SAARC J. Agric., 22(2): 169-180 (2024) https://doi.org/10.3329/sja.v22i2.75962

SAARC Journal of Agriculture
Journal Homage: https://www.banglajol.info/index.php/SJA

Research Article

Selection of the Best Time Series ARIMA Model to Forecast Onion Production in Bangladesh

M.R. Hasan[1*] and N. Khatun[2]

[1]Assistant Professor (Statistics), Department of Public Health, Leading University, Sylhet-3112, Bangladesh
[2]Lecturer, Department of Sociology, Tejgaon Mohila College, Dhaka-1215, Bangladesh

Abstract

Onion is one of the most commonly used spices in Bangladesh for preparing food and illness treatment. Since the demand for onions is rising daily, accurate projections must be made to implement policies based on that need. Hence, this study attempts to select the best Auto-Regressive Integrated Moving Average (ARIMA) model to forecast onion production in Bangladesh. Initially, this study considered Bangladesh's yearly onion production (in hg/ha) dataset from 1971 to 2017. This study then applied the Box-Jenkins approach to build the model and forecast the onion production in Bangladesh. The data from 2018 to 2022 were utilized to compare the forecasted values. The AIC, AIC_C, and BIC values, error metrics, and residuals of the fitted model were assessed using onion production data. The comparative analysis shows that ARIMA (0, 2, 1) is the best model for precise forecasts of onion production in Bangladesh. The proposed model may help policymakers forecast accurately and know the future trend of onion production in Bangladesh.

Keywords: ARIMA Model, Bangladesh, Box-Jenkins approach, Forecast, Onion

Introduction

Bangladesh is considered an agricultural country because of its monsoon climate and proximity to rivers. Of all the minor crops, onion (*Allium cepa* L.) is the most important seasonal crop for the human diet. Onion is essential in the human diet for its nutritional, aromatic, and medicinal qualities (Huda et al., 2008; Mishra et al., 2013; Sharma et al., 2017). One of the primary objectives of Sustainable Development Goals (SDGs) is to

[*]Corresponding author: rashidulhasan067@gmail.com

Received: 03. 09. 2024 *Accepted:03. 12. 2024*

achieve food security and ensure better nutrition, and onions are crucial to maintaining the people's proper nutritional status. Onion production is considered more profitable than competing spices and is a highly commercial, income-generating, and employment-generating activity (Anjum & Barman, 2017). It occupies the highest position among all other spice goods in Bangladesh in production (BBS, 2018). Due to its favorable agro-climatic conditions, onions are produced, processed, and marketed in Bangladesh. Onion is produced throughout Bangladesh, with the most extensive cultivation occurring in the following regions: Faridpur, Dhaka, Rajshahi, Cumilla, Jashore, Dinajpur, Mymensingh, Rangpur, and Pabna. Due to the rapid population growth, evolving consumption patterns, and the diverse applications of spices in various food items, the demand for onion in Bangladesh is consistently increasing.

Bangladesh's economy heavily depends on onions because they are widely consumed and have many applications and financial advantages. However, the Bangladeshi market faces difficulties such as odd price swings, intricate dynamics between supply and demand, and dishonest intermediary operations that hurt customers and growers by negatively impacting onion prices. However, insufficient government incentives make many farmers hesitant to grow onions. Bangladesh produced over 1866 thousand tons of onions on 459 thousand acres of land between 2016 and 2017, according to BBS (2018). However, the country must import 700–800 thousand tons of onions annually to meet the 2400 thousand tons of demand, primarily from India, the world's second-largest producer (Hossain, 2019). Such a large amount suggests that Indian onions greatly influence the domestic market price. Additionally, consumers' real incomes are directly impacted by the regular fluctuations in onion prices, which pushes them to adjust their consumption budgets. In this sense, boosting domestic production is crucial for helping consumers balance their food baskets on the one hand and decreasing reliance on imports on the other.

It is difficult for Bangladesh to produce even 15% of the country's yearly onion needs domestically, according to Hossain et al. (2017). The author further reported that the yield level is relatively low (about 370–500 kg/ha) compared to the larger yield (1000–1200 kg/ha) of onions grown in other countries. Lack of high-quality seed, improved varieties, and incorrect irrigation of onion production are all linked to reduced yield levels, which in turn cause an imbalance between supply and demand. Furthermore, such an imbalance frequently drives up the price of onions, directly affecting consumers' real incomes and causing them to adjust their consumption budget (Kondal, 2011). Therefore, increasing domestic production in the upcoming years will be crucial to help consumers balance their food basket and lessen their reliance on imports. Furthermore, it can only be accomplished with a precise forecast of onion production. Consequently, this study attempts to determine which Auto-Regressive Integrated Moving Average (ARIM) model is most suitable for predicting Bangladesh's onion production. Such an attempt may enable the government to design appropriate production schedules and trade policies to balance demand and supply in the domestic market through self-production rather than resorting to import.

The Box-Jenkins ARIMA model has been widely used in the field of forecasting econometric time series, inventory, and price modeling. Furthermore, this method is frequently used in the agricultural crop production process to foresee the yield of important crops including soybean, rice, wheat, jute, tea, palm oil, and so on. However, very limited use of ARIMA was found in forecasting onion production in Bangladesh. Several comparable studies (Kondal, 2011; Mishra et al., 2013; Sharma et al., 2017; Warade, 2016) that were carried out in India may be found in the literature. India leads the global onion market as an exporting nation. Thus, the policy-making in importing nations like Bangladesh may not be appropriate for the forecast based on the Indian context. Similar research was done nationally by Hossain et al. (2017); however, this study's shortcoming was that it only looked at onion production data from 1971 to 2013. Considering this limitation, an attempt has also been made in our work to forecast onion production spanning a large data set over 47 years using the most popular ARIMA model. However, several studies have been conducted in Bangladesh and abroad to analyze the forecast behavior of time series data by applying the ARIMA model. For instance, Mila and Parvin (2019) used the ARIMA model to forecast the area, production, and yield of onions in Bangladesh. Time series analysis, more specifically the ARIMA technique, was used by Kumar et al. (2024) to forecast onion prices in Gujarat, India's central wholesale market. Finally, this paper is organized under the following sections: Introduction, Materials and Methods, Results and Discussions, Conclusions and Policy Implications.

Materials and Methods

The Data

This study used data on yearly onion production (in hg/ha) in Bangladesh from 1971 to 2017, collected from the Food and Agricultural Organization (FAO) website (https://www.fao.org/faostat). The forecasting was made for 2018 to 2027. The data from 2018 to 2022 were used to validate the forecasting.

Box-Jenkins Approach

The Box-Jenkins approach is applied to univariate time series modeling and is particularly well suited to short-term forecasting (Ray & Bhattacharyya, 2020). The steps for the Box-Jenkins approach are identification, estimation, diagnostic checking, and forecasting. Before constructing the Box-Jenkins technique, it is essential to examine the stationarity of the time series. In this work, several well-known statistical tests, such as the Augmented Dickey-Fuller (1979) test and the Phillips-Perron (1988) test, were employed to examine the stationarity of time series. The identification step estimated the model's order using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Parameters were estimated using the Maximum Likelihood (ML) method after determining the model's order. The best model has been chosen using model selection criteria such as Akaike Information Criteria (AIC), Akaike Information Corrected Criteria (AIC_C), and Bayesian Information Criteria

(BIC). The standardized residuals plot, ACF plot of residuals, and histogram with a normal curve of residuals were used to perform diagnostic checking. The Ljung-Box test (Ljung and Box, 1978) was also used to test the residuals' autocorrelation. In this case, the null hypothesis is $H_o : \rho_1 = \rho_2 = ... = \rho_k = 0$ that is the autocorrelation among the residuals is zero is tested with the Ljung-Box statistic $Q^* = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\rho_k^{\ 2} \sim \chi^2(m)$. Where n is the number of observations used to estimate the model, ρ_k is the sample autocorrelation at lag k, and m is the number of lags being tested.

The accuracy of the selected model was measured using the Mean Absolute Percentage Error (MAPE). The MAPE is perhaps the most commonly used indicator of forecasting accuracy (Armstrong & Collopy, 1992; Goodwin & Lawton, 1999) and is defined by the following formula $MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100$. Where n is the number of observations, y_t is the observed value at time t, and $\hat{y}_t$ is the predicted value by the model for time t. Finally, the forecasted and actual values were compared to check the validity of the chosen model (Ray et al., 2016).

ARIMA Model

A generalization of an auto-regressive moving average model in time series analysis is called an auto-regressive integrated moving average (ARIMA) model. This study took a non-seasonal ARIMA (p, d, q) model into consideration because the data show no seasonal variation. The model can be derived as follows:

A time series process $\{Y_t\}$ is called an auto-regressive process of order p denoted by AR (p) and is defined by $Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \varepsilon_t$

If $\{\varepsilon_t\}$ is a white noise with mean zero and variance σ^2 then $\{Y_t\}$ is called a moving average process of order q denoted by MA (q) and is defined by $Y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + ... + \beta_q\varepsilon_{t-q}$.

The combination of AR and MA models are known as autoregressive moving average (ARMA) model. An ARMA (p, q) model is defined as

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + ... + \beta_q\varepsilon_{t-q}$$

Where, Y_t is the original time series.

For every t we assume that ε_t is independent of $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$.

Again, a time series $\{Y_t\}$ is said to follow an ARIMA model if the d th difference $W_t = \nabla^d Y_t$ is a stationary ARMA process.

If $\{W_t\}$ follows an ARMA (p, q) model, $\{Y_t\}$ is said to be an ARIMA (p, d, q) process. Thus, an ARIMA ($p, 1, q$) process is defined by

$$W_t = \alpha_1 W_{t-1} + \alpha_2 W_{t-2} + \ldots\ldots + \alpha_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots\ldots + \beta_q \varepsilon_{t-q}$$

Where, $W_t = Y_t - Y_{t-1}$

Statistical Software and Packages Used

The analysis of this study was carried out entirely using the open-source statistical programming software R Studio (version 4. 3. 0) for Windows. Several library packages were utilized for analysis, including TSA, forecast, MASS, and t series.

Results and Discussions

Plot of the Original and Differenced Time Series

The original onion production data are shown graphically in Figure 1. It is clear from Figure 1 that the yearly onion production in Bangladesh fluctuated over the study period of 1971 to 2017. It started in 1971 and fell till 1977, then reached a short peak in 1978 and 1980, and after that, the production again fell dramatically. So, there is a decreasing trend up to the year 1981. Between 1982 and 2000, onion production was almost equal, and then it fell again for a short time. After 2003, there was a dramatic increase in onion production, which led to an increasing trend; that is, the variation of onion production was unstable, leading to the data series not being stationary. One of the main causes of this increase is the revolution in the agricultural sector in Bangladesh.

The first step in building an ARIMA model is to identify whether the variable being forecasted is stationary in a time series. To obtain a stationary series, we take the difference 'd' of the original time series. The first and second differenced onion production series are presented in Figure 2. Figure 2(b) shows that the second differenced onion production data is stationary in its mean and variance. Thus, to stabilize the variance and to make the series stationary, the second difference is enough that the difference order is 2, and it is said that integrated order d = 2. But before moving further, we first test the stationarity (unit root problem) of the second differenced data series using the Augmented Dickey-Fuller and Phillips-Perron tests.

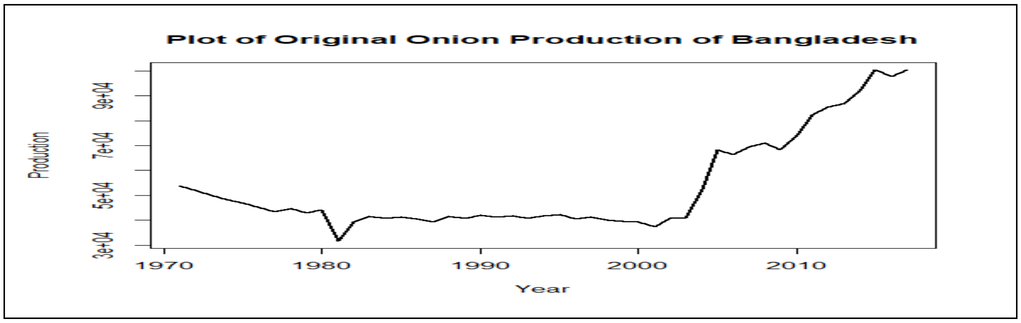


Fig. 1. Time series (original series) plot of onion production in Bangladesh

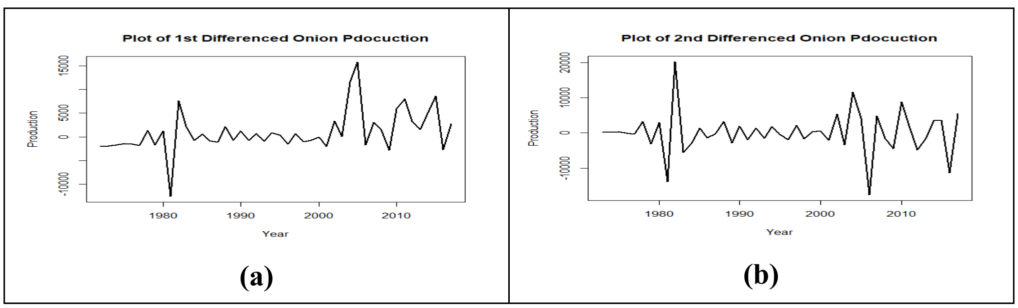


Fig. 2. Time series plot of first and second differenced onion production in Bangladesh

Stationarity Test

The test statistics for the Augmented Dickey-Fuller test and Phillips-Perron tests is

H_0: The series has a unit root.

H_1: The series does not have a unit root. That is the data series is stationary.

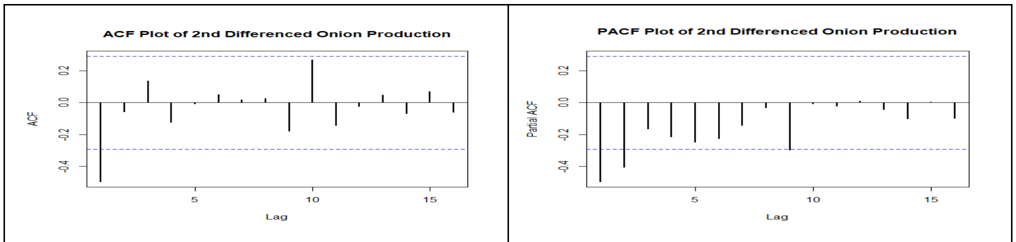
Table 1. ADF and PP tests for second differenced onion production series

Tests	Calculated value	Truncation lag parameter	P-value	Comment
ADF Test	-4.989	3	0.01	Stationary
PP Test	-56.838	3	0.01	Stationary

The p-values (=0.01) of the ADF and PP tests declared that the second differenced onion production data series is stationary, suggesting that there is no unit root.

Identification of Tentative Model

To detect AR and MA terms, the ACF and PACF plots of the second differenced onion production data are given below (Figure 3):



(a) (b)

Fig. 3. ACF and PACF plots of second differenced onion production in Bangladesh

In PACF, there are significant negative spikes at lag 1 and lag 2, and not much else. Clearly, Figure 3(b) has a cut-off on the PACF curve after 2nd lag, which is primarily an AR (2) process. So, AR (2) may be effective for onion production in Bangladesh. Also, in ACF, there is a significant negative spike at lag 1 and not much else. Clearly, Figure 3(a) has a cut-off on the ACF curve after 1st lag, which is primarily an MA (1) process. So, MA (1) may be effective for onion production in Bangladesh. The alternative negative and positive ACF and exponentially decaying PACF indicate an autoregressive moving average process.

Selecting the Best ARIMA Model from Tentative Models

The best ARIMA model is selected based on the lowest AIC, AIC_C, and BIC values. The ARIMA model with auto-regressive of order p = 2, difference of order d = 2, and moving average of order q = 1, and all possible tentative models with appropriate combinations of AR (p) and MA (q) terms are given in Table 2.

Table 2 shows that the ARIMA (0, 2, 1) is the best-selected model for forecasting onion production in Bangladesh based on the lowest AIC, AIC_C, and BIC values.

Table 2. Several ARIMA models and their AIC, AIC_C, and BIC values

Tentative Models	AIC	AIC_C	BIC
ARIMA (2, 2, 1)	-80.87	-79.87	-73.65
ARIMA (1, 2, 1)	-82.87	-82.28	-77.45
ARIMA (0, 2, 1)	-84.14	-83.86	-80.53
ARIMA (2, 2, 0)	-74.41	-73.83	-68.99

Parameter Estimation and Diagnostic Test of Models

The parameter of the fitted model was estimated using the maximum likelihood method, and the diagnostic test of the fitted model's residuals was performed using the Ljung-Box test (Table 3).

Table 3. Summary and Ljung-Box test of the fitted model

Fitted Model	Summary statistics				Ljung-Box test		
	Parameter	Estimated value	SE	P value	Lag used	Q^*	P value
ARIMA (0, 2, 1)	*q*	-0.907	0.063	<0.001	9	3.169	0.923

SE: Standard Error

The p-value (=0.923) of the Ljung-Box test strongly recommends no autocorrelation among the residuals of the fitted ARIMA model.

To further investigate the distribution of residuals in the fitted model, the residual plot, ACF and PACF plots of residuals, and histogram with a normal curve are presented below (Figure 4):

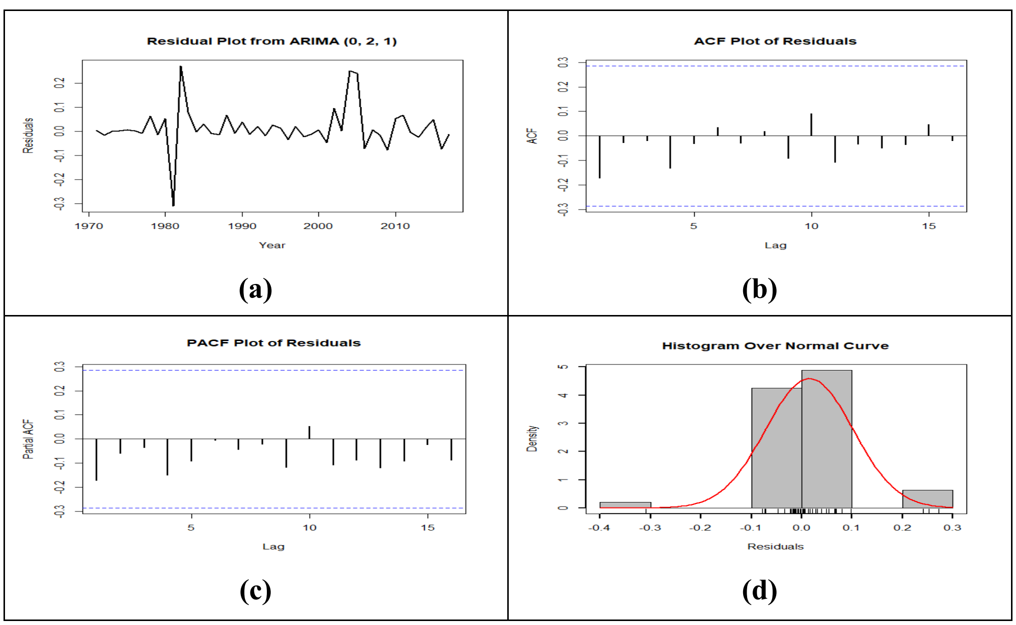


Fig. 4. Several plot of residuals and histogram with a normal curve

The residual plot (Figure 4(a)) illustrates that the standard errors are roughly constant in their mean and variance over time. However, there has been a higher variation in the past and most recent decades. Figures 4(b) and 4(c) show all the autocorrelation and partial autocorrelation coefficients are within the 95% confidence intervals, providing reliable evidence of no autocorrelation among the residuals of the fitted ARIMA model. The histogram with a normal residual curve (Figure 4(d)) indicates that the residuals of the fitted models are roughly normally distributed. To further investigate the adequacy of the fitted model, forecasting criteria values were presented in the following table.

Table 4. Forecasting criteria values of the fitted model

Fitted Model	Forecasting criteria			
	RMSE	MAE	MAPE	MASE
ARIMA (0, 2, 1)	4169.599	2479.866	4.795	0.864

RMSE: Root Mean Square Error, MAE: Mean Absolute Error, MAPE: Mean Absolute Percentage Error, MASE: Mean Absolute Scaled Error

The MAPE of the ARIMA (0, 2, 1) model is less than 10, which indicates that the model is highly accurate (Lewis, 1982). Therefore, ARIMA (0, 2, 1) is the best-fitted model and is adequate to forecast the onion production in Bangladesh. The study conducted in Bangladesh by Hossain et al. (2017) supports this outcome.

Forecasting

Forecasting onion production was made after selecting the best-fitted model. The first 47 years of data (1971–2017) were used to build the model. The model was then used to forecast onion production for 2018–2027. The five years of data (2018–2022) were used to compare the predicted and actual values to validate forecasting. Table 5 shows the best-fitted ARIMA (0, 2, 1) model's forecast value and 95% confidence interval (Figure 5). Table 5 shows that the ARIMA (0, 2, 1) model forecasts an increase in onion production from 2018 to 2027. The forecast and actual values are quite similar and in the 95% confidence interval. Figure 5 also shows that the forecast data series (dark blue color) fluctuates very little from the original data series (red color). Therefore, the forecasted onion production series is a much more accurate representation of the actual onion production series.

Table 5. Forecasted value of the onion production in Bangladesh

Year	Forecast value	Actual value	95% CI for forecast value	
			LCL	UCL
2018	104297	97348	87399	124462
2019	108245	104541	83303	140654
2020	112342	105455	80317	157137
2021	116595	116768	77786	174765
2022	121008	122542	75481	193996
2023	125589	--------	73294	215194
2024	130342	--------	71173	238703
2025	135276	--------	69085	264885
2026	140397	--------	67015	294132
2027	145711	--------	64953	326878

CI: Confidence Interval, LCL: Lower Confidence Limit, UCL: Upper Confidence Limit

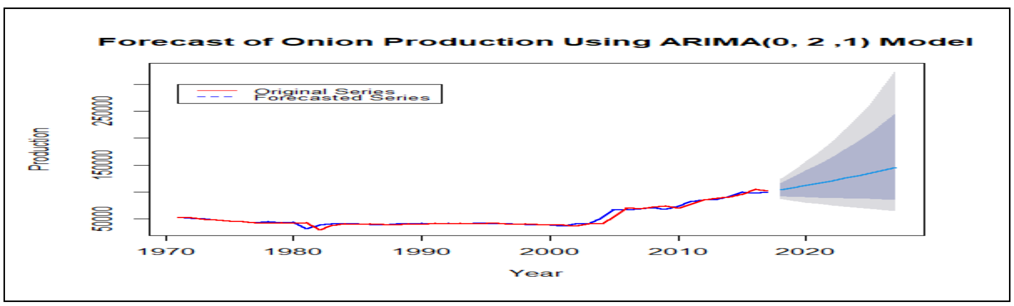


Fig. 5. Forecast from the best-fitted ARIMA model

Conclusions and Policy Implications

A time series model forecasts future values by identifying patterns in a variable's past movement. This study is intended to select the best time series ARIMA model to forecast onion production in Bangladesh. The Box-Jenkins ARIMA approach and several model selection criteria, such as AIC, AIC_C, and BIC, served the purpose. The experimental findings of this study conclude that ARIMA (0, 2, 1) is the best model to forecast onion production in Bangladesh. The graphical comparison between the observed and projected onion production shows slight variation, which indicates that the fitted model is forecasted well. The predicted values help businessmen or decision-makers identify whether to import or export the onion to meet the country's needs. The outcomes of the forecasting model also offer valuable insights for policymakers to stabilize onion prices, support farmers, and protect consumers' interests in the Bangladeshi market.

References

Anjum, A., and Barmon, B. K. (2017). Profitability and comparative advantage of onion (Allium cepa) production in Bangladesh: An economic study in some selected areas. *The Agriculturist*s, 15(2): 66-78. http://dx.doi.org/10.3329/agric.v15i2.35466

Armstrong, J. S., and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1): 69-80. https://doi.org/10.1016/0169-2070(92)90008-W

BBS. (2018). *Yearbook of Agricultural Statistics of Bangladesh-2017*. Statistics and Informatics Division, Bangladesh Bureau of Statistics, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka, Bangladesh. https://bbs.portal.gov.bd

Goodwin, P., and Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting,* 15(4): 405-408. https://doi.org/10.1016/S0169-2070(99)00007-2

Hossain, I. (2019). Onion market on fire: The price of 1 kg onion is now the same as 1 kg of broiler chicken. *Dhaka Tribune*. https://www.dhakatribune.com/bangladesh/nation/192054/

Hossain, M. M., Abdulla, F., and Parvez, I. (2017). Time series analysis of onion production in Bangladesh. *Innovare Journal of Agricultural Science,* 5(1): 1-4. https://www.researchgate.net/publication/310734848

Huda, F. A., Islam, M. S., Biswas, H., and Islam, M. S. (2008). Impact assessment study on selected spice crops under action plan in Bangladesh. *Progressive Agriculture,* 19(2): 229-241. http://dx.doi.org/10.3329/pa.v19i2.16965

Kondal, K. (2011). Growth rate of area, production and productivity of onion crop in Andhra Pradesh. *Indian Journal of Applied Research,* 4(3): 4-6. https://doi.org/10.15373/2249555 x/mar2014/2

Kumar, M., Sharma, R. K., Bhopala, U. D., and Suthar, B. (2024). Forecasting prices of onion in major wholesale markets of Gujarat. *International Journal of Statistics and Applied Mathematics,* 9(1S): 22-29. https://doi.org/10.22271/maths.2024.v9.i1Sa.1539

Lewis, C. D. (1982). Industrial and Business forecasting methods. *Journal of Forecasting,* 2(2): 194-196. https://doi.org/10.1002/for.3980020210

Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika,* 65(2): 297–303. https://doi.org/10.1093/biomet/65.2.297

Mila, F. A., and Parvin, M. T. (2019). Forecasting area, production and yield of onion in Bangladesh by using ARIMA model. *Asian Journal of Agricultural Extension, Economics & Sociology, 37*(4): 1–12. https://doi.org/10.9734/ajaees/2019/v37i430274

Mishra, P., Sarkar, C., Vishwajith, K. P., Dhekale, B. S., and Sahu, P. K. (2013). Instability and forecasting using ARIMA model in area, production and productivity of onion in India. *Journal of Crop and Weed,* 9(2): 96-101. https://www.researchgate.net/publication/275768973

Ray, S., and Bhattacharyya, B. (2020). Statistical modeling and forecasting of ARIMA and ARIMAX models for food grains production and net availability of India. *Journal of Experimental Biology and Agricultural Sciences,* 8(3): 296–309. http://dx.doi.org/10.18006/2020.8(3).296.309

Ray, S., Bhattacharyya, B., and Pal, S. (2016). Statistical modeling and forecasting of food grain in effects on public distribution system: An application of ARIMA model. *Indian Journal of Economics and Development,* 12(4): 739–744. http://dx.doi.org/10.5958/2322-0430.2016.00199.2

Sharma, H., Parihar, T. B., and Kapadia, K. (2017). Growth rates and decomposition analysis of onion production in Rajasthan State of India. *Economic Affairs,* 62(1): 157-161. http://dx.doi.org/10.5958/2230-7311.2017.00051.4

Warade, S. V. (2016). Forecasting of onion prices in Maharashtra: An approach to support vector regression and ARIMA model. *International Journal of Social Science and Humanities Research,* 4(2): 350–355. https://www.researchpublish.com/upload/book/Forecasting%20Of%20Onion%20Prices-3238.pdf