided secondary annual data on area and production of potatoes in Bangladesh from 1970–71 to 2019–20, using the most recent accessible criteria for selecting a model, such as AIC, BIC, RMSE and others. The ARIMA (0, 2, 2) model is the box-Jenkins ARIMA model with the best selection for forecasting potato output throughout Bangladesh. When considering the area of the potato, the mixed model, i.e., ARIMA (0, 2, 3), beats the univariate ARIMA (0, 2, 2) model. The mixed model's 95 percent confidence interval of the prediction value is shorter than that of the ARIMA model. As a result, the forecasting performance of the mixed model outperforms that of econometric models such as ARIMA; this might be due to the inclusion of explanatory factors such as area. The comparison of the real and predicted series demonstrates that the model used to estimate potato production in Bangladesh is statistically sound. The models forecast well at acceptable levels. As a result, depending on Bangladesh's expected potato production, these models can be used for policy objectives.

Keywords: Potato, production, area, ARIMA model, mixed model, forecasting.

INTRODUCTION

Potatoes (*Solanumtuberosum* L.), a staple crop in many regions of the globe and an essential component of most of the global food supply, are the world's fourth most vital food crop following rice, maize, and wheat (Ezekiel *et al.*, 2013; Zhang *et al.*, 2016).In 2019, world production of potatoes was 354 million tonnes, led by China with 21.31% of the total. Other major producers were India (14.15%), Russia (6.22%), Ukraine (5.71), United States (5.43%) and Germany (2.99%). Potato is considered as a potential crop to meet the challenges of the 21st century in Bangladesh and has achieved remarkable success in potato production making it ranked 7th with producing 2.72% potatoes of the total in the world. It remains an essential crop in Europe, where per capita production is still the highest in the world, but the most rapid expansion over the past few decades has occurred in southern and eastern Asia (FAOSTAT, 2019). Potato production is increasing day by day in Bangladesh due to suitable varieties as well as various production packages and regular input supply. It is the third most important crop after paddy and wheat (Kundu and Kabir, 2012). The total production and area in 2018-19 were 9.65 million metric tons and 0.47 million hectares, respectively (BBS, 2020). It constitutes the second largest crop in terms of cultivated area (after rice), corresponding to 3% of Bangladesh's total area and providing as many as 6.0% of the daily per-capita calories and protein consumed in rural areas because of the abundant supply. It is regarded as a partial substitute of rice in many households (BBS, 2017; Hossain *et al.*, 2016). Potato is produced once a year in Bangladesh. Farmers face many hardships and struggle to achieve satisfactory yields during the cultivation period. But when it comes to marketing, their production revenue is insufficient and as a result the farmer falls into liability again (Hossain *et al.*, 2016). Most of the farmers in

Bangladesh are uneducated. The distribution of market statistics needs to be increased so that farmers can get a reasonable price for potatoes.

Although Bangladesh has attained food security status in terms of quantity but the nutritional security remains questionable (Talukder *et al*., 2005). According to estimates, more than 20 million people in Bangladesh are adversely affected by chronic vitamin A, iron, and zinc deficiencies, which more strongly affect pregnant women and infants (FAO, 2015). Considering the population growth and knowing that rice is the major staple food in Bangladesh, an urgent need exists for specific examination of alternative crops that can stabilize the country's economy and change its dietary patterns with increased nutritional value. So, it is high time to grow high productivity crops, like potato to intensify the agricultural production further (Uddin *et al*., 2015). Potatoes are the world's largest non-cereal crop, after rice, wheat, and maize. Because of being high-yielding, nutrient-rich and climate-resilient nature, they can replace cereal crops (Hong *et al*., 2017). From a nutritional point of view, potatoes are one of the cheapest sources of carbohydrates and provide an admirable amount of vitamin B (Hossain *et al*., 2016). Potatoes provide 6 percent of the daily per capita calorie and protein intake in rural areas and a much higher percentage of total food intake in the winter months when they provide seasonal abundance (Scott, 1985). So, potato might be a good source of nutrients for Bangladesh's undernourished and malnourished people. Therefore, a comprehensible plan is required to make the crop popular and sustainable.

Developing agricultural technology for expanding production requires long- and short-term export and import policies and for this credible pre-harvest crop yield estimates are needed (Kumar *et al*., 2020). Several methods for estimating agricultural production are available, including ARIMA and mixed model approaches, which have been utilized for research and crop productivity prediction. Prediction can be used to support decision-making and to plan for the future in a more effective and efficient manner. The Auto Regressive Integrated Moving Average (ARIMA) model is utilized as a benchmark model in this study to compare the regression model's prediction ability. According to Biswas *et al*., (2013), the ARIMA (2,1,3) model is the best fit for the study of gross cultivated area, whereas the ARIMA (2,1,1) model is the best fit for the series of production of rice. The model has a high level of accuracy for future projections of rice acreage and production in West Bengal, India. Sarika *et al*., (2011) used this methodonpigeon pea production in India. Sugarcane acreage, production, and productivity in Tamilnadu, India, were estimated using the ARIMA model by (Suresh *et al*., 2011). The ARIMA model was used by Kumari *et al*., (2014) to estimate the rice yield in India. ARIMA was used by Naveena *et al*., (2014) to forecast the production of coconuts in India. Rathod *et al*., (2018) used the ARIMA model to predict mango and banana yields in Karnataka, India.

(Hamjah, 2014) suggests ARIMA (2,1,2) for Aus, Aman, and Boro ARIMA (1,1,3) rice production in Bangladesh. This study compared original and forecasted series. Both sets of data looked the same, proving that the fitted model accurately predicts Bangladeshi rice production. The analysis shows the ARIMA model's short-term forecasting accuracy. Hamjah (2014) also revealed that ARIMA(2,1,3), ARIMA(3,1,2), and ARIMA(1,1,2) are the best models for predicting mango, banana, and guava production in Bangladesh. Several studies in Bangladesh have used the ARIMA model to predict crop production. Hossain *et al*., (2016) predicted Bangladesh's potato production. In their research, the ARIMA (0, 2, 1) model predicted future potato production well. Hossain *et al*., (2015) suggested using ARIMA to predict Bangladesh's tea production. This study examined Bangladesh's secondary tea production from 1972 to 2013. Based on the lowest values of AIC, AICC, and BIC, ARIMA (0, 2, 1) best predicts Bangladesh's tea production. Rahman *et al*., (2013) determined the best Auto-Regressive Integrated Moving Average (ARIMA) model for chickpea, field pea, and pigeon pea pulse production in Bangladesh. The Box-Jenkins ARIMA model and deterministic growth models were used to predict Bangladesh's pulse production. ARIMA (0, 1, 0), ARIMA (1, 1, 1), and ARIMA (1, 1, 3) were the best models for chickpea, pigeon pea, and field pea pulse production, respectively. Rahman (2010) examined the best-fitted ARIMA model to forecast boro rice production in Bangladesh from 2008-2009 to 2012-13. The best ARIMA models for local, modern, and total boro rice production are (0,1,0), (0,1,3), and (0,1,2). In Bangladesh, the mixed model forecasts Aus, Aman, Boro, and Potato better than ARIMA. Boro rice production is higher in this study. Aus, Aman, and Boro rice production is expected to equal national production in 2015-2016 (AMIS, 2017).

To meet up the nutritional demand of rising population and to have a positive impact on economy by exporting potatoes, an urgent need exists to adopt appropriate policy recommendations by forecasting the production of potatoes. In this regard, the current study examines the existing trend in potato production and an effort is made to forecast production for the five leading years. The model developed for forecasting is an Autoregressive Integrated Moving Average (ARIMA) model and a dynamic regression model. After choosing the best model, forecasting is done with the model. The main reason for choosing this model in this study for forecasting production is due to the fact that this model assumes and takes into account the non-zero autocorrelation between the successive values of the time series data. The open source computer language R version 4.0.2 was used for all statistical research.

MATERIAL AND METHODS

This study considers published secondary data on annual potato production in Bangladesh which was collected from 1970-71 to 2019-20 from the annual report published by BBS.

The auto-regressive integrated moving average (ARIMA) and mixed model approaches were then used to forecast crop production at the national levels. Finally, for the most accurate forecast, integrated crop prediction techniques are required. For short-term prediction, auto-regressive integrated moving average (ARIMA) is usually better than a deterministic growth model in time series models. The best auto-regressive integrated moving average model for time series prediction was fitted using the Box-Jenkins technique.

Identification, parameter estimate, and diagnostic checking are the three processes in the prediction approach. The data were plotted to check for any unexpected observations. Possible data transformation is required to provide stable variance among time series data. If the time series is not stationary, seasonal and non-seasonal differencing must be used to ensure that the series remains stationary. When stationarity has been achieved, the autocorrelation and correlogram should be examined to determine whether any pattern remains. Previously, the sequence of differences was determined, and the differenced univariate time series could be calculated using both time-domain and frequency-domain methods (Cressie, 1988). The model's parameters must be estimated in the second stage. The parameters of the models were estimated using the maximum likelihood estimation approach. The third stage is whether the preferred model is fit for the data. The goodness of fit for the ARIMA model (Sarda and Prajneshu, 2002) was assessed using Akaike's Information Criterion (AIC) and the Schwartz-Bayesian Criterion in diagnostic testing (SBC). The true ARIMA (p, d, q) model must be adopted with competence such that the residuals predicted by this model are white noise. As a result, the residual autocorrelations must be predicted for model diagnostic checking. Under the null hypothesis that the autocorrelation co-efficient is zero, they may also be tested using the chi-square test (Ljung and Box, 1978).

The following is the study's ARIMA model:

$$\Delta^d Production_t = \alpha_0 + \sum \alpha_p Production_{t-p} + \sum \lambda_q Error_{t-q} + Error_t$$

Where Δ refers to the differencing of the time seriesThe orders of autoregressive, differencing, and moving average are p, d, and q, respectively.

Pankratz (1991) defined the mixed model method as a dynamic regression model that may be used to determine the contribution of independent factors in determining the variable of interest (response variable). A dynamic regression model (mixed model) is a regression model that allows for the inclusion of lagged values of the independent variables. When the independent variable changes, this model is used to predict what will occur to the predicted variable. The dynamic regression model is a mixed model strategy for detecting the role of independent variables in determining the variable of interest, i.e., the response variable (AMIS, 2017). The first step in selecting the best dynamic regression model is to fit a multiple regression model that looks like this:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + Z_t$$

Where, Y_t is the response variable; X_{t-k} is the independent variable with the time-lag $k = 0, 1, 2, \ldots, K$; $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients and Z_t is an ARIMA process. The backward time lags i.e., $X_{t-1}, X_{t-2}, X_{t-3}, \ldots, X_{t-k}$ are used in explaining the movement in Y_t. The parameters of mixed models are estimated using maximum likelihood method (Akpan *et al*., 2016).

If the regression errors are non-stationary, the second step requires the application of variable differencing. For the errors, the model was fitted again using the lower order ARIMA model. If the errors now perform stationarity, the third stage is to diagnose the number of lagged independent factors that will impact the prediction variable. The fourth step is to compute the errors in the regression model and choose the best ARMA model for the error series. The fifth stage involves updating the entire model using the new ARMA model for errors and the transfer function model for independent variables. The last step is to collect the residuals and see if the fitted model is enough (Makridakis *et al*., 1998). As the study used a mixed model technique to consider area, the functional form of the study's final mixed model is as follows:

$$\Delta^d Production_t = \alpha_0 + \sum \alpha_p Production_{t-p} + \beta_i Area_{t-i} + \sum \lambda_q Error_{t-q} + Error_t$$

Where, β_i is the i^th^ coefficient of the lagged independent variables; α_p is the p^th^ coefficient of lagged response variable; λ_q is the q^th^ coefficient of lagged error and α_0 is the drift.

Dynamic regression allows the errors from a regression to contain autocorrelation. To emphasise this change in perspective, ε_t will be replaced with η_t in the equation. The error series η_t is assumed to follow an ARIMA model. For example, if η_t follows an ARIMA(1,1,1) model, then

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \eta_t$$
$$(1 - \phi_1 B)(1 - B)\eta_t = (1 + \theta_1 B)\varepsilon_t$$

This model has two error terms; the error from the regression model, which we denote by η_t, and the error from the ARIMA model, which we denote by ε_t. Only the ARIMA model errors are assumed to be white noise.

When we estimate the parameters from the model, we need to minimize the sum of squared ε_t values. If we minimise the sum of squared η_t values instead (which is what would happen if we estimated the regression model ignoring the autocorrelations in the errors), then several problems arise.

1. The estimated coefficients $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ are no longer the best estimates, as some information has been ignored in the calculation;
2. Any statistical tests associated with the model (e.g., t-tests on the coefficients) will be incorrect.
3. The AICc values of the fitted models are no longer a good guide as to which is the best model for forecasting.
4. In most cases, the pp-values associated with the coefficients will be too small, and so some predictor variables will appear to be important when they are not. This is known as “spurious regression”.

Minimizing the sum of squared ε_t values avoids these problems. Alternatively, maximum likelihood estimation can be used; this will give similar estimates of the coefficients. An important consideration when estimating a regression with ARMA errors is that all of the variables in the model must first be stationary. Thus, we first have to check that y_t and all of the predictors $(X_{1,t}, X_{2,t}, \ldots, X_{k,t})$ appear to be stationary. If we estimate the model when any of these are non-stationary, the estimated coefficients will not be consistent estimates (and therefore may not be meaningful). One exception to this is the case where non-stationary variables are co-integrated. If there exists a linear combination of the non-stationary y_t and the predictors that is stationary, then the estimated coefficients will be consistent (Harris and Sollis, 2003).

The Ljung-Box test reveals that the residuals follow white noise, hence the optimal model is one that has the biggest R^2 and the lowest root mean squared error (RMSE), mean absolute percent error (MAPE), and Bayesian information criterion (BIC).

RESULTS AND DISCUSSION

A stationary time series is one whose characteristics do not depend on the time in which the series is observed. Thus, according to the time series with the trend, or with seasonality, the trend is not stationary and seasonality will affect the value of the time series at different times. A time series data is a two-dimensional plot. Potato production fluctuates over the time (Figure 1). The data series may not be stationary if it fluctuates over the time.

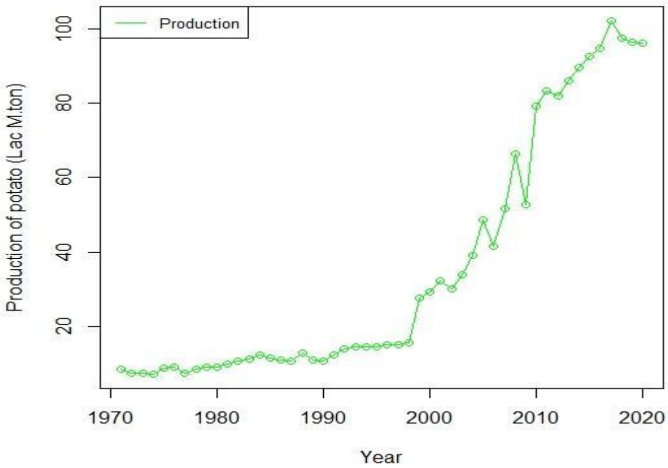

Figure 1: Potato production during the year 1970-71 to 2019-20

In this study, the stationarity of the time series of potato (in LacM.Tons) from financial year 1970-71 to 2019-20 have been checked on the basis of Augmented Dickey-Fuller and Phillips-Perron test. After second differencing the Augmented-Dickey-Fuller (ADF) test with P(|τ|≥-5.7037)<0.01 and after first difference Phillips-Perron (PP) test with P(|τ|≥-66.2)<0.01 at 5% level of significance adequately mentioned that the data series are stationary and suggest that there is no unit root. At second difference, potato productions in distinct year’s data are stationary which is presented in Table 1.

Table 1: The stationarity checking of the series of potato production (M. Ton)

Difference	Augmented Dickey-Fuller test			Phillips-Perron test		
	Statistic	P-value	Comments	Statistic	P-value	Comments
No	-1.53	0.76	Not Stationary	-3.43	0.91	Not Stationary
First	-2.06	0.55	Not Stationary	-66.2	0.01	Stationary
Second	-5.70	0.01	Stationary	-	-	-

It is clear that the potato production data series shows increasing pattern until 2017 and then decreasing till to date but the variance is not stable (Figure 1) and so we examine whether the difference of the potato production data series shows stable variance i.e., the differenced data series becomes stationary (Figure 2a). To stabilize the variance and make the data stationary, second difference is enough. The alternative positive and negative ACF (Figure 2b) and exponentially decay PACF (Figure 2c) indicate an autoregressive moving average process. The PACF with significant spike at lag 1 and lag 2 and ACF exponentially decaying towards 0 suggest that no autoregressive order and second order moving average are effective on potato production in Bangladesh.

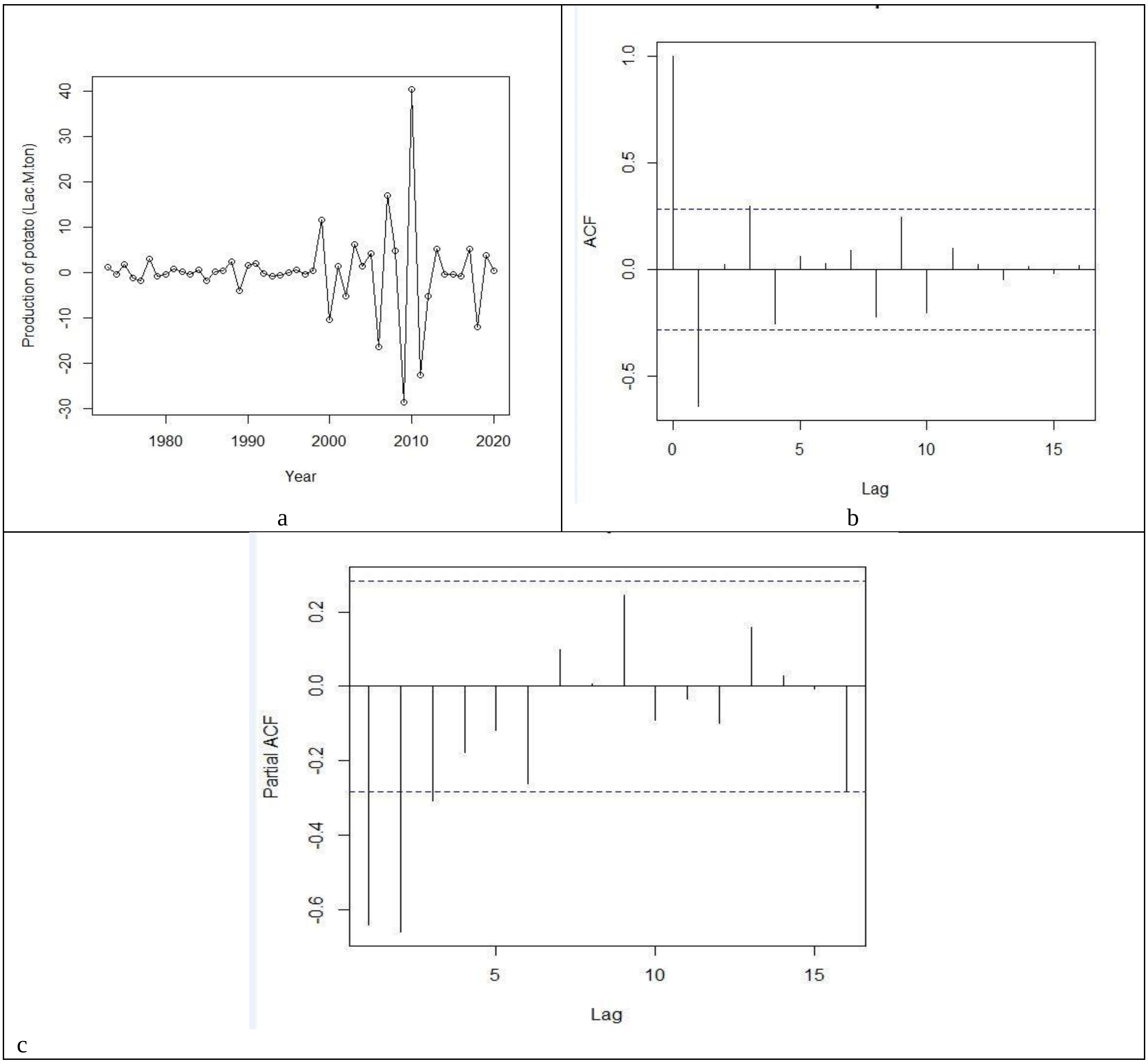

Figure 2: (a) Time series of 2nd differenced plot (b) Correlogram of ACF plot and (c) Correlogram of PACF plot at 2nd differenced potato production of Bangladesh

The above findings can be verified using trial and error methods.

Table 2: Model selection

Model	AICc	Model	AICc
ARIMA(2,2,2)	303.11	Mixed ARIMA(0,2,2)	300.00
ARIMA(0,2,0)	352.64	ARIMA(0,2,3)	302.34
ARIMA(1,2,0)	329.82	ARIMA(1,2,1)	305.16
ARIMA(0,2,1)	311.25		
ARIMA(1,2,2)	302.35		

So, it is clear that ARIMA(0,2,2) model with AIC=299.46, AICc=300.00 and BIC= 305.07 is the best selected model for forecasting potato production in Bangladesh. The estimates of the parameters of the fitted ARIMA (0,2,2) model are shown in Table 3.Note that previous studies identified ARIMA (0,1,1) (AMIS, 2017) and ARIMA (0,2,1) (Hossain *et al*., 2016) for potato production. The difference may be due to possible structural difference over time.

Table 3: Summary statistics and forecasting criteria of the fitted ARIMA (0,2,2) model

Coefficients	Estimates	SE
Moving average 1	-1.3919	0.1179
Moving average 2	0.6152	0.1162

Now checking the stationarity of all the variables are needed in case of the parameter estimation of dynamic regression model. From the previous findings, the response variable is not stationary and the regressor variable is also non stationary (Table-4).

Table 4: The stationary checking of area (Hectre)

Difference	Augmented Dickey-Fuller test			Phillips-Perron test		Comments
	Statistic	P-value	Comments	Statistic	P-value	
No	-1.73	0.68	Not Stationary	-5.51	0.91	Not Stationary
First	-3.12	0.13	Not Stationary	-54.34	0.01	Stationary
Second	-5.55	0.01	Stationary	-	-	-

Considering result of Augmented Dickey-Fuller test, the area follows ARIMA (0, 2, 3) process. Hence the mixed model would follow either ARIMA (0, 2, 2) or ARIMA (0, 2, 3).

Table 5: Model selection for dynamic regression

Model	AICc
ARIMA(0, 2,2)	266.19
Mixed ARIMA(0, 2, 3)	266.17

From Table 5, mixed ARIMA (0, 2, 3) model with AIC=264.74, AIC_C=266.17and BIC= 274.1is the best selected model for forecasting potato production in Bangladesh. The estimates of the parameters of the fitted mixed ARIMA (0, 2,3) model are shown in Table 6. A previous study (AMIS, 2017) suggested a mixed model for potato production. The difference is mainly due to the length of time series considered for analysis.

Table 6: Summary statistics and forecasting criteria of the fitted ARIMA (0, 2, 3) model

Coefficients	Estimates	SE
Moving average 1	-1.5202	0.1477
Moving average 2	0.8692	0.2566
Moving average 3	-0.2611	0.1575
Regressor (Area)	14.4134	1.9803

To select the best model between the traditional ARIMA and mixed ARIMA, we have to rely on maximum value of R^2 and the lowest value of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Bayesian Information Criteria (BIC) where Ljung-Box tests reveal that the residue follows the white noise. The ARIMA model and mixed-model for forecasting potato production based on separate model selection criteria are shown in Table 7.

Table 7: Criteria for model selection to check the best fitted model

Model	R^2	RMSE	MAPE	AIC	BIC	Ljung-Box Q	
						Statistic	P-value
ARIMA (0, 2, 2)	0.979	4.92	9.76	299.46	305.1	6.36	0.6066
Mixed-model (0, 2, 3)	0.990	3.28	6.31	264.74	274.1	4.56	0.6013

From Table 2 we noticed that ARIMA (0, 2, 2) is the best selected model for forecasting potato production. The ARIMA model depends on crop production time series data (response variable) but does not include crop area data. Alternatively, the mixed model method is a dynamic regression model that creates the role of independent variables in determining the response variable. So mixed-model depends on crop production time series data (response variable) including crop area time series data (independent variable). On the basis of different model selection criteria, the mixed-model is the best fitted model for predicting potato production. Instead of ARIMA (0, 2, 2) the mixed model is the best fit model for predicting potato production in Bangladesh. Mixed model for potato production incorporates area of the production, and ARIMA (0, 2, 2) as the explanatory variables.

From Figure 3a., it is suggesting that there is no significant pattern and hence, there is no autocorrelation among the residuals in different graphical tests of the residuals for the fitted ARIMA (0, 2, 2) model.

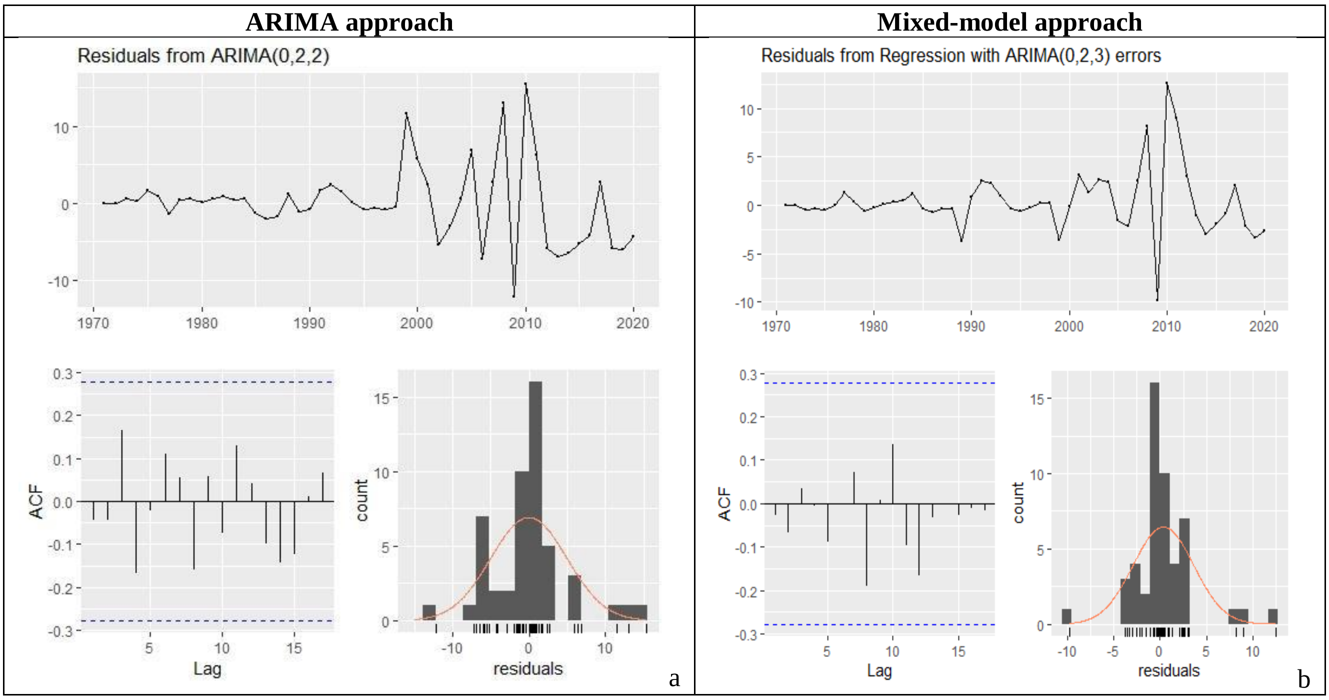

Figure 3: Several plots of residuals and Histogram with Normal Curve; a. ARIMA approach; b. Mixed-model approach

Normal Curve of Histogram is applied to test the normality statement of the residuals of the fitted model. The Histogram with Normal Curves of the residuals of the fitted ARIMA (0, 2, 2) and mixed model is given in Figure 3 (a, b). Histograms with Normal curves almost suggest that the residues of the fitted ARIMA (0, 2, 2) and mixed models are normally distributed. Although the mixed model can be used to determine the force of an independent variable in determining the response variable, it is clear that the applied mixed model is better fitted on a normality basis rather than the ARIMA (0, 2, 2) model and it forecasts potato production adequality in Bangladesh.

There are different types of potatoes available in Bangladesh which has different planting times. Hence forecasting is needed for different stages. It is important to predict at least successive stages of crop production such as planting, growth and harvesting time. At the planting stage ARIMA may be used to predict crop production. On the contrary, mixed model methods may be used to predict crop production at the growing stage. Based on the most useful predicting criteria, the predicted values of ARIMA and mixed model methods and the 95% confidence level for five fiscal years for potato production are shown in Table 8.

Table 8: Predict production of potato (in Lac M.Ton) from the year 2020-21 to 2024-25

Year	ARIMA (0, 2, 2)			Mixed-model (0,2,3)		
	Forecast	Lower limit	Upper limit	Forecast	Lower limit	Upper limit
2020-21	97.90	87.85	107.96	98.84	92.00	105.68
2021-22	97.09	85.32	108.86	99.25	91.66	106.84
2022-23	96.28	81.84	110.71	99.34	89.87	108.82
2023-24	95.46	77.55	113.38	99.44	88.08	110.80
2024-25	94.65	72.60	116.70	99.53	86.25	112.81

According to Table 7, potato production in 2020-21 will be 97.90 Lac.M.tons, and if present trends continue, potato production in Bangladesh will be 94.65 LacM.tons, according to the ARIMA model estimate. However, using land area and lag of land area production as independent variables, the mixed-model technique predicts that potato production in 2020-21 will be 98.84 LacM.tons. The length of the projected value confidence interval is less for the mixed-model approach than for the ARIMA model.

Figure 4 depicts a graphical comparison of actual production data with predicted production data. It is clear that the original potato production data for ARIMA initially indicates equal production, then shows a tiny increased trend, and then again after 1999 shows a substantially rising trend in Bangladesh potato production. The predicted potato production data differs somewhat from the original series in that it depicts production in the same way as the original data (Figure 4). Similarly, using the mixed-model approach, the original and predicted potato production data initially indicate equal production, but after some time, the forecasted potato production data shows a declining trend rather than the actual data. As a result, the confidence interval of mixed-model forecast potato production data is closer than the confidence interval of ARIMA forecast potato production data. Finally, we may infer that a mixed-model approach, rather than ARIMA, will provide more precise interval prediction for potato production (Figure 4).

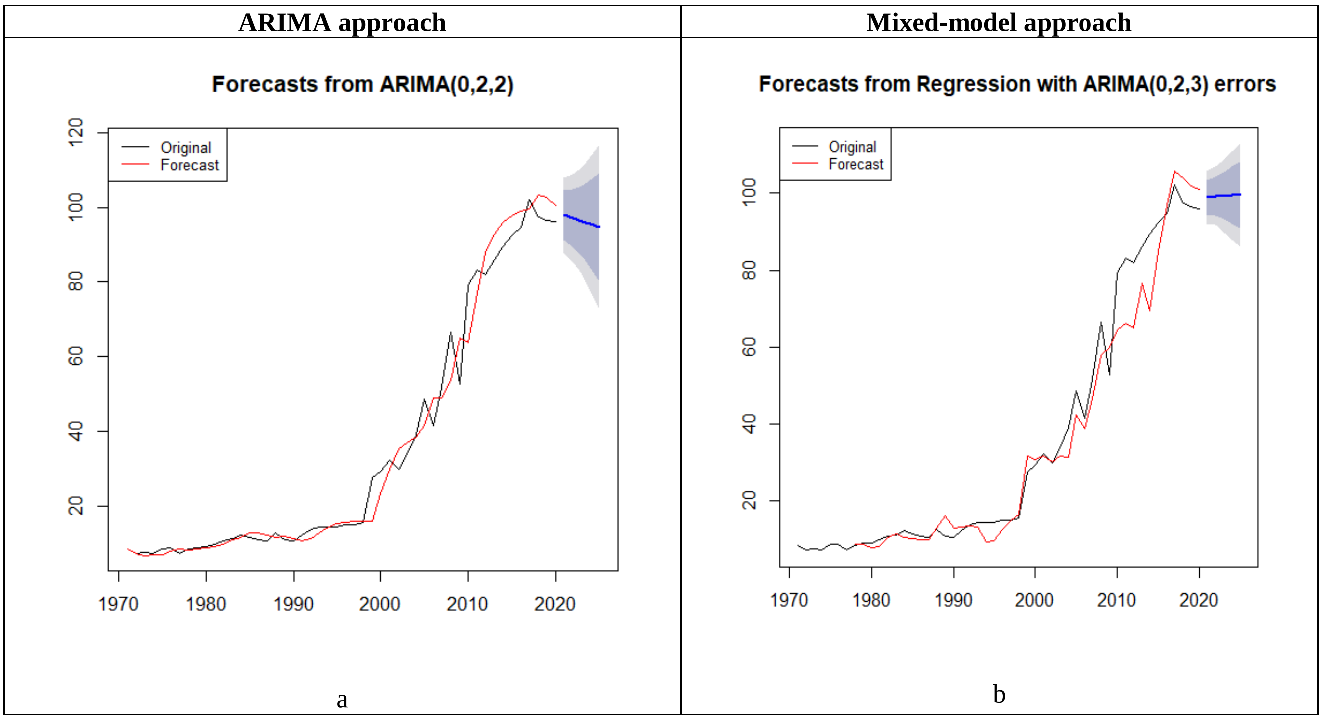

Figure 4: Comparison of original and forecast values of potato production. (a). ARIMA; (b). Mixed-model approach

CONCLUSION

A time series model detects patterns in a variable's previous movement and utilizes that knowledge to forecast future values. This study attempted to design the best model to estimate potato production in Bangladesh using the most recent accessible model selection criteria such as R^2, RMSE, MAPE, and BIC, among others. ARIMA (0, 2, 2) is the best selected Box-Jenkins ARIMA model for forecasting potato production across Bangladesh. When the area of the potato is included, the mixed-model outperforms the univariate ARIMA model.

However, generated models must be checked and remade after a few years to account for structural changes caused by unavoidable situations. According to the ARIMA model, Bangladesh's potato production will be 97.90 LacM.tons in 2020-21, and assuming the same conditions persist, Bangladesh's potato production will be 94.65 LacM.tons in 2024-25. However, using land area and lag of land area as explanatory factors, the mixed-model method predicts that maize production in 2020-21 will be 98.84 Lac M.tons, increasing to 99.53 Lac M.tons in 2024-25. The mixed-model method has a shorter confidence interval (containing both the upper and lower bounds) than the ARIMA model implying that the mixed-model outperforms other econometric models such as ARIMA in terms of predicting performance; the explanation for this might be the inclusion of exogenous variables such as area. The models projected to a good level both during and after the prediction period.

Considering the Bangladeshi context this modeling exercise may be done at multiple stages, i.e., during plantation and before harvesting. When the data on area has not been collected yet ARIMA may give an initial guideline. Afterwards, final forecasting may be made using mixed method approach. Potato being an important crop, to take appropriate marketing strategies and plan for storage facilities, thus supplying both the farmer and marketing actors forecasting practices are recommended on a regular basis. As there has been no research on forecasting potato production in other studies except AMIS (2017) using mixed model approach in Bangladesh relative comparison was not possible, which is the single limitation of this study.

REFERENCES

- Akpan, E. A., & Moffat, I. U. (2016). Dynamic Time Series Regression: A Panacea for Spurious Correlations. *International Journal of Scientific and Research Publications*, 6(10).
- AMIS. (2017). Chapter VI. Model based forecasting. *Strengthening Agriculture Market Information System (AMIS) in Bangladesh project*. Analytical report on methodologies of crop estimation and forecast and private stock of food grain survey 2016-17. Statistics and Informatics Division (SID), Bangladesh Bureau of Statistics (BBS). Ministry of Planning. 2017, 38-59.
- BBS. (2020). *Yearbook of Agricultural Statistics of Bangladesh*. Bangladesh Bureau of Statistics, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.
- BBS. (2017). *Yearbook of Agricultural Statistics of Bangladesh*. Bangladesh Bureau of Statistics, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.
- Biswas, R., & Bhattacharyya, B. (2013). ARIMA modeling to forecast area and production of rice in West Bengal. *Journal of Crop and Weed*, 9(2), 26-31.
- Cressie, N. (1988). A Graphical Procedure for Determining Non-stationary in Time Series. *JASA*. [83: 1108-15].
- Ezekiel, R., Singh, N., Sharma, S., & Kaur, A. (2013). Beneficialphytochemicals in potato — a review. *Food Research international*, 50, 487–496.
- FAOSTAT. (2019). "Potato production in 2019; Region/World/Production Quantity/Crops from pick lists". UN Food and Agriculture Organization, Statistics Division (FAOSTAT). 2019. Retrieved 24 July 2021.
- FAO (Food and Agriculture Organization of the United Nations). (2015). The State of Food and Agriculture, Rome.
- Hamjah, M. A. (2014). Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. *J Econ Sustain Dev*, 5(7).
- Hamjah, M. A. (2014). Rice production forecasting in Bangladesh: An application of Box-Jenkins ARIMA model. *Mathematical theory and modeling*, 4(4), 1-11.
- Hong, Z., Fen, X., Yu, W., Hong-hai, H., & Xiao-feng, D. (2017). Progress of potato staple food research and industry development in China. *J Integr Agric*, 16, 2924-2932.
- Harris, R., & Sollis, R. (2003). *Applied time series modelling and forecasting*. Chichester, UK: John Wiley & Sons.
- Hossain, M. M., & Abdulla, F. (2016). Forecasting potato production in Bangladesh by ARIMA model.*Journal of Advanced Statistics*, 1(4).
- Hossain, M. M., & Abdulla, F. (2015). Forecasting the Tea Production of Bangladesh: Application of ARIMA Model, *Jordan Journal of Mathematics & Statistics*, 8(3), 257-270.
- Kumar, K. S., & Shitap, M. (2020). Statistical evaluation of stepwise regression method and autoregressive integrated moving average method for forecasting of groundnut (arachishypogaea l.) productivity in Junagadh district of Gujarat. *Int J Curr Microbiol App Sci*, 9(11), 84-93. https://doi.org/10.20546/ijcmas.2020.911.009.
- Kumari, P., Mishra, G. C., Pant, A. K., Shukla, G., & Kujur, S. N. (2014). Autoregressive Integrated Moving Average (ARIMA) approach for prediction of rice (*Oryza sativa* L) yield in India. *The Bioscan*, 9(3), 1063-1066.
- Kundu, B. C., & Kabir, K. H. (2012). Potato introduction of potato varieties for processing. In "Proc. Wrksp. Potato Processing in Bangladesh". Hussain, M. M. (eds.). TCRC and AFE, Dhaka. P. 29-38.
- Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of fit in Time Series Models. *Biometrika*, 65, 297-303.
- Makridakis, S., Wheelwright, S. C., Wiley., & Sons, R. J. (1998). *A review of: "Forecasting:*

Methods and Applications". pp.642. Third edition, ISBN 0-471-53233-9." IIE TRANSACTIONS, 31(3), p. 282.

- Naveena, K., Rathod, S., Shukla, G., & Yogish, K. J. (2014). Forecasting of coconut production in India: A suitable time series model. *International Journal of Agricultural Engineering*, 7(1), 190-193.
- Pankratz, A. (1991). *Forecasting with dynamic regression models.Forecasting with Dynamic Regressions Models*, 3nd Ed., New York, John Wiley and Sons, 1991. DOI: 10.1016/0169-2070(92)90081
- Rahman, N. A. F. (2010). Forecasting of boro rice production in Bangladesh: An ARIMA approach. *J Bangladesh Agril Univ*, 8(1), 103-112.
- Rahman, N. M. F., Aziz, M. A., Rahman, M. M., & Mohammad, N. (2013). Modeling on Grass Pea and Mung Bean Pulse Production in Bangladesh Using ARIMA Model. *Journal of Agriculture and Veterinary Science*, 6(1), 20-31.
- Rathod, S., & Mishra, G. C. (2018). Statistical Models for Forecasting Mango and Banana Yield of Karnataka, India. *J Agr Sci Tech*, 20, 803-816.
- Sarda, C., & Prajneshu. (2002). Modeling and Forecasting Country's Pesticide/ Consumption Data using ARIMA Time Series Approach. *Annals of Agricultural Research*, 23(4), 719-722.
- Sarika, Iquebal, M. A., & Chattopadhyay, C. (2011). Modelling and forecasting of pigeonpea (*Cajanuscajan*) production using autoregressive integrated moving average methodology. *Ind J Agric Sci*, 81(6), 520-523.
- Scott, G. (1985). *Marketing Bangladesh's Potatoes: Present Patterns and Future Prospects*, Lima: International Potato Center.
- Suresh, K. K., & Priya, S. R. K. (2011). Forecasting sugarcane yield of Tamilnadu using ARIMA models. *Sugar Tech*, 13(1), 23-26.
- Talukder, R. K. (2005). Food security, self-sufficiency and nutrition gap in bangladesh, bangladesh development studies. *BIDS*, 31, 35–62.
- Uddin, M. A., Rahman, K. S., Rahman, M. M., Mohammad, N., & Islam, A. F. M. T. (2015). Development of union level digital databases and maps of maize growing areas at Pirgonj in Thakurgaon district. *Bangladesh J Agril Res*, 40(4), 693-702.
- Zhang, D. Q., Mu, T. H., & Sun, H. N. (2016). Domestic and abroad research progress of potato tuber-specific storage protein patatin. *Scientia Agricultura Sinica*, 49, 1746–1756.