Dhaka Univ. J. Sci. 70(1): 8-13, 2022 (January)
ISSN 1022 - 2502, E-ISSN 2408-8528
DOI: https://doi.org/10.3329/dujs.v70i1.60375

An Econometric Analysis to Forecast the Food Grain Production in Bangladesh by Using ARIMA and VAR Models

Md. Mehedi Hossain, Murshida Khanam* and Salma Akhter

Department of Statistics, University of Dhaka, Dhaka-1000, Bangladesh

(Received: 19 September 2021 ; Accepted : 6 April 2022)

Abstract

In this study, an Econometric analysis has been conducted to identify the important factors that affect the food grain productions in Bangladesh. Here, we have considered time series data for the years from 1989- 1990 to 2019-2020. Vector Autoregressive (VAR) Model and Autoregressive Integrated Moving Average (ARIMA) model have been considered in this study. Both these models have been considered to forecast the productions of food grains in Bangladesh. The forecasting performances of these two models have been compared by using RMSE, MAE, and MAPE. It has been found that the VAR model is better than the ARIMA model to forecast the food grain production. On the other hand, it has been come out from the analysis that there is no significant impact of chemical fertilizer on the food grain production, but irrigation area has significant impact on the food grain production. Among the three variables: food grain production, irrigation area and chemical fertilizer, there exists short run relationship.

Keywords: ARIMA, VAR, RMSE, MAPE, MAE, Food Grain Production.

I. Introduction

Bangladesh's Economy is growing. Some factors affect the enlargement of economy of our country. Agriculture is an important component of GDP. Agriculture is also the core sector for the economy of Bangladesh since our independence (1971). In FY 2018-19, out of total GDP, the contribution of agriculture is around 13.02 percent (NATIONAL ACCOUNTS STATISTICS)[1]. In Fiscal year 2019-20 (p), the production of Aus rice is 27.54 lac metric tons, the production of Aman rice is 140.63 lac metric tons, the production of Boro rice is 195.60 lac metric tons, the production of wheat is 10.16 lac metric tons and the production of Maize is 35.69 lac metric tons[1]. Bangladesh Bureau of Statistics (BBS) estimated that the Food Grain production was approximately 388.14 lakh metric tons in the Fiscal Year 2016-2017 [2]. The total amount of Food Grains procured internally in 2016-2017 was 13.83 lakh metric tons[2].

Since the population of Bangladesh is increasing so the prime goal of agricultural production is to provide sufficient food. In that case forecasting about food grain production will help to a great extent for policy formulation. To do this, various time series models can be used, such as, ARIMA model, Exponential Smoothing, VAR model, Neural Network etc.

In the present study, we have used the popular ARIMA and VAR models. Because the main focus of this paper is to make an Econometric analysis to forecast the food grain production in Bangladesh by using ARIMA and VAR models. In this research, we have considered three major food grains which are rice (aus, amon and boro), wheat and maize.

There are so many literatures have been found which are related with the present research.

Sultana and Khanam[3] conducted a study to forecast the rice production of Bangladesh. The data of yearly rice production from 1972 to 2013 has been used. They found that the ARIMA model is superior to ANN model to forecast the rice production of Bangladesh.

Parvin and Khanam[4] carried out a study for forecasting the price of jute goods in Bangladesh. They used the data from the year 1980-1981 to 2013-2014 from Bangladesh Jute Mills Corporation. It has been come out that the ARIMA model is more competent than VAR model to forecast the price of jute goods in Bangladesh.

Hossain et al. [5] compared ARIMA and ANN model regarding estimating accuracy by using jute production in Bangladesh. They found that the ANN model is superior to the ARIMA model for forecasting the jute production in Bangladesh[5].

In 2017, a research has been carried out to forecast the food grain production in Odisha. In this study, ARIMA model has been used. It has been investigated that the best selected ARIMA model is ARIMA (2,1,0) for Kharif food grain production and ARIMA (1,1,0) for rabi food grain production[6].

The accuracy of two models ARIMA and VAR has been compared in forecasting milk production in India. The data has been collected from FAOSTAT and NDDB. The data collection period is 1961 to 2012 and 1991 to 2012 respectively for FAOSTAT and for NDDB. The result found that ARIMA (1,1,1) is the more suitable model for forecasting the milk production[7].

A study has been accomplished to investigate the forecasting performance of the water melon production in Bangladesh. The supply model, log-linear model, ARIMA model and

*Author for correspondence. e-mail:murshida@du.ac.bd

MARMA model have been used in this study. The findings showed that the ARIMA model is the finest model regarding the forecasting behavior of the water melon production in Bangladesh[8].

Objectives of the study

The important objectives of this research have been mentioned in the followings:

a. To identify the factors affecting food grains productions in Bangladesh.
b. To model annual food grain productions using VAR model to check short run relationship among the variables.
c. To fit suitable ARIMA model for food grains production in Bangladesh.
d. To forecast the future food grain productions using VAR and ARIMA model and also to compare forecasting performances of both the models.
e. To put some recommendations for the future food grain productions in Bangladesh based on the findings of the study.

II. Methodology

In this study, ARIMA and VAR models, two extensively used models for time series are considered.

ARIMA Model

ARIMA model is known as the Box-Jenkins (BJ) methodology. In this model, Y_t has been explained by using the past values of Y itself and random disturbance term[9].

VAR Model

VAR model has been used to internment the linear interdependencies among various time series[10]. Sims (1980) developed this VAR model. The univariate autoregressive (AR) model is generalized by the VAR model by permitting for more than one progressing variable. All variables are introduced into the model in the same manner in a VAR model. Each variable has an equation with the explanation of its progression based on its variables. The evolution of a set of k variables has been explained by the VAR model, and these k variables are known as endogenous variables. The endogenous variables are decorated as a linear function of only their preceding values over the sample period. The variables are collected in a (k×1) vector y_t, where y_t is the observation at time t.

Evaluation of Forecasting Model

The accuracy of the models has been compared by using three statistics regarding the forecasting approach of the models.

The Root Mean Square Error (RMSE) can be defined as-

$$RMSE = \sqrt{E\left(\left(f_t - Y_t\right)^2\right)}$$

The Mean Absolute Error (MAE) used to measure how closely the prediction goes to the outcome. It is given as-

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|f_t - Y_t\right|$$

The goodness of fit of a model has been evaluated by the Mean Absolute Percentage Error (MAPE). It can be defined as-

$$MAPE = \frac{\frac{1}{n}\sum_{t=1}^{n}\left|f_t - Y_t\right|}{Y_t} * 100$$

where, Y_t = actual observation and f_t = fitted value[3].

III. Data and Variables

Data has been obtained from Bangladesh Economic Review (BER, June 1989-June 2020)[11]. Data for the years June 1989-1990 to 2019-2020 is used for the analysis. According to the availability of the data, three variables are considered here: they are- food grain production, area under irrigation and use of chemical fertilizer. The dependent variable of the present study is Food grain production.

IV. Results

In any kinds of econometric research, the following criteria need to be checked as well as they need to be satisfied. These are: normality, multicollinearity, heteroscedasticity and autocorrelation.Through the widely used Jarque-Bera test statistic, the normality assumption of residual has been checked[12]. It has been investigated that the p value of the Jarque-Bera test statistic is 0.8444 which is greater than 0.05. Therefore, we can accept the null hypothesis (Residuals follow normal distribution.). Accordingly, it can be said that the residuals follow normal distribution for the required logarithmic transformed data sets. Here ln(production) of the food grains is the dependent variable, ln(Irrigation Area) and ln(Chemical fertilizers) are independent variables.

It has been found that the variance inflation factor (VIF) for the transformed variables are less than 10, so it can be said that the multicollinearity is not present in the data set[9].

To detect the heteroscedasticity, Breusch-Pagan-Godfrey (BPG) test has been performed in this study. Because this is expansively used test to check the heteroscedasticity. It has been found that the p-value of the BPG test is 0.3616 (logarihmic transformed data sets) which is greater than 0.05. So, we can accept the null hypothesis (Variance of the residual is homoscedastic). Accordingly, we can conclude that the data of the present study is homoscedastic[9].

To test the autocorrelation we used the Durbin Watson test. It has been found that the p-value of the Durbin Watson test is 0.0000 which is less than 0.05. We cannot accept H_0 (There is no autocorrelation). So, it can be said that the residuals are autocorrelated. After taking the first difference, the Durbin Watson test has been performed. It has been found that the p-value is 0.0502 and it is greater than 0.05. That means, we can accept H_0. So we can infer that the residuals are not autocorrelated[9].

The Augmented Dickey Fuller (ADF) test has been used to check whether the data set is stationary or not [13].

The following hypothesis to be tested-
H_0: The data is non-stationary
H_1: The data is stationary.

Table 1. Results of Augmented Dickey-Fuller (ADF) unit root test.

Series	Difference	ADF test statistic value	p-value	Conclusion
ln(Production)	First difference	-5.107	0.0000	Stationary
ln(Irrigation area)	First difference	-4.321	0.0004	Stationary
ln(chemical fertilizer)	First difference	-6.453	0.0000	Stationary

Table 1 reveals evidently that after taking first difference all the variables ln(Production), ln(Irrigation area), ln(chemical fertilizer) become stationary at 5% level of significance.

In order to select the number of lags, AkaikeInformation Criterion (AIC), Hannan-Quinn Information Criterion (HQ) have been calculated[14]. To select the number of lags in a model, these criteria are generally used.

Table 2. Lag Selection by using different criterion.

Lag	**Akaike Information Criterion(AIC)**	**Hannan-Quinn Information Criterion(HQ)**	**LR**
0	-8.29299	-8.25119	NA
1	-8.84701*	-8.6798*	32.405*
2	-8.72247	-8.42985	14.762
3	-8.64164	-8.22362	15.898
4	-8.59988	-8.05646	16.914

From the Table 2, it can be said that the value of AIC and HQ are minimum at lag length 1. In LR test the maximum value is 32.405 at lag 1. So, our optimal lag length is 1 for approaching the analysis.

To investigate co-integration, the Johansen tests[12] for co-integration has been conducted. The findings of the test suggested that there is no co-integration in the data. That means, there exists short run relationship among the variables. Consequently, we have applied the VAR model.

To apply VAR model three equations have been considered.

The equation 1 can be written by taking Δln (production$_t$) as dependent variable.

$$\Delta\ln(production_t)$$
$$=\alpha_1+\sum_{i=1}^{k}\beta_{1i}\Delta\ln\left(irrigationarea_{t-i}\right)$$
$$+\sum_{i=1}^{k}\gamma_{1i}\Delta\ln\left(c.fertilizer_{t-i}\right)$$
$$=+\sum_{i=1}^{k}\delta_{1i}\Delta\ln\left(production_{t-i}\right)+\varepsilon_{1t}$$

The equation 2 can be written by taking Δln (irrigationarea$_t$) as dependent variable.

$$\Delta\ln(irrigationarea_t) = \alpha_2 + \sum_{i=1}^{k}\beta_{2i}\Delta\ln\left(irrigationarea_{t-i}\right) + \sum_{i=1}^{k}\gamma_{2i}\Delta\ln\left(c.fertilizer_{t-i}\right) = + \sum_{i=1}^{k}\delta_{2i}\Delta\ln\left(production_{t-i}\right) + \varepsilon_{2t}$$

The equation 3 can be written by taking Δlnc. fertilizer$_t$ as dependent variable.

$$\Delta\ln(c.fertilizer_t) = \alpha_3 + \sum_{i=1}^{k}\beta_{3i}\Delta\ln\left(irrigationarea_{t-i}\right) + \sum_{i=1}^{k}\gamma_{3i}\Delta\ln\left(c.fertilizer_{t-i}\right) = + \sum_{i=1}^{k}\delta_{3i}\Delta\ln\left(production_{t-i}\right) + \varepsilon_{3t}$$

Here k is the value of the lagged period, ε_{it} is the error of the required respective VAR model equation. Here, Δ is the difference operator, which indicates stationarity of the series.

Table 3. VAR model output of the lagged values of coefficients of ln(irrigation area), ln(production) and ln(c.fertilizer).

Dependent variable	Coefficients	Values	p-values	Standard error
Δln(production$_t$)	Constant	0-.6782842	0.112	0.4267444
	Δln(production(−1))	0.9733783	0.000*	0.2154705
	Δln(irrigationarea(−1))	0.7771144	0.002*	0.2567102
	Δln(c. fertilizer(−1))	0-.1182494	0.291	0.1119821
Δln(irrigationarea$_t$)	Constant	0.2517115	0.547	0.4176538
	Δln(production(−1))	0-.218587	0.300	0.2108805
	Δln(irrigationarea(−1))	0.7035	0.005*	0.2512418
	Δln(c. fertilizer(−1))	0.0312269	0.776	0.1095967
Δln(c.fertilizer$_t$)	Constant	0.2340502	0.758	0.7585222
	Δln(production(−1))	0-.8027446	0.036*	0.3829908
	Δln(irrigationarea(−1))	0.6080546	0.183	0.4562928
	Δln(c. fertilizer(−1))	0.2075193	0.297	0.1990441

* means the value is significant at 5% level of significance.

The output of the VAR model is presented in Table 3. The findings represented lagged values of coefficients of ln(irrigation area) and the past values of the ln(production) have significant impact on current year production. Also there exists significant impact of past values of ln (irrigation area) on current year irrigation area. Production has significant impact on chemical fertilizer at 5% level of significance. The study did not find significant impact of chemical fertilizer on production or irrigation area.

To perform ARIMA model, at first we need to identify the numbers of AR and MA terms. The numbers can be easily identified from the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the differenced series.

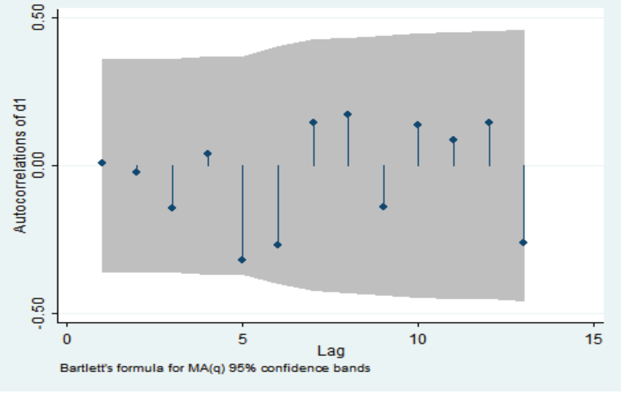

Fig. 1. ACF plot of differenced series to identify the numbers of AR and MA terms in ARIMA model.

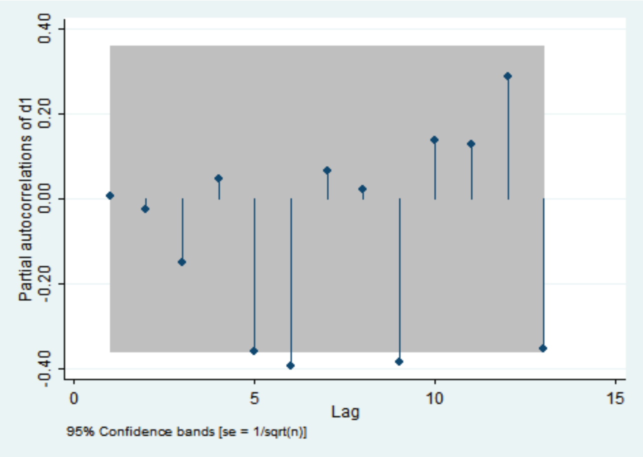

Fig. 2. PACF plot of differenced series to identify the numbers of AR and MA terms in ARIMA model.

The lowest AIC value has been used to select the best possible model among ARIMA(5,1,0), ARIMA(6,1,0) and ARIMA(9,1,0). If the model has the lowest AIC value, the chosen model is the most excellent model. The model with the lowest AIC value is the ARIMA (5,1,0), which is the appropriate model for the present study.

In the present study VAR and ARIMA models have been fitted. Then we have performed diagnostic performance. After satisfying all the diagnostic checks we have used both the models for forecasting future values of food grain production in Bangladesh.

Table 4. Forecasted Values of Food Grain Productions for both VAR and ARIMA models.

Year	VAR	ARIMA
2020-2021	457.6554	462.4348
2021-2022	470.1585	476.6216
2022-2023	491.1714	480.4437
2023-2024	510.6223	491.0715
2024-2025	521.0365	492.3147
2025-2026	536.3514	496.6917
2026-2027	559.8364	497.3020
2027-2028	578.6115	500.2195
2028-2029	596.6115	500.1262
2029-2030	619.0839	501.4302
2030-2031	644.6986	501.4121
2031-2032	667.2803	502.2161

[Here, productions are in lakh metric ton]

From Table 4. we obtained forecasted values for the next 12 years for food grain production in Bangladesh. Both the model resulted that the food grain production is increasing over the years.

To check the accuracy of the forecasted food grain productions using VAR and ARIMA models, we have estimated the three indices: i) RMSE, MAE and MAPE. The lower the value of the indices the better the estimate is. It has been found that the results are similar for these three indices regarding VAR and ARIMA models.

Table 5. Estimation and Comparisons of Forecasting Performances.

Models	RMSE	MAE	MAPE
VAR	9.573699	7.704773	1.925743
ARIMA	10.85535	9.9292	2.522392

From Table 5, we can say that all the three indices of forecasting evaluation technique are small for VAR compared to ARIMA model. So, VAR model is regarded as the best forecasting model to forecast the future food grain productions in Bangladesh comparing with the ARIMA model.

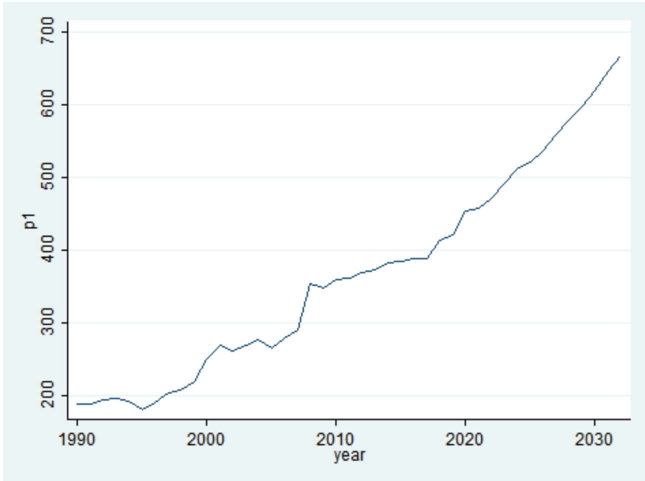

Fig. 3. Forecast Graph of Food Grains Production using VAR model.

Now from the above figure, we can easily deduce that, the food grain productions in Bangladesh has a rising trend. We forecast the values for next 12 years and it is up to the year 2032. There is a little amount of fluctuations. But overall the productions of food grains have an increasing trend.

V. Conclusion

In this study, we have found the best forecasting model through VAR and ARIMA model to forecast the productions of food grains in Bangladesh. We compare the forecasting performances of these two models by using RMSE, MAE and MAPE of prediction. It has been come out that the VAR model is the best forecasting model for food grain production in Bangladesh. Because the values of RMSE, MAE and MAPE are the lowest for the VAR model comparing with the ARIMAmodel.

From the Johansen test for co-integration, it has been found that there exists no co-integrating equation. That means, there exists short run relationship among the variables: food grain production, area under irrigation and use of chemical fertilizer.

It has been found that the chemical fertilizer has no significant impact on food grain production. This result is little bit unusual. This may be due to the fact that our sample represents such type of results.

On the other hand, the irrigation area has significant impact on the productions of food grains. Generally, it is expected that increasing of irrigation area also increases the production. So, in the context of general sense, this result is very much logical. The past values of the production have significant impact on the present value of the food grain production. Because the present situation sometimes dependent on the past situation. Even then the past value of the irrigation area also has significant impact on the present value of the irrigation area. This result is also logical as the present situation has some similar pattern with past situation.

From the forecasted values, it may be expected that the food grain production will be increased for the next 12 years.The findings regarding the best forecasting model contradicts with most of the findings of the literatures which have been mentioned in this paper. Because in most of the literatures, findings suggested that the ARIMA model is the best model. But from this study, it has been come out from the analysis that the VAR model is the better choice for the forecasting purposes comparing with the ARIMA model regarding food grain production in Bangladesh.

Further Scope of the Study

In the current study, we have considered merely two influencing factors for food grains productions such as – chemical fertilizers and irrigation area. Also there are some other factors that can also influence the food grain production such as-seed productions of the food grain, total cropped area, types of soil, rainfall, temperature, humidity etc. So, further research can be conducted by considering those factors to make more reliable result for policy making. Even then the present study used ARIMA model and VAR models. So, further study can be accompanied for forecasting food grain production in Bangladesh by using Neural Network model, Exponential smoothing technique etc.

References

1. National Accounts Statistics, Provisional Estimates of GDP, 2019-20 and Final Estimates of GDP, 2018-19., https://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd
2. Bangladesh Economic Review, 2017.
3. Sultana A. and M., Khanam, 2020. Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models. *Dhaka University Journal of Science,* **68(2),** 143-147.
4. Parvin, S. and M., Khanam, 2018. Comparison between ARIMA and VAR model regarding the Forecasting of the price of Jute Goods in Bangladesh. *Dhaka University Journal of Science,* **66(2),** 91-94.
5. Hossain, M.M., A. Faruq, and Z., Hossain, 2017. Comparison of ARIMA and Neural Networks Model to Forecast the Jute Production in Bangladesh. *Jahangirnagar University Journal of Science*, **40(1),** 11-18.
6. Dash, A., D. S., Dhakre, D.S. and D., Bhattacharya, 2017. Forecasting of food grain production in Odisha by fitting ARIMA model. *Journal of Pharmacognosy and Phytochemistry.* **6(6),** 1126-1132.
7. Deshmukh, S. S. and R., Paramasivam, 2016. Forecasting of milk production in India with ARIMA and VAR time series models. *Asian J. Dairy & Food Res,* **35(1)***,* 17-22.
8. Khanam, M. and U., Hafsa, 2013. Market Model Analysis and Forecasting Behavior of Watermelon Production in Bangladesh. *Bangladesh Journal of Scientific Research,* **26(1&2)***,* 47-56.
9. Gujarati, D. N. 2003. Basic Econometrics, 4th Edition New Work: McGraw Hill Book Co.
10. Sims, C. A., 1980. Macroeconomics and Reality. *Econometrica,***48***,* 1-48.
11. Bangladesh Economic Review (BER, June 1989-June 2020), Economic Advisor’s wing, Finance Division, Ministry of Finance, Government of People’s Republic of Bangladesh.
12. Carlos, M. J. and A. K., Bera,1987. A test for normality of observations and regression residuals. *International Statistical Review,* **55 (2),** 163–172.
13. Dickey, D. A., and W.A., Fuller, 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association, 74,* 427- 431.
14. Johansen, S., 1991. Estimation and Hypothesis Testing of Cointegration Vectorsin Gaussian Vector Autoregressive Models. *Econometrica,* **59 (6)***,* 1551–1580. JSTOR 2938278.