JnUJSci

Jagannath University Journal of Science
Volume 09, Number I, June 2022, pp. 1−7
https://jnu.ac.bd/journal/portal/archies/science.jsp
ISSN 2224-1698

Select an Appropriate ARIMA Model for Forecasting Future Behavior: Considering Four Pharmaceuticals Company of DSE

Research Article

Shahanaj Parvin

Department of Statistics, Jagannath University, Dhaka-1100, Bangladesh

Received: 19 July 2022, Accepted: 17 May 2023

ABSTRACT

Stock market plays a vital role on the providence of any country. This study attempts to select an appropriate ARIMA (Autoregressive Integrated Moving Average) model for forecasting the future behavior of four pharmaceuticals company; ACI Pharmaceuticals Limited (ACI), Beximco Pharmaceuticals Limited (BPL), RENETA and Square Pharmaceuticals Limited (SPL). Daily stock price data has been collected from January 2011 to December 2018. There are three parameters p, d and q in ARIMA model namely ARIMA (p,d,q).Box-Ljung (1978) diagnostic test was performed to select the p, d, q parameter that best fit to the data. At first checking the stationarity by Augmented Dickey Fuller test where the alternative hypothesis is the data is stationary. After taking first difference almost all the variable is stationary. Based on different selection criterion such as AIC, BIC, HQC, RMSE and MAE we can select the best forecasting ARIMA model for predicting stock market in each company under investigation. The outcome of the analysis indicated that ARIMA (0,1,1) model for ACI, ARIMA (0,1,1) model for BPL, ARIMA (0,1,1) for RENETA and ARIMA (0,1,2) for SPL are the finest forecast models respectively.

Keywords*: ARIMA, Augmented dickey fuller test, Forecasting, Dhaka stock exchange (DSE), Information criterion*

1. Introduction

Stock market plays an important role to any country for developing the economy. We can easily visualize countries growth rate through stock exchange rate. Stock exchange is that place where financial securities issued by company are bought and sold (www.bqprime.com). The first Stock Exchange of Bangladesh is Dhaka Stock Exchange (DSE) which is currently well assembled stock market.

Rationale of this study is data availability. That's why present study considered four Pharmaceuticals Company of DSE, which is the sample in this study. This study is a research exercise about the selection of ARIMA model. This study will enrich the present literature.

The main objective of this study is to find out an appropriate ARIMA model for forecasting the future behavior of four pharmaceuticals company; ACI Pharmaceuticals Limited (ACI), Beximco Pharmaceuticals Limited (BPL), RENETA and Square Pharmaceuticals Limited (SPL).

So the variables of the study are ACI, BPL, RENETA and SPL. Several works have been done related to stock market as well as ARIMA model in different time periods. Some of them are discussed below-

*Corresponding author: Shahanaj Parvin
Email: shahnaj26@gmail.com

In 2020 Zou conducted a study entitled "Research on GDP Forecast of Ji'an City Based on ARIMA Model". He found that ARIMA(0,2,1) model is the best model to forecast the GDP of Ji'an city from 2019 to 2023.

Rana (2019) conducted a study based on GDP movement data in Nepal which is executed by ARIMA modeling process. He showed that ARIMA (0,1,2) model is the most preferable model for explaining the GDP movement in Nepal.

In 2018 Parvin and Khanam studied forecasting behavior of price of Jute Goods (Hessian, Sacking, C.B.C, others) in Bangladesh by ARIMA model and VAR model. In their study ARIMA model is well fit for forecasting purpose compare with VAR model.

To forecast the selected macroeconomic time series variables (GDP, FDI and Remittance) Hanif and Khanam (2017) examined VAR model, ARIMA model and Holt's linear trend model. They have been found that the VAR model is the much more efficient model than others two.

Adebayo and Sivasamy (2014) conducted a study on ARIMA model. In this paper they tried to select finest ARIMA model for predicting stock market considering Botswana and Nigeria. On the basis of AIC, BIC, SQC, RMSE and MPE they shown that for Botswana and Nigeria stock market series ARIMA (3,1,1) and ARIMA (1,1,4) are the best forecast models respectively.

Khanam and Hafsa (2013) studied to predict watermelon production in Bangladesh where they used four different types of model (Supply model, Log-linear model, ARIMA model and MARMA model). And it was establish that ARIMA model is one step ahead than the MARMA model for forecasting purpose.

Another study conducted by Paul et al. in 2013 based on average daily share price index data series of Square pharmaceuticals Limited companies in Bangladesh. In their study they establish that ARIMA (2,1,1) model is the finest model for Square Pharmaceutical Limited data series considering forecasting purpose.

ARIMA as a standard model serve a satisfactory inflation forecasting in Nigeria (2012, Uko and Nkoro). This study used ARIMA, VAR and VECM model in prediction of inflation in Nigeria.

Alnaa and Ahiakpor (2011) conducted a study based on ARIMA model to predict inflation in Ghana. Box-Jenkins approach used to select satisfactory ARIMA model in this study. According to the study the AIC value establish that ARIMA (6,1,6) model is the satisfactory model.

In 2011 Datta found that ARIMA (1,0,1) model is the finest model for predicting inflation data in Bangladesh. He examined ARIMA model in forecasting the inflation of Bangladesh economy.

Merh et al. (2011) conducted a study based on stock market forecasting using ANN and ARIMA model. In this study ARIMA (1,1,1) gives satisfactory result than ANN(4-4-1). Future index value of sensex (BSE 30) has been used in this study.

To find out the best fitted model Rahman (2010) has conducted a study by using ARIMA model in predicting Boro rice production in Bangladesh. In the different step of analysis shown that ARIMA (0,1,0), ARIMA (0,1,3) and ARIMA (0,1,2) are the finest model for local, modern and total Boro rice production respectively.

Another study done by Haider and Kabir (2009) based on ARIMA model using Dhaka Stock Exchange (DSE) data series. They described that ARIMA (3,1,2) model is the most convenient model for index series and ARIMA (3,1) for return series.

2. Materials and Methods

2.1 Augmented Dickey Fuller test

Augmented Dickey-Fuller (ADF) test is very well known method of testing non-stationary status which is developed by Dickey and Fuller (1979). This test is simply called Dickey Fuller (DF) test. The following equations are estimated in order to run a DF test.

$\Delta Y_t = \Omega Y_{t-1} + \epsilon_t$ No drift, no intercept (1)

$\Delta Y_t = \beta_0 + \Omega Y_{t-1} + \epsilon_t$ Intercept, no drift term (2)

$\Delta Y_t = \beta_0 + \beta_1 t + \Omega Y_{t-1} + \epsilon_t$ with intercept and trend (3)

If white noise,ϵ_t has no autocorrelation then the Dickey-Fuller test is valid only. The probability of rejecting a correct null hypothesis is high if $\epsilon_t$is auto-correlated. There are three cases which are given as follows:

$\Delta Y_t = \Omega Y_{t-1} + \sum \alpha_i \Delta Y_{t-1} + \epsilon_t$ No trend, no intercept (4)

$\Delta Y_t = \beta_0 + \Omega Y_{t-1} + \sum \alpha_i \Delta Y_{t-1} + \epsilon_t$ Intercept, no trend term (5)

$\Delta Y_t = \beta_0 + \beta_1 t + \Omega Y_{t-1} + \sum \alpha_i \Delta Y_{t-1} + \epsilon_t$ With intercept and trend (6)

Now we set up the hypothesis -

H_0: $\Omega = 1$ ($Time\ series\ has\ unit\ root$)

H_1: $\Omega \neq 1$(Time series has no unit root)

Time series has unit root means the time series is non-stationary. Now we know that decision rule is based on either p-value or critical value. If we want to take decision by p-value than the decision rule is that if

P-value < level of significance

- Reject the null hypothesis

Another way to test the above null hypothesis is that if

Absolute ADF test statistic value > critical value

- Reject the null hypothesis otherwise not.

2.2 ARIMA model

Box-Jenkins method (1970) is used in this study of modeling and predicting non-seasonal time series data of four pharmaceuticals company last closing price in Bangladesh. In this method ARIMA model describes by several steps:

Step I: Choosing an appropriate ARIMA process

Step II: Fitting the data to it

Step III: Use the fitted model, and then

Step IV: Forecasting.

ARIMA model consists of three components such as p, d and q, namely we can write ARIMA (p,d,q) where p represents the order of the autoregressive parameters in the first component, d stands for the parameters' order of differentiation and lastly q indicates order of moving average parameters. This method requires the time series to initially be stationary in order to simulate the data. The first component, i. e, AR term, uses the p lags of time series to improve forecast. An AR (p) model has the form:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t$$

Where

$Y_t$is the dependent variable at time t.

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$are the dependent variables at time lags t-1, t-2,…,t-p respectively.

$\beta_1, \beta_2, \dots, \beta_p$are the coefficients to be estimated and $\epsilon_t$is the random error term at time t.

The next part is the integrated process. To make it stationary if the time series needs to be differenced for "d" times, then it is integrated of order "d" or I(d) (Gujrati,2003). The last component for the process is moving average process. MA (q) takes into account the q number of lags of errors for the improvement of the forecast. The model looks like:

$$Y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_q \epsilon_{t-q}$$

Where

$Y_t$is the dependent variable at time t.

$\phi_1, \phi_2, \dots, \phi_q$are the coefficients to be estimated and ϵ_t is the random error term at time t.

$\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$are the errors in previous time periods that are incorporated in the response Y_t.

Therefore the combined ARMA (p, q) can be written as-

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_q \epsilon_{t-q}$$

For the identification of a proper model through Box-Jenkins or ARIMA (p, d, q) procedure and forecasting the four pharmaceuticals company there are four steps: (i) Identification of model (ii) Estimation of model (iii) Diagnostics of model and (iv) Forecasting.

2.3 Statistical Analysis

The study has been directed through the processes listed below that were suggested by Box and Jenkins (1970) with the intention of developing an appropriate ARIMA model.

2.3.1 Identification of Stationarity

The Box-Jenkins model's first step is to determine if the series is stationary or not. If the data is non-stationary due to mean and variance than at first take log transformation and then taking difference transformation. Almost all of the variables in this study being stationary after taking the first difference. Unit Root test is performed for checking stationarity. Augmented Dickey Fuller (ADF) test statistic reveals either the null hypothesis is rejected or not where the alternative hypothesis is the data is stationary. After taking log transformation the ADF test statistic does not reject the null hypothesis but after taking the first difference ADF test statistic is rejected at 5% level of significance.

Variable	ADF test statistic	p-value	Difference	Decision
ACI	-12.724	0.01	First	Stationary
BPL	-12.361	0.01	First	Stationary
RENETA	-12.143	0.01	First	Stationary
SPL	-12.402	0.01	First	Stationary

Table 1: ADF test result

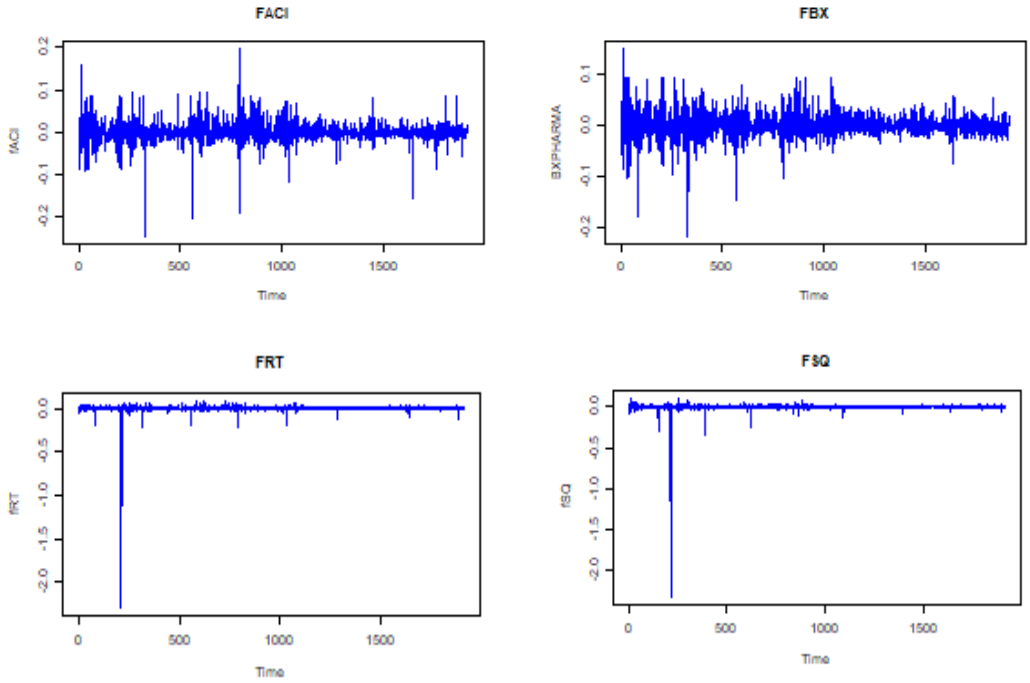


Figure 1: Stationary time series plot of ACI, BPL, RENETA and SPL

2.3.1 Model Estimation and Diagnostic checking

After the data being stationary the next step is to select lag order of ARIMA (p, d, q) model. ACF plot and PACF plot has been drawn for each variable. We can estimate the finest ARIMA model on the basis of AIC (Akaike Information Criterion) value, BIC (Bayesian Information Criterion) value, adjusted R square value, standard error of regression etc. In this study best ARIMA model is selected based on AIC value. Minimum AIC value indicates the model is best. Different lag order ARIMA model is taken and then examines the AIC value for each model. For ACI we see that AIC value is minimum for ARIMA (0,1,1), similarly for BPL ARIMA (0,1,1), for RENETA ARIMA (0,1,1) and finally for SPL ARIMA (0,1,2) exhibits minimum AIC value.

Several diagnostic tests must be run at this stage in order to forecast using the final model. A very strict assumption is that residuals are assumed to have white noise and uncorrelated is implied before choosing the most accurate model.

Figure 2 present ACF and PACF of the respective ARIMA (p, d, q) model given below

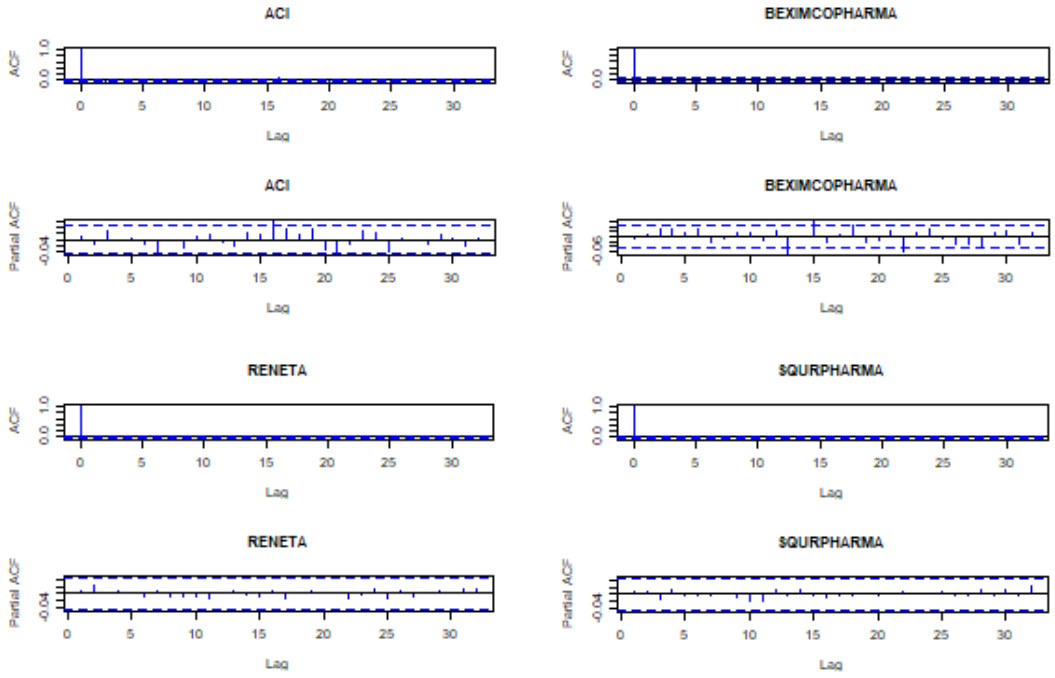
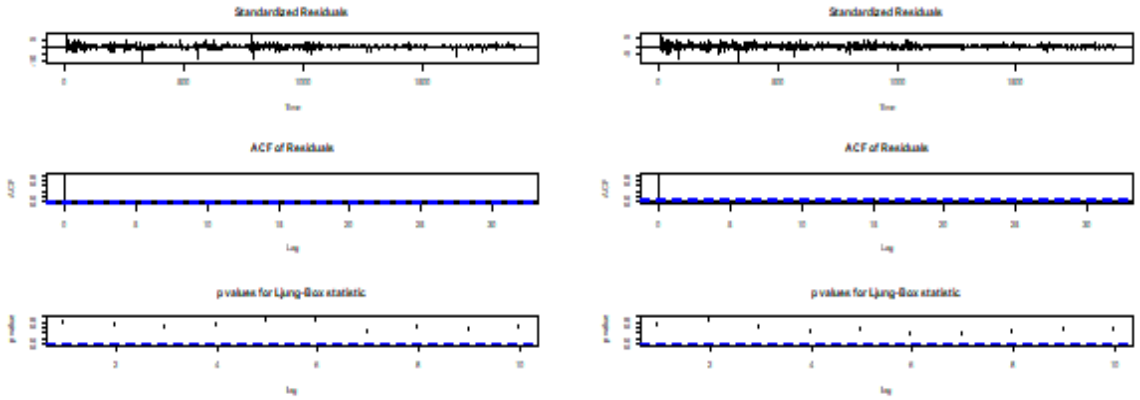


Figure 2: ACF and PACF plots of respective variables.

By examining the residuals we can assess the adequacy of Box-Jenkins model. For these reason we observe ACF and PACF functions. If the autocorrelations and partial autocorrelations are statistically zero than the residuals are randomly distributed. If they are not, the fitted model is not correct. Residual autocorrelation for fixed lags should be examined. That's why model residual autocorrelations for 30 lags have been observed in this study.

After choosing the model then we need to justify either the model is valid or not.



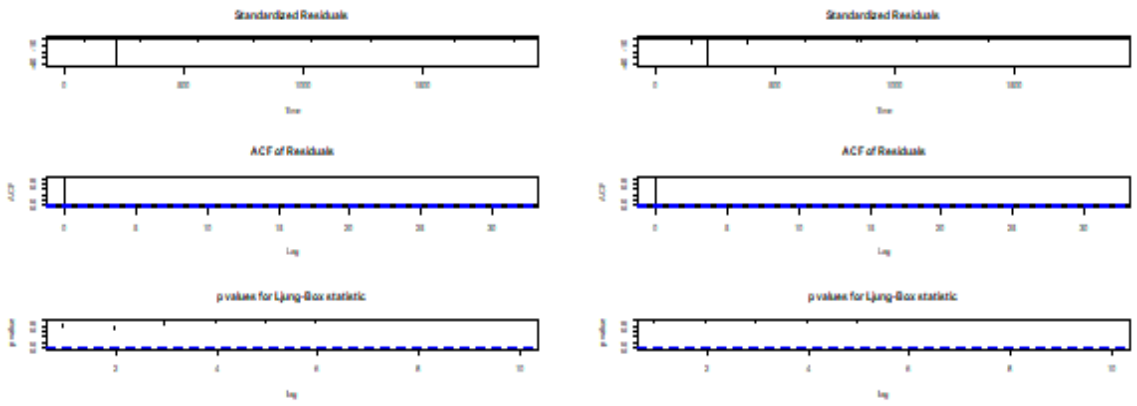


Figure 3: Residuals plot of respective variables

From figure 3 ACF plot of residuals (2nd band) we see that for almost all the variables residuals are into the confidence band that means residuals are independent. From the p-values plot we also see that at 5% level of significance for all the test statistic p-value is greater than the level of significance, that is we accept the test statistic for all values. On the basis of above discussion we can say that these models are valid for the above data set respectively.

2.3.2 Forecasting

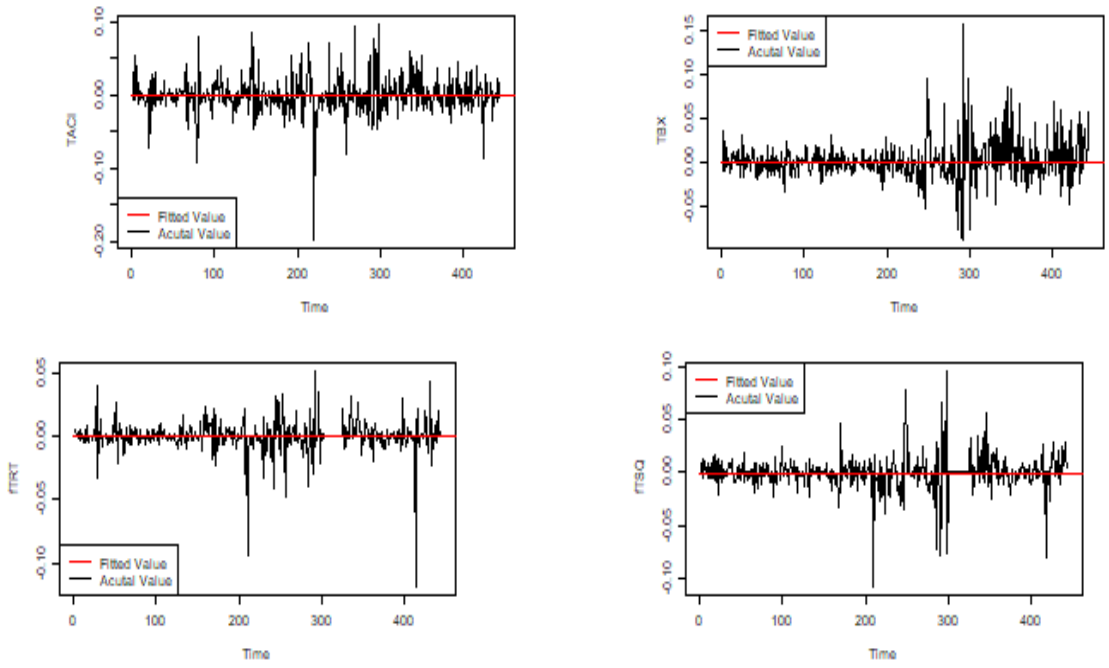


Figure 4: Actual and fitted values plot respectively

Figure 4 exhibits the actual and forecasted values of respective variable using appropriate ARIMA (p, d, q) model. For all the variables fitted value are almost stable because actual values are close to each other. For that reason, all the fitted line as like as straight line.

3 Conclusion

- From the above analysis it can be concluded that for forecasting ACI, BPL and RENETA daily stock price ARIMA (0,1,1) model is the most convenient model considering minimum AIC value which is -8847.14, -8839.19 and –5620.33 respectively.
- For forecasting SPL daily stock price ARIMA (0,1,2) model is the most convenient model considering minimum AIC value which is -5589.78.

In addition to the ARIMA model, there are many more econometric forecasting models such as Vector Auto-regression (VAR) model, Autoregressive Conditional Heteroskedasticity (ARCH) model, Neural networks and so on. In this study we only use ARIMA model and try to select best performing ARIMA model in apprehending share prices movements in Bangladesh. So further study might be executed by using other forecasting models mentioned above.

References

Adebayo FA, Sivasamy R. 2014. Forecasting Stock Market Series with ARIMA model.*Journal of Statistical and Econometric Methods*, 3(3):65-77.

Alnaa SE, Ahiakpor F. 2011. ARIMA (Autoregressive Integrated Moving Average) approach to predicting inflation in Ghana.*Journal of Economics and International Finance*, 3(5):328-336.

Box GEP, Jenkins GM. 1970. Time series analysis, forecasting and control.*San Francisco*: Holden Day.

Brooks C, Tsolacos S. 2000. Forecasting models of retail rents.*Environment and Planning A*, 32 (10): 1825- 1839.

Datta K. 2011. ARIMA Forecasting of Inflation in the Bangladesh Economy.*The IUP Journal of Bank Management*, X(4):7-15.

Dickey DA, Fuller WA. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root.*Journal of American Statistical Association*, 74(366):427-431.

Gujarati DN. 2003. Basic Econometrics, Fourth edition, *Tata McGraw Hill Publishing Company Limited*, New Delhi.

Haider AS, Kabir MR. 2009.Forecasting Dhaka Stock Exchange (DSE) return: An Autoregressive Integrated Moving Average (ARIMA) approach. *North South Business Review*, 3(1).

Hanif A, Khanam M. 2017. Analyzing the impact of Remittance and FDI on GNP in Bangladesh: An Econometric Time Series Analysis.*South Asian Journal of Population and Health*, 10(1&2): 97-107.

Hyndman R.2002. Box-Jenkins modeling. In H. Daellenbach,&R.Flood, (Eds.). The informed student guide to management science.London: Thomson.

Khanam M, Hafsa U. 2013.Market Model Analysis and Forecasting Behevior of Watermelon Production in Bangladesh.*Bangladesh Journal of Scientific Research*, 26(1&2): 47-56.

Merh N, Saxena VP, Pardasani KR. 2011. Next Day Stock market Forecasting: An Application of ANN & ARIMA. *The IUP Journal of Applied Science*, 17(1): 70-84.

Parvin S, Khanam M. 2018. Comparison between ARIMA and VAR Model Regarding the Forecasting of the Price of Jute Goods in Bangladesh.*The Dhaka University Journal of Science*, 66(2): 91-94.

Paul DJC, Haque MS, Rahman MM.2013. Selection of best ARIMA model for forecasting average daily share price index of pharmaceuticals companies in Bangladesh: A case study on Square Pharmaceutical Limited. *Global Journal of Management and Business Research Finance*, 13(3).

Rahman, N.M.F. 2010.Forecasting of Boro Rice production in Bangladesh: An ARIMA approach.*J.BangladeshAgril.Univ*., 8(1): 103-112.

Rana SB. 2019.Forecasting GDP movements in Nepal using Autoregressive Integrated Moving Average (ARIMA) modeling process.*Journal of Business and Social Sciences Research*, 4(2).

Said SE, Dickey DA. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrica*, 71(3): 599-607.

Sims, C.A.1980. Macroeconomics and Reality, *Econometrica*, 48:1-48.

Uko AK, Nkoro E. 2012. Inflation forecasts with ARIMA, Vector Autoregressive and Error Correction Models in Nigeria. *European Journal of Economics, Finance and Administrative Sciences*, 50:71-87.