EBAUB Journal Volume 2 January 2020

ISSN: 2617 - 8338

EBAUB Journal

An Academic Journal of
EXIM Bank Agricultural University Bangladesh

Forecasting Mango Production in Bangladesh: An Econometric Approach

[1]Mousumi Saha*, [2]Md. Israfil Haque, [3]Md. Sazzadur Rahman Sarker

[1]Department of Statistics, EXIM Bank Agricultural University Bangladesh
[2]Faculty of Agricultural Economics and Rural Development, EXIM Bank Agricultural University Bangladesh
[3]Department of Agricultural Statistics, Bangladesh Agricultural University

ARTICLE INFO

Received date: August 07, 2019
Accepted date: Nov. 22, 2019

ABSTRACT

A timely and reliable system of forecasting mango production well in advance is of prime importance to farmers and other people who are dependent on horticultural sector. This study was undertaken to estimate growth trend and to forecast production of mango in Bangladesh. For this purpose, secondary data was collected of mango production for the period 1961 to 2017 from the Food and Agriculture Organization (FAO). The study found that the year 1990 is the breakdown year. Hence, the data was taken from 1991 to 2017. The result depicts that the growth-rate of mango production follows exponential trend and production of mango grew at a rate of 9.7% per year for the period, 1991 - 2017. The ARIMA model was used to forecast production of mango. The study reveals that among various models, ARIMA (0,1,2) model has highest value of R^2 and lowest value of root mean square error (RMSE), mean absolute percent error (MAPE) and appeared to be the best model. The predicted mango production for the year of 2025 worked out as 2680711 tons, which means that an increased output of mango would be available for consumption as well as for export. This paper underlines the need for taking measures to increase export of mango by improving its quality, packaging and complying with international standards.

Key words: Mango production, Growth trend, Forecasting, ARIMA model

CORRESPONDENCE
* mousumisaha559@gmail.com
Department of Statistics, EXIM Bank Agricultural University Bangladesh, Chapainawabganj 6300

1. INTRODUCTION

Bangladesh ranks first in terms of mango-grown area with an area of 37,830 hectares and third in terms of production with an annual production of 11,61,685 metric ton (MT) (BBS, 2016). It has a unique position among the fruits grown in Bangladesh in respect of nutritional quality, taste, consumer's preference etc. Mango grows almost all over the country but commercial and good quality mangoes grow in the Northwestern districts of the country. The leading mango growing districts of the country are Rajshahi, Chapainawabganj and greater Dinajpur. More than 500 varieties of sweet edible mangoes are found in Rajshahi and Chapainawabganj district. The Khirshapat mangoes from Chapainawabganj, popularly known as Himsagar, have received the Geographical Indication (GI) as the third product from Bangladesh, following Ilish fish and Jamdani sari. Khirshapat mangoes are one of the most popular and tastiest varieties of mangoes that are produced in Bangladesh, making up about 20-25% of the total mango

To Cite: Saha, M., Haque, M.I. & Sarker, M.S.R. (2020). Forecasting Mango Production in Bangladesh: An Econometric Approach. *EBAUB J.*, 2, 65-70.

production in the country each season. It is estimated that around 85% people of the mentioned districts are directly or indirectly dependent on mango cultivation and business (Dhaka Tribune, 2019). Therefore, the first objective of this research is to estimate growth trend in the production of mango.

Mango production provides more income to the farmers than any other crops. Because of hardy nature of the tree and low maintenance cost, mango cultivation has come up reasonably well even under dry land cultivation in Bangladesh. Receiving the GI certification means, it is known in the global market. It also inspires the farmers to boost production as well as strengthen the mango-related economy of the country. In this context, a study of production performance of mango in Bangladesh is quite essential. Therefore, the ultimate objective of our study is to study production potentiality of mango through forecasting.

In literature, a bulk of studies (e.g., Arivarasi & Ganesan, 2015; Badar et al., 2015; Khayati, 2015; Hamjah, 2014; Sabir & Tahir, 2012; Awal & Siddique, 2011; Badmus & Ariyo, 2011; Suleman & Sarpong, 2011) has been done to forecast of various crops such as wheat, rice, sugarcane, maize, cotton, and vegetables. Arivarasi & Ganesan (2015) forecasted vegetable trends and production of two vegetable growing zones in the Chennai state of India using ARIMA model. Badar et al. (2015) used ARIMA model to forecast area, production and yield of major food crops such as wheat, rice, and maize in Pakistan. Hamjah (2014) has used Box-Jenkins ARIMA model to forecast different types of seasonal rice production in Bangladesh. The study of Awal and Siddique (2011) was carried out to estimate growth pattern and also examine the best ARIMA model to efficiently forecast Aus, Aman and Boro rice production in Bangladesh. Rathod and Mishra (2017) carried out weather-based modeling for forecasting area and production of mango in Karnataka. These plenty of works have provided a fruitful contribution to literature but gap still prevails to the forecasting of the production of mango in Bangladesh. Studying the relevant works for forecasting different types of agricultural products by using Box-Jenkins (1970) ARIMA model, idea has been generated about forecasting mango production in Bangladesh. In this research, mango production has been forecasted for the next 5 years to provide deep insight to bridge such research gap. In fact, forecast is needed by national governments for the establishment of various policy decisions related to storage, distribution, pricing, marketing, import-export etc.

The paper has been organized in the given manner. The methodology section defines the data sources and techniques used in the study. After that, empirical results are given. Finally, the paper deals with conclusions and policy recommendations.

2. MATERIALS AND METHODS

2.1 Data Sources

The time series data of mango production was collected from the Food and Agriculture Organization (FAO) during the period 1961-2017.

In order to fulfill the objectives of the study, the successive methodologies have been used:

2.2 Time Series Plot

Time series plot is a visual way to check the trend of time series data. Such a plot gives an initial clue about the likely nature of the time series under study. The plotted figure may show an upward or a downward trend that means the mean of the time series has been changing. Therefore, a visual plot of the data is usually the first step in the analysis of any time series.

2.3 Growth Trend

The growth-rate of mango production follows exponential trend by pursuing curve fitting regression. The exponential growth model is defined as

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \quad (1)$$

Where, Y_t is the production of mango in year t, β_2 is the exponential growth rate (r) and u_t is disturbance term.

Given the type of data, nature of research and reliability of forecast, ARIMA model has been selected from amongst available time series models for forecasting production of mango.

2.4 ARIMA Model

A non-seasonal ARIMA model is denoted by ARIMA (p,d,q) (Box and Jenkins, 1970). Where, p is the order of the auto regressive process, d is the order of integration, ie., the number of differences to make the series stationary, q is the order of the moving average process.

The general form of ARIMA is:

$$\Delta^d Y_t = \theta + \alpha_1 \Delta^d Y_{t-1} + \alpha_2 \Delta^d Y_{t-2} + \dots + \alpha_p \Delta^d Y_{t-p} - \left(\varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \cdots + \varphi_q u_{t-q}\right) + u_t \quad (2)$$

Where, integrated of order 'd' series denoted by I(d) is:

$$\Delta^d Y_t = \Delta^{d-1} Y_t - \Delta^{d-1} Y_{t-1} \quad (3)$$

The auto-regressive model of order 'p' denoted by AR (p) is:

$$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + u_t \quad (4)$$

Where, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ are past series values (lags), the α is the coefficient to be estimated by auto-regressive model and u_t is a random variable with zero mean and constant variance (i.e., it is white noise).

The moving average model of order 'q' or MA (q) which can be written as:

$$Y_t = u_t + \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \cdots + \varphi_q u_{t-q} \quad (5)$$

Where, φ is coefficient in the moving average (MA) model. A moving average (MA) model is simply a linear combination of white noise error terms.

This ARIMA model (2) has been employed for analyzing quantitative relationship of data and for forecasting future trend mango production for the period, 2018 to 2025. Correlogram test and Augmented Dickey – Fuller (ADF) are to be applied to confirm reliability and fitness of the selected model.

At first the time series is differenced at'd' times to make it stationary. Graphically stationarity is to be tested by using correlogram. Correlogram of a time series is a graph of autocorrelation at various lags. The autocorrelation function (ACF) at lag k, denoted by ρ_k, is defined as

$$\rho_k = \frac{\sum_{t=k+1}^{n}(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})}{\sum_{t=1}^{n}(Y_t-\bar{Y})^2} \quad (6)$$

Where, $\bar{Y}$ is the sample mean of Y. If ρ_1 is nonzero, it means that the series is serially correlated in first order. If ρ_k drops to zero after a small number of lags, it is a sign that the series obeys a low-order moving-average (MA) process. A rule of thumb is to compute ACF up to one-third or one-quarter the length of the tie series (Gujarati, 2003). Since we have 27 observations, by this rule lags of 7 to 9 quarters will do. A formal application of Augmented Dickey-Fuller (ADF) is to be used for analysis of stationarity. The ADF test equation is:

$$\Delta^d Y_t = \beta_0 + \lambda t + \phi \Delta^d Y_{t-1} + \sum_{i=1}^{p} \alpha_i \, \Delta^d Y_{t-i} \quad (7)$$

The null and alternative hypothesis of the test is:

$H_0: \phi = 0$ (The time series is non-stationary)

$H_0: \phi \neq 0$ (The time series is generated by a stationary process)

The test statistic is

$$t = \frac{(\hat{\phi})}{S.E.(\hat{\phi})}$$

Dickey and Fuller obtained critical values based on Monte Carlo simulations. If the calculated t-value is lower than the critical value, the null hypothesis that the series is non-stationary should be rejected. Acceptance of the null hypothesis means that the time series at least one unit root. The test is therefore repeated on the differenced series of Y_t until the null hypothesis of non-stationarity is rejected.

Then ARMA (p,q) is applied. The best ARIMA model is established on the basis of maximum value of R^2 and minimum value of root mean squared error (RMSE), mean absolute percent error (MAPE) where Ljung-Box *Q*-statistic test reveals that whether the residual follows white noise.

The *Q*-statistic at lag k is a test statistic for the null hypothesis that there is no autocorrelation up to order k and is computed as:

$$Q_{LB} = T\,(T + 2) \sum_{j=1}^{k} \frac{\tau_j}{T-j}$$

Where, τ_j is the *j*-th autocorrelation and T is the number of observations.

3. RESULTS AND DISCUSSION

3.1. Time Series Plot

To visualize the basic pattern of the data, the time series of mango production has been represented by a graph, where the observations are plotted against corresponding time. The variable 'mango production' has been log-transformed to avoid the violation of normality assumption. By referring Figure 1, it is visually evident that the growth trend of mango production is declining before the year 1990 and after that period it is gradually increasing. Accordingly, the year 1990 is the break-down year. The upward trend suggests that the mean of mango production have been changing since 1991. Hence, to estimate the growth trend, data of mango production has been taken from 1991 to 2017. Therefore, before beginning the formal analysis, a time series plot gives an initial idea of the data series.

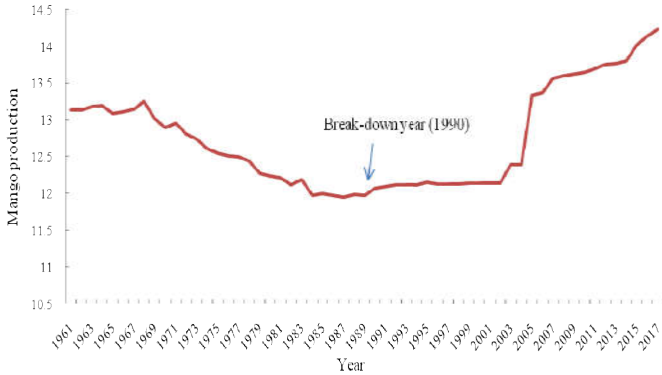

Fig 1 Time series plot of mango production.

3.2. Growth Trend

Growth trend has been estimated by employing exponential model.

Table 1 Estimated values of exponential model.

Items	Coefficient	Std. Error	t-Statistic	P-value	R^2
Constant	11.572	0.118	97.880	0.000	0.873
Trend	0.097	0.007	13.088	0.000	

The results (Table 1) illustrate that production of mango grew at a rate of 9.7% per year for the period, 1991 - 2017. The Standard Error (SE) of the slope coefficient was 0.007**,** which is very low and confirms reliability of results. Estimated coefficients are found significant at one percent level of significance. The calculated value of R^2 was 0.873 which shows that 87.3% regressand (Y_t) is explained by the regressor (t), thereby confirming reliability of the estimated model.

3.3. Forecasting Mango Production

The accuracy of forecasting of mango production depends on the time series data, which must be stationary. To find out whether the time series is stationary or not, the correlograms upto 9 lags have been shown in Figure 2 and Figure 3. The solid vertical line in the diagram represents the zero axis; observations above the line are positive values and those below the line are negative values. The area between two dashed lines denotes 95% confidence interval. For the correlogram of a stationary time series, the coefficients of autocorrelation (ACF) and partial autocorrelation (PACF) at various lags hover around zero. From Figure 2, two facts is found out: First, the autocorrelation coefficient starts at a very high value at lag 1 (0.860) and declines very slowly. ACF up to 5 lags are statistically significant, as they all are outside 95% confidence bounds. Second, after the first lag, the PACF drops dramatically and all PACFs after lag 1 are statistically significant. Hence it is a non-stationary time series and it must be first differenced to see if the first differenced time series is stationary or not.

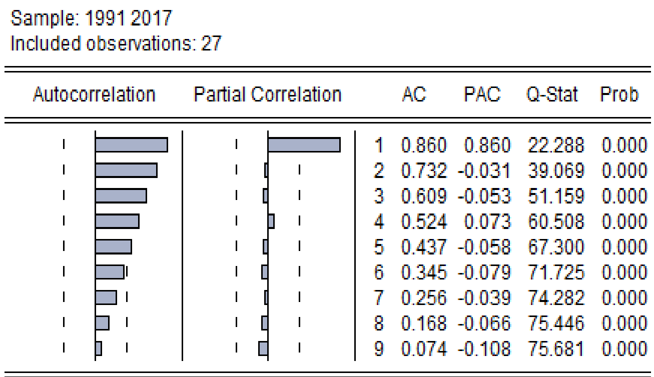

Fig 2 Correlogram and partial correlogram of mango production.

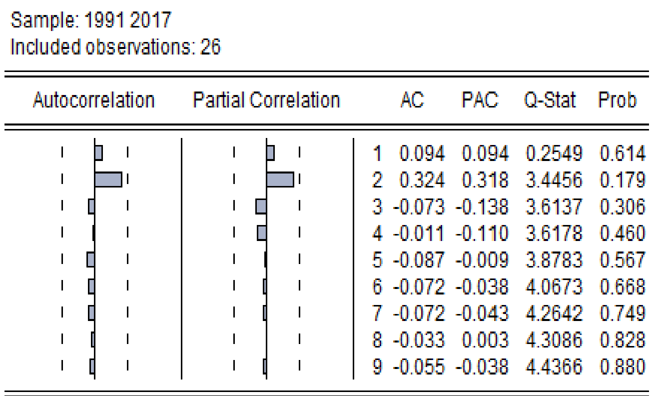

Fig 3 Correlogram and partial correlogram of first differences of mango production.

From Figure 3, there is a much different pattern of ACF and PACF. All the spikes of ACF and PACF for the time series are inside the 95% confidence bounds. These are the pictures of correlogrm of stationary time series. The result indicates that the time series of mango production is stationary in its first difference.

Table 2 Augmented Dickey-Fuller (ADF) Test Result.

Item	Difference	Augmented Dickey-Fuller Test		Comment
		t-statistic	p-value	
Mango production (Y_t)	Level	-1.076	0.914	First differenced stationary
	First difference	-5.220	0.002	

The ADF test result (Table 2) shows that the first difference of time series is stationary and therefore, gives the similar result that of correlogram test. So the selected value of 'd' is '1'. The value of parameters 'p' and 'q' was selected through trial and error basis. Time series data for the period, 1991-2017 was analyzed by employing ARIMA model all the way through using SPSS (Statistical package for Social Sciences).

Table 3 Results of ARIMA Model.

Model	R^2	RMSE	MAPE	Ljung-Box Q	
				Statistic	p-value
ARIMA (1,1,0)	0.965	83288.956	10.002	15.828	.536
ARIMA (0,1,1)	0.965	83430.572	10.003	15.155	.584
ARIMA (1,1,1)	0.967	83642.157	9.619	14.968	.527
ARIMA (2,1,0)	0.967	83102.021	9.015	12.403	.716
ARIMA (0,1,2)	0.967	82665.141	8.821	10.846	.819
ARIMA (2,1,1)	0.967	84674.694	8.954	12.011	.678

By referring Table 3, among various models, ARIMA (0,1,2) model has highest value of R^2 and lowest value of RMSE, mean absolute percent error (MAPE) and appeared to be the best model. Moreover, the Ljung-Box test suggests that the ACF of residuals for the model at lag times is insignificant.

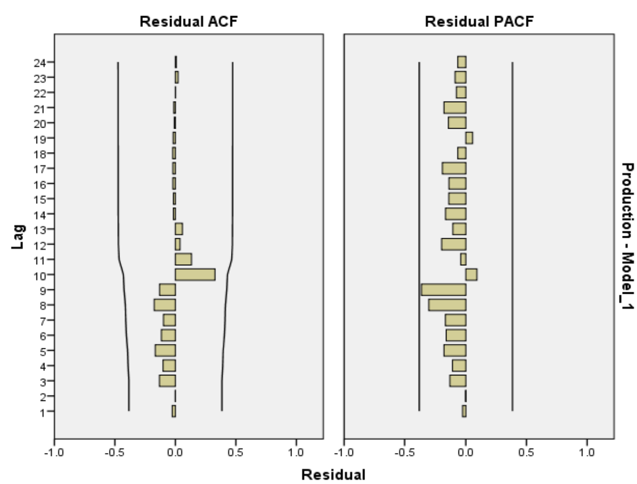

Fig 4 Correlogram and partial correlogram of residuals.

Table 4 Estimated values of ARIMA (0,1,2) Model.

Parameters	Estimate	Standard Error	t-statistic	p-value
θ	-10719430.58			
φ_1	0.03	0.21	0.15	0.88
φ_2	-0.27	0.21	-1.26	0.22

The correlogram and partial correlogram, given in Figure 4, indicates that there is no autocorrelation exists among the residuals of ARIMA (0,1,2) model. The estimated values for the ARIMA (0, 1, 2) model is presented in Table 4.

Table 5 Forecasting of mango production in Bangladesh for the period 2018 to 2025.

Year	Forecasted values	Lower 95% limit	Upper 95% limit
2018	1660021	1488603	1831439
2019	1786646	1548106	2025186
2020	1922222	1602962	2241482
2021	2063172	1679828	2446517
2022	2209496	1771343	2647650
2023	2361194	1874363	2848025
2024	2518265	1987200	3049331
2025	2680711	2108823	3252599

.

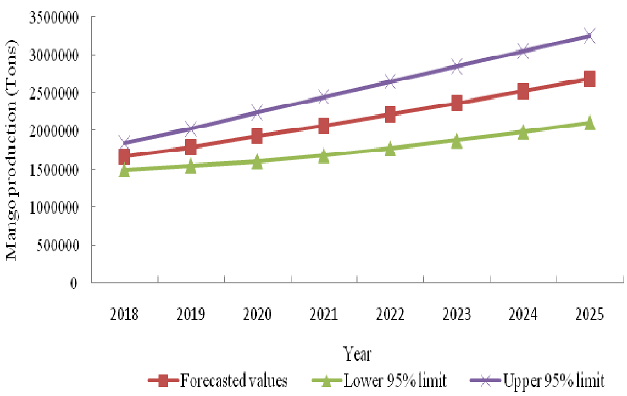

Fig 5 Forecasting of mango production in Bangladesh for the period 2018 to 2025.

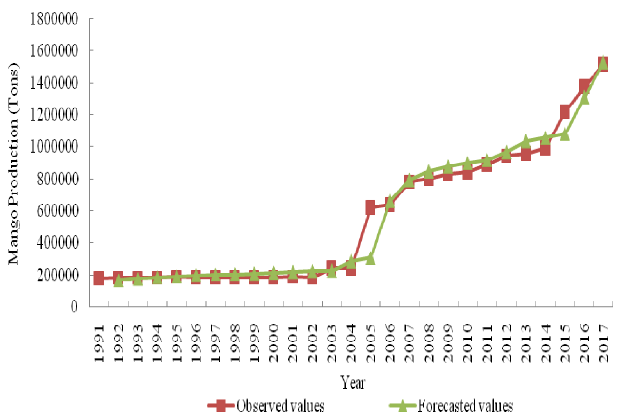

Fig 6 Observed and predicted (forecasted) values of mango production for the period 1991 to 2017.

The lagged values of moving average terms (φ_1, φ_2) are insignificant which indicates the error terms are not auto-correlated at lag 1 and lag 2 and the time series is stationary for the ARIMA (0, 1, 2) model. Hence, ARIMA (0, 1, 2) is found as the best fit for forecasting.

A forecast for mango production (with 95% confidence intervals) is generated by using ARIMA (0, 1, 2) model for the period 2018 to 2025. Forecasts (with their upper & lower limits at 95% confidence intervals) are presented in Table 5 and plotted in Figure 5. Data presented in Table 5 shows that production of mango would increase and predicted production of mango (forecasted values) will range between 1660021 and 2680711 tons during 2018-2025. This means that an increased quantity of mango will be available in future for domestic consumption and export.

The reliability of the estimated model is also checked for the period of 1991-2017. The observed and predicted (forecasted) values are presented in Table 6 and plotted in Figure 6. Minor differences were observed, however, the results reconfirmed that the estimated model is the best fit for forecasting production of mango in Bangladesh.

4. CONCLUSION

The present study is undertaken to estimate growth trend and to forecast production of mango in Bangladesh. Firstly, Growth trend of mango production is estimated by employing exponential model which results that production of mango grew at a rate of 9.7% per year for the period, 1991 - 2017. Then, ARIMA methodology is used to search out the best model for the market on the basis of R^2, root mean square error (RMSE) and mean absolute percent error (MAPE). The best model is found as ARIMA (0,1,2) which illustrates that the production of mango would increase and predicted production of mango (forecasted values) will range between 1660021 and 2680711 tons during the period, 2018-2025. Therefore, Bangladesh needs well-organized and properly managed mango orchards that can ensure sustainable production of the fruit and prove instrumental in enhancing export of mango from Bangladesh. Following measures are suggested for the rationalization of mango production and marketing to the development of mango-producing regions.

i. The forecast establishes that increased production of mango will be forthcoming in future. Given the trend, Government should create supportive infrastructure for handling surplus production of the fruit.
ii. Improved infrastructure (better roads, refrigerated transportation and cold storages) will ensure increased marketed surplus. A well conceived production and marketing plan can guarantee prosperous future for mango.
iii. There is need to launch a campaign for boosting export of Mango. New markets should be identified and a culture of value addition should be promoted. In order to increase export from Bangladesh, it also needs to ensure the authenticity of quality certificate.
iv. Bangladesh needs to modernize and update the existing production, harvesting, and post harvest management practices.

v. Packaging is also another important issue and it needs to focus on packaging industry as a priority area.

6. REFERENCES

Arivarasi, R & Ganesan, M. (2015). Time Series Analysis of Vegetable Production and Forecasting Using ARIMA Model. *Asian Journal of Science and Technology*, 6(10), 1844-1848.

Awal, M. A. & Siddique, M.A.B. (2011). Rice Production in Bangladesh Employing by ARIMA Model. *Bangladesh Journal of Agricultural Research,* 36(1), 51-62.

Badmus, M.A & Ariyo, O.S. (2011). Forecasting Cultivated Areas and Production of Maize in Nigeria Using ARIMA Model. *Asian Journal of Agricultural Sciences*, 3(3), 171-176.

Badar, N., Fatima., H., Jabbar, A. & Asif, M. (2015). Major Food Crops Production and Yield Forecast in Pakistan. *Pakistan Journal of Agricultural Research*, 28(3), 295-303.

BBS, (2016). *Statistical Yearbook of Bangladesh*, Statistical Division, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka.

Box, G. & Jenkins, G. (1970). *Time series of analysis, forecasting and control*. San Franscico, HoldenDay, California, USA.

Dhaka Tribune (2019). Himsagar Mango Receives GI Recognition. *Dhaka Tribune*, 26 January.

Gujarati, D. (2003). Basic Econometrics, 4th edition, McGraw-Hill.

Hamjah, M. A. (2014). Rice Production Forecasting in Bangladesh: An Application of Box-Jenkins ARIMA Model. *The International Institute for Science, Technology and Education, Mathematical Theory and Modeling,* 4(4), 1-11.

Khayati, A. (2015). Forecasting Major Vegetable Crops Productions in Tunisia. *International Journal of Research in Business Studies and Management*, 2(6), 15-19.

Rathod S. & Mishra G.C. (2017). Weather based modeling for forecasting area and production of mango in Karnataka. *International Journal of Agriculture, Environment and Biotechnology,* 10(1), 149-162.

Sabir, H.M. & Tahir, S.H. (2012). Supply and Demand Projections of Wheat in Punjab for the Year 2011-12. *Interdisciplinary Journal of Contemporary Research in Business*, 3(10), 800-808.

Suleman, N. & Sarpong S. (2011). Forecasting Milled Rice Production in Ghana Using BoxJenkins Approach. *International Journal of Agricultural Management & Development (IJAMAD)*, 2(2), 79-84.