Predicting the Stock Market Index Using Stochastic Time Series ARIMA Modelling : The Sample of BSE and NSE

* *C. Viswanatha Reddy*

Abstract

The stock market is basically volatile, and the prediction of its movement will be more useful to the stock traders to design their trading strategies. An intelligent forecasting will certainly abet to yield significant profits. Many important models have been proposed in the economics and finance literature for improving the prediction accuracy, and this task has been carried out through modelling based on time - series analysis. The main aim of this paper was to check the stationarity in time series data and predicting the direction of change in stock market index using the stochastic time series ARIMA modelling. The best fit ARIMA (0,1,0) model was chosen for forecasting the values of time series, that is, *BSE_CLOSE* and *NSE_CLOSE* by considering the smallest values of AIC, BIC, RMSE, MAE, MAPE, standard error of regression, and the relatively high adjusted R^2 values. Using this best fitted model, the predictions were made for the period ranging from January 7, 2018 to June 3, 2018 (22 expected values) using the weekly data ranging from January 6, 2014 to December 31, 2017 (187 observed values). The results obtained from the study confirmed the prospectives of ARIMA model to forecast the future time series in short-run and would assist the investing community in making profitable investment decisions.

***Keywords* : *BSE_CLOSE, NSE_CLOSE*, ARIMA model, forecasting, AIC, BIC, MAPE**

JEL Classification : C53, C58, E37, G170

Paper Submission Date : September 10, 2018 ; Paper sent back for Revision : July 18, 2019 ; Paper Acceptance Date : July 25, 2019

Stock market forecasting is an exercise to determine the future value of its performance index, that is, SENSEX, NIFTY. The successful prediction of any market's future index or a stock's future price will be more useful to the investing community to design optimal trading strategies that could yield significant profits. So, in the recent past, the concept of forecasting the stock market and its returns gained a lot of attention of the researchers. It may be because of the fact that if the directions of change in the market movements are successfully predicted, the investors may be better guided. Sometimes, the forecasted trends of the market will help the policy makers and regulators of the stock market in making curative decisions. The profit making investments and day to day operations in the capital market depend heavily on the forecasting ability.

Many practicing investors like Warren Buffett and other market researchers have proposed several models using various analytical methods, that is, fundamental analysis, technical analysis, and analytical techniques, etc. to give more or less exact forecasting. In addition to the above methods of forecasting, some traditional time series models were also used for it. Mainly, there are two kinds of time series models for forecasting, that is, linear models and non - linear models. Some of the examples of linear models are moving average, exponential smoothing, time series regression, etc. One among the most common and popular linear models is the

* *Professor & Head,* Department of Business Management, Rayalaseema University, Kurnool - 518 007, Andhra Pradesh. (E-mail : vsrits@gmail.com) ; ORCID id : https://orcid.org/0000-0002-2716-0624

DOI : 10.17010/ijf/2019/v13i8/146301

autoregressive integrated moving average (ARIMA) model proposed by Box and Jenkins (1976). In this paper, a modest attempt has been made to select the best fitted ARIMA model from different stochastic models that satisfies all the criteria of goodness of fit statistics for making the predictions and also to forecast the future values of stock market indices.

Review of Literature

It is pertinent to review the accessible literature connected to time series modeling and forecasting using ARIMA model. Most of the literature is focused on the identification of suitable ARIMA time series model and forecasting the gold price, exchange rates, oil palm prices, inflation rates, electricity consumption, etc. Only few studies are available relating to the forecasting of stock prices and stock market indices. Hence, this paper has been mainly devoted to the studies related to the determination of best ARIMA time series model and forecasting of future stock prices and stock market indices.

Meyler, Kenny, and Quinn (1998) developed the ARIMA time series predicting model for predicting the inflation in Ireland. In their study, they focused on maximizing the power of forecasting by minimizing forecast errors. Contreras, Espinola, Nogales, and Conejo (2003) examined the trends in daily prices of electricity in spot and forward contracts for mainland Spain and Californian markets and provided the best - suited ARIMA method to predict the next day electricity prices.

Nochai and Tidia (2006) conducted a study with an objective to find an appropriate ARIMA model for forecasting three types of oil palm prices by considering the minimum mean absolute percentage error. The empirical analysis of the study showed that ARIMA (2,1,0), (1,0,1), and (3,0,0) are the best models for forecasting the farm prices of oil palm, wholesale price of oil palm, and pure oil price of oil palm, respectively.

In a study conducted by Jarrett and Kyper (2011) using the data developed by Pacific - Basin Capital Markets (PACAP) and the SINOFIN Information Services Inc. demonstrated the usefulness of ARIMA - intervention time series analysis as both an analytical and forecast tool. The study indicated the usefulness of the developed model in explaining the rapid decline in the values of the price index of Shanghai market during the world economic decline in China in 2008. The authors concluded that the daily stock price index contained an autoregressive component; hence, it was better to forecast the stock returns using ARIMA model.

Banerjee (2014) used the ARIMA model for predicting stock market indices and also highlighted that they have an undue influence on the progress of the Indian economy. The study dealt with the identification of the best fit ARIMA model and after that predicted the SENSEX using the justified model.

Adebiyi and Adewumi (2014) presented the procedure for developing ARIMA models for forecasting share prices during the short-run. The results of the study explained the power of ARIMA models in predicting the stock prices in the short-run, which would help the investors in their decisions. A study was conducted by Jadhav, Kakade, Utpat, and Deshpande (2015) for forecasting the Indian share market using ARIMA model and said that artificial neural networks (ANNs) are universal approximates that can be applied to a wide range of time series for forecasting futuristic values in share market and give bright scope for investment. However, in their study, they proposed a novel hybrid model of ANN using ARIMA model instead of only artificial neural network for improving the predictive performance.

Guha and Bandyopadhyay (2016) examined the application of ARIMA time series model to forecast the future gold prices based on the past data from November 2003 to January 2014 to mitigate the risk in purchase of gold and, hence, to give guidelines for the investor when to buy or sell the yellow metal. The authors opined that nowadays, gold has gained importance as one of the investment alternatives ; it has become necessary to predict the price of gold with an appropriate method.

Savadatti (2017) carried out a study to identify the best fitted ARIMA models for forecasting the area, production, and productivity of food grains for 5 years. Based on univariate time series analysis, the study identified ARIMA (2,1,2), ARIMA (4,1,0), and ARIMA (3,1,3) models for forecasting the data on area, production, and productivity of food grains, respectively and these models were found to be adequate. The forecast values indicated that production and productivity increased during the forecast period but that of area exhibited near stagnancy, calling for timely measures to enhance the supply of food grains to meet the increasing demand in the future years.

Dikshita and Singh (2019) examined the different volatility estimators for forecasting volatility with high accuracy by traders, option practitioners, and various players of the stock market. The study evaluated the efficiency and bias of various volatility estimators based on various error measuring parameters, that is, ME, RMSE, MAE, MPE, MAPE, MASE, and ACFI. The study identified the Parkinson estimator as the most efficient volatility estimator. The study also suggested that the forecasted values were accurate based on the values of MAE and RMSE.

It may be concluded that many researchers have conducted studies to give reasons for the selection of ARIMA model for forecasting the time series data of a single variable with better accuracy. However, no researcher has focused on forecasting stock market indices in the Indian context. The present work is an effort to forecast the indices of BSE and NSE based on the past 187 weeks using the best fitted ARIMA model.

Statement of the Problem

Due to them being dynamic and non - linear in nature, it is very tricky to predict the stock exchange movements precisely. However, it is necessary to forecast and uncover non - linearity of the stock market to enable individual and institutional investors to design appropriate trading strategies and to achieve better results out of their investment endeavors. Hence, stock market forecasting has become a significant theme and has motivated researchers to build improved forecasting models. There are quite a few methods of statistical forecasting, that is, regression analysis, classical decomposition method, Box and Jenkins and smoothing techniques, with different degrees of accuracy. The accuracy of a forecasting model is based on the minimum errors of forecasting, that is, root mean square error, mean absolute error, standard error of regression, adjusted R-square, Akaike information criterion, Bayesian information criterion, etc. Among several methods of time series forecasting, the Box and Jenkins method is quite accurate compared to other methods and may be applicable to all types of data movements. This paper is an attempt to test the stationarity in the given time series and selecting the best suitable ARIMA model (also known as Box - Jenkins methodology) for short-term forecasting of BSE and NSE. The results obtained from the study can aid the investors in their investment decision - making process.

Objectives of the Study

The objectives of the study are listed below :

(1) To test the stationarity of the time series data compiled for the study, that is, weekly closing index values of BSE *(BSE_CLOSE)* and NSE *(NSE_CLOSE)*.

(2) To choose the optimum ARIMA model for estimating the series.

(3) To forecast the indices of BSE and NSE using the selected time - series ARIMA model.

Research Methodology

(1) Research Design : Keeping in view of the above listed objectives of the study, an exploratory research design and stochastic modeling has been adopted. Exploratory research is one which interprets the already available information and it lays particular emphasis on the analysis and interpretation of the available secondary data. Stochastic modeling is used for selecting the best ARIMA model and forecasting the time series using the selected model.

(2) Sources of Data : The data required for the present study is secondary in nature and has been compiled from an online source, that is, yahoofinance.com. The weekly closing indices of BSE and NSE are obtained from the website for the period from June 6, 2014 to June 3, 2018 (209 observations). From this range of data, I have taken the sample data ranging from June 6, 2014 to December 31, 2017 (187 observations) for making predictions ranging from January 7, 2018 to June 3, 2018 (22 observations).

(3) Hypothesis : The null hypothesis is generally defined as the presence of a unit root and the alternative hypothesis is stationarity (or trend - stationary).

- **H_{01}** : $\delta = 1$, there is unit root and the series (*BSE_CLOSE* and *NSE_CLOSE*) is non stationary.
- **H_{a1}** : $\delta < 1$, there is no unit root and the series (*BSE_CLOSE* and *NSE_CLOSE*) is stationary.

Analysis and Results

To select the best fitted ARIMA model, among several experiments conducted, many statistical tools are to be applied, that is, root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), Akaike information criterion (AIC), Bayesian information criterion (BIC), etc.

The RMSE has been used as a standard metric to measure the model performance in stock market forecasting. While applying the RMSE, the underlying assumption is that the errors are unbiased and follow a normal distribution. It provides a complete picture of the error distribution and its value should be relatively low (Draxler, 2014). The RMSE can be calculated by using the following formula :

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x}_i)^2}{n}} \qquad \text{.................(1)}$$

Mean absolute error measures the average magnitude of the errors in a set of predictions without considering their directions. It is the average over the test sample of the absolute differences between prediction and actual observations where all individual differences have equal weight. Hence, its value should be low. The MAE coefficient is given by the following equation (Draxler, 2014) :

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}_i| \qquad \text{.................(2)}$$

The mean absolute percentage error is a measure of prediction accuracy of a forecasting method. It usually expresses the forecasting accuracy of a model in percentage terms ; hence, its value should be maximum. The MAPE formula as stated by Tofallis (2015) is :

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \left| \frac{x_i - \bar{x}_i}{x_i} \right| \qquad \text{.................(3)}$$

Bayesian information criterion, also known as Schwarz information criterion (SIC), is a criterion for model selection among a finite set of models. It is based on the likelihood function, and it is closely related to Akaike information criterion (AIC). Mathematically, the BIC is an asymptotic result derived under the assumption that the data series is exponentially distributed. The BIC was developed by Schwarz (1978), who gave a Bayesian argument for adopting it.

$$BIC = \log\left(\frac{rss}{n}\right) + \frac{k}{n}\log n \qquad \text{.................(4)}$$

where, rss is residual sum of squares ; k is the number of coefficients estimated, that is, $1 + p + q + P + Q$; and n is the number of observations.

(1) Augmented Dickey - Fuller Unit Root Test : The initial phase of structuring the ARIMA model is to recognize whether the variable being predicted is stationary in time series or not. Most forecasting methods assume that a distribution has stationarity. A time series has stationarity if a shift in time does not cause a change in the shape of the distribution, that is, the mean and auto-covariance of the series do not depend on time (Tsay, 2005). Unit roots are one cause for non-stationarity. An absence of stationarity can cause unexpected behaviors in data series. Most real-life data sets just are non - stationary and we should make it stationary in order to get any useful predictions from it. Augmented Dickey - Fuller (ADF) unit root test examines whether a time series variable is non - stationary and possesses unit root. A common example of a non-stationary series is the random walk. We may write the random walk model (RWM) with stochastic process as (Garekos & Gramacy, 2013 ; Rao & Mukherjee, 1971) :

$$Y_t = \delta Y_{t-1} + u_t \ (-1 \leq \delta \leq 1) \qquad \text{.................(5)}$$

where,

t = time measured chronologically ; and

u_t = white noise error term.

For theoretical reasons, we manipulate equation (5) by subtracting Y_{t-1} from both the sides to obtain :

$$Y_t - Y_{t-1} = \delta Y_{t-1} - Y_{t-1} + u_t$$

$$Y_t - Y_{t-1} = (\delta - 1) Y_{t-1} + u_t \qquad \text{..................(6)}$$

which can be written as :

$$\Delta Y_t = \beta Y_{t-1} + u_t \qquad \text{..................(7)}$$

where, $\beta = (\delta - 1)$, and Δ = first difference operator.

In practice, instead of estimating equation (5), we estimate equation (7) and test the hypothesis (null) that $\beta = 0$. If $\beta = 0$, then $\delta = 1$, that is, we have a unit root, meaning that the time series under consideration is non - stationary. Before we proceed to estimate equation (7), it may be noted that if $\delta = 0$, equation (7) will become :

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \qquad \text{..................(8)}$$

Since u_t is the white noise error term, it is stationary, which means that the first differences of a random walk time series are stationary.

Before running the ADF test, one should inspect the data to figure out an appropriate regression model. We have three versions of the test :

Type 0 No Constant, No Trend $\Delta Y_t = \beta_1 Y_{t-1} + u_t$

Type 1 Constant, No Trend $\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$

Type 2 Constant, Trend $\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 t + u_t$

The Augmented Dickey - Fuller adds lagged differences to the above models (Gujarati, 2004) :

Type 0 No Constant, No Trend $\Delta Y_t = \beta_1 Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + u_t$

Type 1 Constant, No Trend $\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + u_t$

Type 2 Constant, Trend $\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 t + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + u_t$

where,

u_t = Error term; and

ΔY_{t-i} = Lagged differences.

Number of lagged differences to be added in the model is often decided numerically so that the residuals are not serially correlated. Moreover, there are several options for choosing lags : Minimize Akaike's information criterion (AIC), or Bayesian information criterion (BIC), or drop lags until the last lag is statistically significant.

ADF test with intercept was applied on both series to test the data for stationarity. The null hypothesis is tested through t - statistics, which is given by the following formula :

$$t = \frac{\hat{\delta} - \delta_{H0}}{SE \text{ of } \hat{\delta}} \qquad \text{..................(9)}$$

If 't' calculated is greater than the critical value, we do not reject the null hypothesis and the series under consideration would be non - stationary and has a unit root. On the other hand, if 't' calculated is less than the critical value, we reject the null hypothesis and the series under consideration would be stationary and does not have the unit root. First, the series should be tested on level and if it does not become stationary, then we should test the series at the first and second difference sequentially. 'p' – value is also used to reject or accept the null hypothesis. If the 'p' – value is less than 0.05 ($p < 0.05$), the null hypothesis is rejected and vice - versa.

The ADF Test (Table 1) at level depicts that the calculated 't' – value is greater than the critical values at 1%, 5%, and 10% levels of significance. At levels, both the underlying series (*BSE_CLOSE* and *NSE_CLOSE*) are non - stationary. The p - value of the series is also greater than 0.05. Hence, we do not reject the null hypothesis (H_{01}) and accept the alternative hypothesis (H_{a1}) that the series has a unit root. When the series (Y_t) is non - stationary, it must be differenced 'd' times before it becomes stationary, then it is said to be integrated of order 'd' (Brooks, 2008). The results of ADF test at first difference are presented in the Table 2.

The ADF test (Table 2) at first difference reveals that both the series are stationary at first difference. The calculated value of *DBSE_CLOSE* is -13.97590, which is less than the critical values at all levels of significance. Similarly, the 't' – statistics of *DNSE_CLOSE* is -14.28255, which is also less than the critical values at all levels of significance. Therefore, the null hypothesis (H_{01}) is rejected and it can be concluded that both the series are

Table 1. Augmented Dickey - Fuller Test at Level

H_0 is: *BSE_CLOSE* has a unit root.					H_0 is: *NSE_CLOSE* has a unit root.				
Exogenous: Constant, Linear Trend					Exogenous: Constant, Linear Trend				
Lag Length: 0 (Automatic-based on AIC, maxlag = 14)					Lag Length: 0 (Automatic-based on AIC, maxlag = 14)				
		***t*-statistic**	**Prob.***				***t*-statistic**	**Prob.***	
ADF test statistic		-1.603668	0.7886		ADF test statistic		-1.787570	0.7075	
Test critical values	1% level		-4.002786		Test critical values	1% level		-4.002786	
	5% level		-3.431576			5% level		-3.431576	
	10% level		-3.139475			10% level		-3.139475	
*MacKinnon (1996) one-sided *p* - values.					*MacKinnon (1996) one-sided *p* - values.				
Augmented Dickey - Fuller Test Equation					Augmented Dickey - Fuller Test Equation				
Dependent Variable: *D(BSE_CLOSE)*					Dependent Variable: *D(NSE_CLOSE)*				
Method: Least Squares					Method: Least Squares				
Date:06/22/18 Time:05:51					Date:06/22/18 Time:07:28				
Sample (adjusted): 6/15/2014 6/03/2018					Sample (adjusted): 6/08/2014 6/03/2018				
Included Observations: 208 after adjustments					Included Observations: 209 after adjustments				
Variable	**Coefficient**	**Std. Err.**	***t*-stat**	**Prob.**	**Variable**	**Coefficient**	**Std. Err**	***t*-stat**	**Prob.**
BSE_CLOSE(-1)	-0.028791	0.017953	-1.603	0.1103	*NSE_CLOSE*(-1)	-0.032916	0.018414	-1.787	0.0753
C	713.4920	449.9443	1.585	0.1143	*C*	246.6873	137.8652	1.789	0.0750
@TREND ("6/08/2014")	1.538760	0.901803	1.706	0.0895	@TREND ("6/08/2014")	0.540804	0.300068	1.802	0.0730
R-squared	0.015160	Mean dependent var		49.112	*R*-squared	0.017105	Mean dependent var		15.235
Adj *R*-squared	0.005552	*S.D.* dependent var		520.18	Adj *R*-squared	0.007105	*S.D.* dependent var		158.44
S.E. of regression	518.7404	AI Criterion		15.355	S.E. of regression	157.8433	AI Criterion		12.975
Sum squared resid	55163775	Schwarz criterion		15.403	Sum squared resid	5132391	Schwarz criterion		13.023
Log likelihood	-1593.920	HQ Criterion		15.374	Log likelihood	-1352.922	HQ Criterion		12.994
F-statistic	1.577813	Durbin-Watson Stat		1.9243	*F*-statistic	1.792443	Durbin-Watson Stat		1.9596
Prob (*F*-statistic)	0.2018922				Prob (*F*-statistic)	0.169141			

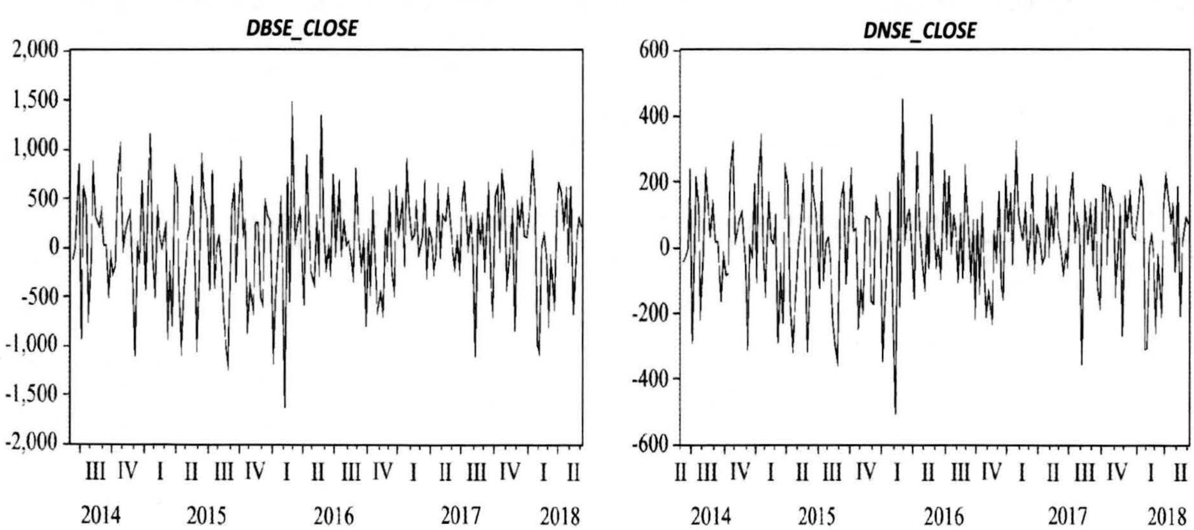
Table 2. Augmented Dickey - Fuller Test at First Difference

H_0 is: *DBSE_CLOSE* has a unit root.				H_0 is: *D(DNSE_CLOSE)* has a unit root.			
Exogenous: Constant, Linear Trend				Exogenous: Constant, Linear Trend			
Lag Length:2 (Automatic-based on AIC, maxlag=14)				Lag Length:2 (Automatic-based on AIC, maxlag=14)			
		***t*-statistic**	**Prob.***			***t*-statistic**	**Prob.***
ADF test statistic		-9.75042	0.0000	ADF test statistic		-9.796126	0.0000
Test critical values	1% level		-4.002449	Test critical values	1% level		-4.003226
	5% level		-3.431896		5% level		-3.431789

10% level -3.139664	10% level -3.139601
*MacKinnon (1996) one-sided *p* - values.	*MacKinnon (1996) one-sided *p*-values.
Augmented Dickey - Fuller Test Equation	Augmented Dickey - Fuller Test Equation
Dependent Variable: *D(BSE_CLOSE)*	Dependent Variable: *D(NSE_CLOSE)*
Method: Least Squares	Method: Least Squares
Date:06/22/18 Time:14:05	Date:06/22/18 Time:14:03
Sample (adjusted): 7/06/2014 6/03/2018	Sample (adjusted): 6/29/2014 6/03/2018
Included Observations: 205 after adjustments	Included Observations: 206 after adjustments

Variable	Coeffi	Std. Err.	*t*-stat	Prob.	Variable	Coeffi	Std. Err	*t*-stat	Prob.
D(NSE_CLOSE(-1))	-1.1881	0.0179	-9.7504	0.000	*D(NSE_CLOSE*(-1))	-1.2034	0.1228	-9.7961	0.0000
D(NSE_CLOSE(-1),2)	0.2015	0.0972	2.0725	0.039	*D(NSE_CLOSE*(-1),2)	0.1979	0.0985	2.0080	0.0460
D(NSE_CLOSE(-2),2)	0.1285	0.0696	1.8452	0.066	*D(NSE_CLOSE*(-2),2)	0.1445	0.0698	2.0705	0.0397
C	-12..4935	74.332	-0.1680	0.866	*C*	-5.9844	22.739	0.2631	0.7927
@TREND ("6/08/2014")	0.6378	0.6143	1.03829	0.300	@TREND ("6/01/2014")	0.1204	0.1867	0.6450	0.5199
R-squared	0.50384	Mean dependent var		-3.149	*R*-squared	0.51073	Mean dependent var		0.3597
Adj *R*-squared	0.49391	*S.D.* dependent var		729.24	Adj *R*-squared	0.50099	*S.D.* dependent var		225.165
S.E. of regression	518.778	AI Criterion		15.364	S.E. of regression	159.057	AI Criterion		13.008
Sum squared resid	53426199	Schwarz criterion		15.445	Sum squared resid	5085154.	Schwarz criterion		13.081
Log likelihood	-1569.904	HQ Criterion		15.397	Log likelihood	-1334.039	HQ Criterion		13.033
F-statistic	50.77432	Durbin-Watson stat		1.9527	*F*-statistic	52.45409	Durbin-Watson stat		1.9890
Prob (*F*-statistic)	0.0000				Prob (*F*-statistic)	0.000000			

Figure 1. Graphs of *BSE_CLOSE* and *NSE_CLOSE* Series at First Difference



stationary at 1%, 5%, and 10% levels of significance, and both the series do not have the unit root. An informal method to test the stationarity also confirms the results of the formal test, that is, the ADF test. Graphs of both the series at first difference do not demonstrate any kind of trend ; there are fluctuations in the graphs. These fluctuations epitomize the stationarity of the underlying series (shown in Figure 1).

(2) Correlogram Analysis : A correlogram (also called auto correlation function plot) is an image of correlation statistics and it gives a summary of correlation at different periods of time, that is, serial correlation. Serial correlation is where an error at one point in time travels to a subsequent point of time. It is a commonly used tool for checking the randomness in a data set. A correlogram contains the autocorrelation function (ACF) and partial autocorrelation function (PACF). Autocorrelation refers to the way the observations in a time series are related to other and is measured by a simple correlation between current observation (Y_t) and the observation 'p' periods (*lag p*) from the current one (Y_{t-p}) (Abdullah, 2012 ; Brooks, 2008). The autocorrelation coefficient at '*lag p*' is given by :

$$r_p = \frac{c_p}{c_0} \qquad \text{................(10)}$$

where,

c_p = the auto - covariance function; and

c_0 = the variance function.

$$c_p = \frac{1}{N}\sum_{t=1}^{N-p}(Y_t - \bar{Y})*(Y_{t+p} - \bar{Y}) \qquad \text{................(11)}$$

$$c_0 = \frac{1}{N}\sum_{t=1}^{N}(Y_t - \bar{Y})^2 \qquad \text{................(12)}$$

The resulting value of 'r_p' will range between -1 and +1.

Partial autocorrelations (PACF) are used to measure the degree of association between Y_t and Y_{t-p} when the effect of other time lags 1, 2, 3, ..., (p-1) are removed. The Figure 2 represents the plot of correlogram (ACF and PACF coefficients) of the time series *BSE_CLOSE* and *NSE_CLOSE* for lags 1 to 20 at the level (zero order difference). We may infer from the correlogram that the ACF of *BSE_CLOSE* and *NSE_CLOSE* were dropped away very gradually; thus, the data in time series is non - stationary. Hence, there is a need to convert non - stationary series into stationary by differencing.

The Figure 3 shows the spikes of correlogram, auto correlation, and partial auto correlation coefficients for the lags 1 to 20 at the first order difference of the time series, that is, *BSE_CLOSE* and *NSE_CLOSE*.

The plots say that the first order difference of the data after transformation is random. If the model is fit, then the residuals of the model would contain the sequence of probable errors. Since spikes of ACFs and PACFs are insignificant, the residuals of the chosen ARIMA model are white noise, and hence, the time series data has become stationary. This is essentially a random walk process and there is no need to think about any other AR(p) and MA(q) models further. Hence, the transformed time series essentially follows an ARIMA (0,1,0) process. The random walk model in stock price and market index forecasting has been commonly used and studied throughout history (Fama, 1965). The random walk model has similar implications as the efficient market hypothesis, suggesting that one cannot outperform the market by analyzing historical prices of a certain stock or index of the overall market.

Fig No.2

Correlogram of BSE_CLOSE and NSE_CLOSE

Date:06/22/18 Time:15:08
Sample: 6/08/2014 6/03/2018
Included observations: 209

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.971	0.971	199.72	0.000
		2	0.941	-0.022	388.24	0.000
		3	0.914	0.036	567.00	0.000
		4	0.892	0.076	738.22	0.000
		5	0.866	-0.087	900.38	0.000
		6	0.843	0.039	1054.6	0.000
		7	0.818	-0.022	1200.8	0.000
		8	0.796	0.008	1339.8	0.000
		9	0.778	0.073	1473.2	0.000
		10	0.764	0.055	1602.5	0.000
		11	0.751	0.025	1728.1	0.000
		12	0.737	-0.011	1849.6	0.000
		13	0.719	-0.069	1965.9	0.000
		14	0.701	-0.012	2077.1	0.000
		15	0.679	-0.083	2182.1	0.000
		16	0.658	-0.017	2280.9	0.000
		17	0.637	0.019	2374.2	0.000
		18	0.616	-0.030	2461.8	0.000
		19	0.593	-0.015	2543.5	0.000
		20	0.567	-0.092	2618.4	0.000

Date:06/22/18 Time:15:13
Sample: 6/01/2014 6/03/2018
Included observations: 210

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.974	0.974	202.12	0.000
		2	0.948	-0.011	394.61	0.000
		3	0.924	0.022	578.36	0.000
		4	0.904	0.055	754.89	0.000
		5	0.882	-0.029	923.93	0.000
		6	0.860	-0.024	1085.3	0.000
		7	0.837	-0.014	1239.0	0.000
		8	0.816	0.020	1385.9	0.000
		9	0.799	0.051	1527.2	0.000
		10	0.783	0.033	1663.8	0.000
		11	0.770	0.039	1796.3	0.000
		12	0.756	-0.010	1924.7	0.000
		13	0.737	-0.087	2047.5	0.000
		14	0.720	0.014	2165.2	0.000
		15	0.699	-0.089	2276.8	0.000
		16	0.679	-0.000	2382.7	0.000
		17	0.660	-0.002	2483.1	0.000
		18	0.640	-0.010	2578.0	0.000
		19	0.618	-0.031	2667.1	0.000
		20	0.591	-0.127	2749.1	0.000

Fig No.3

First Order Difference Correlogram of BSE_CLOSE and NSE_CLOSE

Date:06/22/18 Time:22:58
Sample: 6/08/2014 6/03/2018
Included observations: 208

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.024	0.024	0.1238	0.725
		2	-0.075	-0.075	1.3107	0.519
		3	-0.128	-0.125	4.7840	0.188
		4	0.002	0.002	4.7851	0.310
		5	0.020	0.001	4.8685	0.432
		6	0.034	0.018	5.1165	0.529
		7	-0.041	-0.040	5.4775	0.602
		8	-0.135	-0.131	9.4871	0.303
		9	-0.037	-0.034	9.7861	0.368
		10	0.055	0.029	10.464	0.401
		11	0.091	0.055	12.300	0.342
		12	0.079	0.078	13.684	0.321
		13	-0.015	0.006	13.732	0.393
		14	0.034	0.068	13.990	0.450
		15	-0.045	-0.039	14.444	0.492
		16	-0.035	-0.053	14.728	0.545
		17	-0.002	-0.003	14.728	0.615
		18	-0.100	-0.110	17.015	0.522
		19	0.061	0.086	17.888	0.530
		20	-0.068	-0.069	18.975	0.523

Date:06/22/18 Time:22:54
Sample: 6/01/2014 6/03/2018
Included observations: 209

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.004	0.004	0.0036	0.952
		2	-0.052	-0.052	0.5802	0.748
		3	-0.143	-0.142	4.9294	0.177
		4	0.004	0.002	4.9336	0.294
		5	0.031	0.016	5.1390	0.399
		6	0.019	-0.001	5.2138	0.517
		7	-0.002	0.002	5.2146	0.634
		8	-0.148	-0.143	9.9965	0.265
		9	-0.040	-0.040	10.347	0.323
		10	0.048	0.035	10.862	0.368
		11	0.089	0.048	12.640	0.317
		12	0.094	0.093	14.619	0.263
		13	-0.036	-0.013	14.908	0.313
		14	0.049	0.081	15.458	0.348
		15	-0.052	-0.032	16.085	0.376
		16	-0.031	-0.059	16.307	0.432
		17	-0.016	-0.016	16.363	0.498
		18	-0.084	-0.098	17.996	0.456
		19	0.075	0.090	19.317	0.437
		20	-0.088	-0.076	21.142	0.389

(3) ARIMA Model for Forecasting : The ARIMA model is the composition of a series of steps for discovering the best model, supposing and identifying the different (ARIMA) models using available data in time series, and forecasting the series using the best model. It is one of the well-known techniques for economic forecasting. ARIMA models are extremely capable to produce projections during short-term (Merh, Saxena, & Pardasani, 2010). These are the best composite structural models useful for short-term forecasts (Pai & Lin, 2005). In ARIMA model, the expected value of any variable is a linear combination of past values and errors (Hanke & Wichern, 2005), expressed as follows :

(i) Auto Regressive Model [AR(*p*)] : An AR model is one in which 'Y_t' depends only on its own past values, that is,

$$Y_{t-1}, Y_{t-2}, Y_{t-3}, \text{etc. Thus, } Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots\ldots, \varepsilon_t) \quad \ldots\ldots\ldots\ldots (13)$$

A common representation of an autoregressive model where it depends on '*p*' of its past values called AR(*p*) model is represented below:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \ldots\ldots + \beta_p Y_{t-p} + \varepsilon_t \quad \ldots\ldots\ldots\ldots (14)$$

where Y_t = affecting (dependent) variable at time *t*.

$Y_{t-1}, Y_{t-2}, \ldots\ldots, Y_{t-p}$ = Response variable at time lags *t*-1, *t*-2, ..., *t-p*, respectively.

$\beta_0, \beta_1, \beta_2, \ldots\ldots, \beta_p$ = Coefficients to be estimated.

ε_t = Error term at time *t*.

(ii) Moving Average Model [MA(*q*)] : A moving average model is one when Y_t depends only on the random error terms which follow a white noise process, that is,

$$Y_t = f(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots\ldots) \quad \ldots\ldots\ldots\ldots (15)$$

A common representation of a moving average model where it depends on '*q*' of its past values is called MA(*q*) model and is represented below :

$$Y_t = \beta_0 + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \phi_3 \varepsilon_{t-3} + \ldots\ldots + \phi_q \varepsilon_{t-q} \quad \ldots\ldots\ldots\ldots (16)$$

The error terms ε_t is assumed to be white noise processes with mean zero and variance σ^2.
where ,

Y_t = Response variable (dependent) variable at time *t*,

β_0, = Constant mean of the process,

$\phi_1, \phi_2, \phi_3, \ldots\ldots, \phi_q$ = coefficients to be estimated,

ε_t = Error term at time *t*.

$\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3} \ldots \varepsilon_{t-q}$ = Errors in previous time periods that are incorporated in Y_t.

(iii) Auto Regressive Moving Average (ARMA) Model : There are situations where the time series may be represented as a mix of both AR and MA models referred to as ARMA (*p*,*q*). The general form of such a time - series model, which depends on '*p*' of its own past value and '*q*' past values of white noise disturbances take the form :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + ... + \phi_q \varepsilon Y_{t-q} \qquad ....(17)$$

(iv) Selection of Appropriate ARIMA (*p,d,q*) Model : Model for non - seasonal series is called autoregressive integrated moving average model denoted by ARIMA (*p,d,q*). Here '*p*' is the order of autoregressive part, '*d*' indicates the order of differencing, and '*q*' indicates the order of moving average part. In general, a series which is stationary after being differenced '*d* times' is said to be integrated of order '*d*,' denoted by *I(d)*. If the original series is stationary, $d = 0$ and the ARIMA models reduce to ARMA models. The time series data used for the present study, that is, *BSE_CLOSE* and *NSE_CLOSE* have become stationary after the first order differencing. Since there is no need for further differencing the series, it is necessary to adopt $d=1$ (first difference) for ARIMA (*p, d ,q*) model. To get the appropriate numbers for '*p*' (in AR) and '*q*' (in MA) in the model, we should check the correlogram after first difference in time series (Figure 2). Since there are no significant spikes of ACF and PACF, the residuals of the selected ARIMA model are white noise and there is no need for further consideration of one more AR (*p*) and MA (*q*). To choose one best ARIMA model amongst the numerous combinations present, the following criterions are used :

(a) Comparatively low of Akaike/Bayesian/Schwarz Information criterions (AIC/BIC/SIC).

(b) Comparatively low S.E. of Regression.

(c) Comparatively high adjusted *R* - square (R^2).

(d) Root mean square error (RMSE) should be relatively low.

(e) Mean absolute error (MAE) and mean absolute percentage error (MAPE) should be low.

Table 3 and Table 4 provide the results of various parameters of AR(*p*) and MA(*q*) of the ARIMA model. Using these values, the best fit model for predicting the time series *DBSE_CLOSE* and *DNSE_CLOSE* are identified.

After checking the robustness of the statistics given in the Table 3 and Table 4, it is found that only ARIMA (0,1,0) model convinces all the norms (lowest AIC, BIC, RMSE, MAE, MAPE, standard error of regression, and the relatively high adjusted R^2 values). Hence, this model is considered to be the best predictive model, which is used to forecast the future values of the time series, that is, *BSE_CLOSE* and *NSE_CLOSE*. The prediction equation for this model can be written as :

Table 3. Output for Various ARIMA Parameters for *DBSE_CLOSE*

ARIMA	RMSE	MAE	MAPE	S.E. of Regression	Log Likelihood	Adjusted R^2	AIC	BIC
(0,1,0)	**518.9344**	**407.0126**	**135.6836**	**520.1864**	**-1595.509**	**0.00000**	**15.3510**	**15.36709**
(1,1,0)	520.0480	408.1295	135.3038	522.5642	-1595.448	-0.00916	15.3696	15.41783
(1,1,1)	520.0340	408.1138	135.1785	523.6914	-1595.388	-0.01352	15.3787	15.44291
(2,1,0)	521.2446	409.8329	131.5123	521.2495	-1594.929	-0.00409	15.3647	15.41284
(2,1,1)	521.2519	409.8372	131.5177	522.4525	-1594.900	-0.00873	15.3740	15.43822
(1,1,2)	519.9580	408.0805	135.3919	522.4621	-1594.904	-0.00876	15.3740	15.43826
(2,1,2)	521.2424	409.8305	131.5162	522.5242	-1594.929	-0.00900	15.3743	15.43850

Note. The values in the first row represent the best ARIMA model among different combinations.

Table 4. Output for Various ARIMA Parameters for *DNSE_CLOSE*

ARIMA	RMSE	MAE	MAPE	S.E. of Regression	Log Likelihood	Adjusted R^2	AIC	BIC
(0,1,0)	**158.0640**	**125.1574**	**195.6765**	**158.4436**	**-1354.725**	**0.00000**	**12.9734**	**12.9894**
(1,1,0)	158.3947	125.4862	195.9553	159.2095	-1354.723	-0.00969	12.9925	13.0405
(1,1,1)	158.3945	125.4854	195.9649	159.5926	-1354.717	-0.01455	13.0028	13.0660
(2,1,0)	158.7299	125.8748	196.6628	158.9947	-1354.444	-0.00696	12.9898	13.0378
(2,1,1)	158.7298	125.8755	196.6935	159.3812	-1354.443	-0.01187	12.9994	13.0634
(1,1,2)	158.3971	125.5027	196.4437	159.3827	-1354.445	-0.01189	12.9994	13.0634
(2,1,2)	158.7296	125.8748	196.6631	159.3818	-1354.444	-0.01187	12.9994	13.0634

Note. The values in the first row represent the best ARIMA model among different combinations.

$$Y_t - Y_{t-1} = \mu_t \text{ or equivalently } Y_t = Y_{t-1} + \mu_t \quad \text{................ (18)}$$

...... where the constant term is the average period - to - period change (i.e., the long-term drift) in 'Y'. This model could be fitted as a no - intercept regression model in which the first difference of 'Y' is the dependent variable.

(4) Forecasting Using Selected ARIMA (*p, d, q*) Model : The present study is based on weekly data on the closing indices of BSE *(BSE_CLOSE)* and NSE (*NSE_CLOSE*) covering the period from June 8, 2014 to June 3, 2018, having a total number of 209 observations, of which the period from January 7, 2018 to June 3, 2018 having 22 observations are used for forecasting length.

(i) Results of ARIMA (0,1,0) Model for *BSE_CLOSE* Prediction : The Table 5 exhibits the forecasting results of ARIMA (0,1,0) model, which is regarded as the best fit model for prediction of *BSE_CLOSE* index. The table shows the actual and predicted values of series for the forecast length (22 observations) ranging from January 7, 2018 to June 3, 2018. It is observed from the summary of ARIMA forecasting model that the software has selected the log of dependent variable after first differencing, that is, *DLOG(BSE_CLOSE)* and the forecast length is 22 weeks. The software has estimated nine models and out of them, the best ARMA model selected is (0,0)(0,0). The value of Akaike information criterion of this model is observed to be very small than all other models tested.

The Figure 4 is the graphical illustration and shows the level of accuracy of the selected ARIMA model, which exhibits the predicted performance of the BSE *(BSE_CLOSE)* against the actual performance during the forecasted period. The line of forecasting of *BSE_CLOSE* continues to rise during the forecasting period, that is, from January 7, 2018 to June 3, 2018. When compared to the forecasted performance, the actual performance of *BSE_CLOSE* during the period from February 4, 2018 to March 18, 2018 is quite unsatisfactory. However, the market revived by the end of June 3, 2018.

According to Table 6, ARIMA (0,1,0) is relatively the best model. The model returns the smallest Akaike information criterion of -5.06570, smallest Bayesian or Schwarz information criterion of -5.04836, and relatively smallest standard error of regression of 0.019169. It is also observed from the model selection criteria table (Table 7) that out of nine models verified, ARMA (0,0)(0,0) is found to be the best model as its LogL, AIC, BIC, and HQ coefficients are smaller than the remaining eight models.

(ii) Results of ARIMA (0,1,0) Model for *NSE_CLOSE* Prediction : The Table 8 contains the empirical results of ARIMA (0,1,0), which is regarded as the best fit model for prediction of *NSE_CLOSE* index. The Table shows the

Table 5. Sample Empirical Results of ARIMA (0, 1, 0) of *BSE_CLOSE*

Sample Period	Actual Values	Predicted Values	Summary of the ARIMA Forecasting Model
31st Dec, 17	34153.85	34153.85	Automatic ARIMA Forecasting
7th Jan, 18	34592.39	34209.52	Selected dependent variable: *DLOG(BSE_CLOSE)*
14th Jan, 18	35511.57	34265.28	Date: 07/01/18 Time: 21:56
21st Jan, 18	36050.44	34321.13	Included observations: 186
28th Jan, 18	35066.75	34377.07	Forecast length: 22
4th Feb, 18	34005.76	34433.10	Number of estimated ARMA models: 9
11th Feb, 18	34010.76	34489.22	Number of non-converged estimations: 0
18th Feb, 18	34142.14	34545.43	Selected ARMA model: (0,0) (0,0)
25th Feb,18	34046.94	34601.74	AIC value: -5.05495260101
4th Mar, 18	33307.14	34658.14	
11th Mar, 18	33176.00	34714.63	
18th Mar, 18	32596.53	34771.21	
25th Mar, 18	32968.67	34827.88	
1st Apr, 18	33626.96	34884.65	
8th Apr, 18	34192.64	34941.50	
15th Apr, 18	34415.57	34998.46	
22nd Apr, 18	34969.69	35055.50	
29th Apr, 18	34915.37	35112.64	
6th May, 18	35535.78	35169.87	
13th May, 18	34848.30	35227.19	
20th May, 18	34924.87	35284.61	
27th May, 18	35227.26	35342.12	
3rd June, 18	35443.67	35339.72	

Figure 4. Actual and Forecast Graph of ARIMA (0,1,0) Model

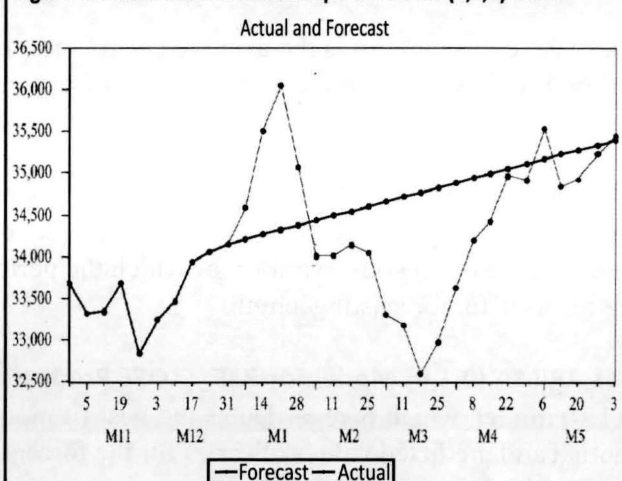


Table 6. ARIMA (0,1,0) Estimation Output with *DLOG(BSE_CLOSE)*

Dep. Variable: *DLOG(BSE_CLOSE)*

Method: Least Squares

Date: 07/01/18 Time : 21:56

Sample (Adjusted): 6/15/2014 12/31/2017

Included observations: 186 after adjustments

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	0.001629	0.001406	1.158693	0.2481
R-squared	0.000000	Mean Dependent variable		0.00162
Adj. *R*-squared	0.000000	*S.D.* dependent variable		0.01916
S.E. of Regression	0.019169	Akaike info criterion		-5.06570
Sum squared residuals	0.067977	Schwarz criterion		-5.04836
Log likelihood	472.1106	Hannan-Quinn criterion		-5.05867
Durbib-Watson Stat	2.044222			

Table 7. Model Selection Criteria Table

Dependent Variable: *DLOG(BSE_CLOSE)*

Date: 07/01/18 Time: 21:56

Sample: 6/08/2014 12/31/2017

Included observations: 186

Model	LogL	AIC*	BIC	HQ
(0,0)(0,0)	472.11059	-5.05495	-5.02026	-5.04089
(0,1)(0,0)	472.16480	-5.04478	-4.99275	-5.02369
(1,0)(0,0)	472.11059	-5.05495	-5.02026	-5.04089
(1,1)(0,0)	473.02099	-5.04323	-4.97386	-5.01512
(0,2)(0,0)	472.69304	-5.03971	-4.97033	-5.01159
(2,0)(0,0)	472.57814	-5.03847	-4.96910	-5.01036
(2,2)(0,0)	474.55985	-5.03827	-4.93422	-4.99611
(2,1)(0,0)	473.47053	-5.03731	-4.95060	-5.00217
(1,2)(0,0)	473.38908	-5.03644	-4.94972	-5.00130

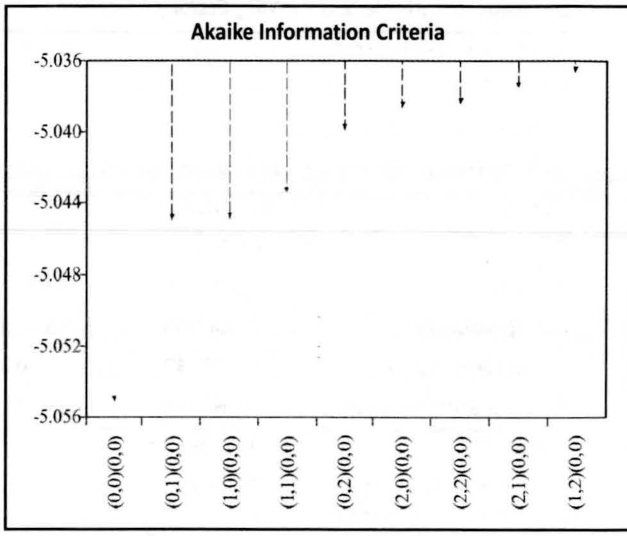


Table 8. Sample Experimental Results of ARIMA (0,1,0) of *NSE_CLOSE*

Sample Period	Actual Values	Predicted Values	Summary of the ARIMA Forecasting Model
31st Dec 17	10558.85	10558.85	Automatic ARIMA Forecasting
7th Jan 18	10681.25	10577.56	Selected dependent variable: *DLOG(NSE_CLOSE)*
14th Jan 18	10894.70	10596.30	Date: 07/01/18 Time: 23:01
21st Jan 18	11069.65	10615.07	Included observations: 187
28th Jan 18	10760.60	10633.87	Forecast length: 22
4th Feb 18	10454.95	10652.71	Number of estimated ARMA models: 9
11th Feb 18	10452.30	10671.59	Number of non-converged estimations: 0
18th Feb 18	10491.05	10690.49	Selected ARMA model: (0,0)(0,0)
25th Feb 18	10458.35	10709.43	AIC value: -5.04795355019
4th Mar 18	10226.85	10728.41	
11th Mar 18	10195.15	10747.41	
18th Mar 18	9998.05	10766.45	
25th Mar 18	10113.70	10785.53	
1st Apr 18	10331.60	10804.64	
8th Apr 18	10480.60	10823.78	
15th Apr 18	10564.05	10842.95	
22nd Apr 18	10692.30	10862.16	
29th Apr 18	10618.25	10881.41	
6th May 18	10806.50	10900.68	
13th May 18	10596.40	10920.00	
20th May18	10605.15	10939.34	
27th May18	10696.20	10958.72	
3rd June 18	10767.65	10978.14	

Figure 5. Actual and Forecast Graph of ARIMA (0,1,0) Model

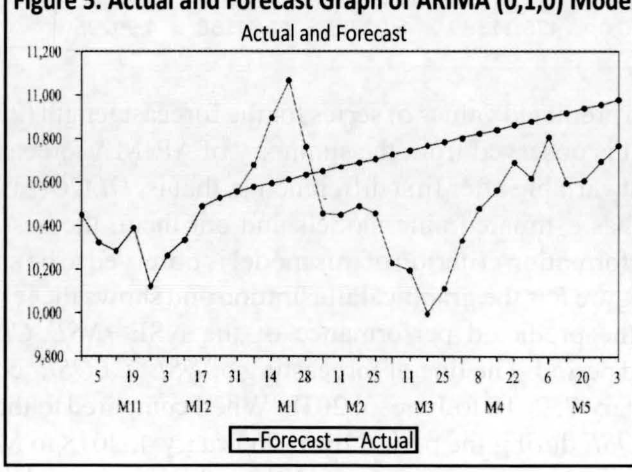


Table 9. ARIMA (0,1,0) Estimation Output with *DLOG(NSE_CLOSE)*

Dependent Variable: *DLOG(NSE_CLOSE)*

Method: Least Squares

Date: 07/01/18 Time:23:13

Sample (Adjusted): 6/08/2014 12/31/2017

Included observations: 187 after adjustments

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	0.001770	0.001407	1.258281	0.2099
R-squared	0.000000	Mean Dependent variable		0.00177
Adj. *R*-squared	0.000000	*S.D.* dependent variable		0.01923
S.E. of Regression	0.019237	Akaike info criterion		-5.05864
Sum squared residuals.	0.068830	Schwarz criterion		-5.04137
Log likelihood	473.9837	Hannan-Quinn criterion.		-5.05164
Durbin-Watson Stat	2.066195			

Table 10. Model Selection Criteria Table

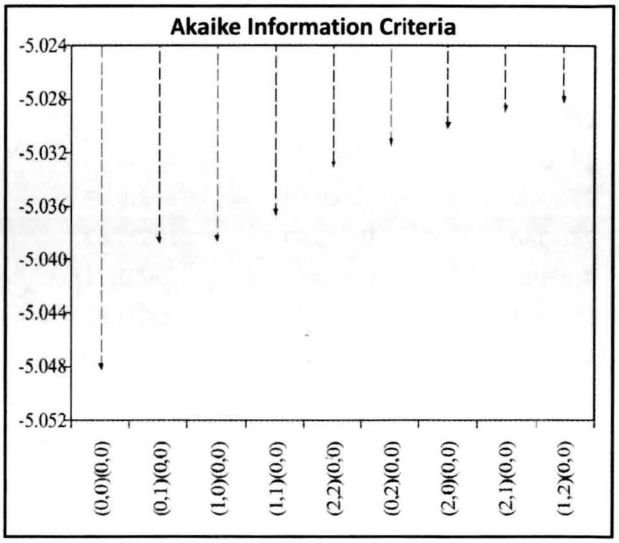
Dependent Variable: *DLOG(NSE_CLOSE)*

Date: 07/01/18 Time: 23:13

Sample: 6/08/2014 12/31/2017

Included observations: 187

Model	LogL	AIC*	BIC	HQ
(0,0)(0,0)	473.98365	-5.04795	-5.01339	-5.03395
(0,1)(0,0)	474.10016	-5.03850	-4.98666	-5.01750
(1,0)(0,0)	474.08806	-5.03837	-4.98653	-5.01737
(1,1)(0,0)	474.89945	-5.03635	-4.96724	-5.00835
(0,2)(0,0)	474.41425	-5.03116	-4.96205	-5.00316
(2,0)(0,0)	474.29483	-5.02989	-4.96077	-5.00188
(2,2)(0,0)	476.56655	-5.03279	-4.92912	-4.99078
(2,1)(0,0)	475.17577	-5.02861	-4.94222	-4.99361
(1,2)(0,0)	475.11389	-5.02795	-4.94156	-4.99295



actual and predicted values of series for the forecast length (22 observations) ranging from January 7, 2018 to June 3, 2018. It is observed from the summary of ARIMA forecasting model that the software has selected the log of dependent variable after first differencing, that is, *DLOG(NSE_CLOSE)* and the forecast length is 22 weeks. The software has estimated nine models and out them, the best ARMA model selected is (0,0)(0,0). The value of Akaike information criterion of this model is observed to be smaller than all other models tested.

The Figure 5 is the graphical illustration and shows the level of accuracy of the selected ARIMA model, which exhibits the predicted performance of the NSE *(NSE_CLOSE)* against the actual performance during the forecasted period. The line of forecasting of *NSE_CLOSE* continues to rise during the forecasting period, that is, from January 7, 2018 to June 3, 2018. When compared to the forecasted performance, the actual performance of *NSE_CLOSE* during the period from February 4, 2018 to March 18, 2018 is quite unsatisfactory. However, the market revived by the end of May 6, 2018, which continued till June 3, 2018.

According to Table 9, ARIMA (0,1,0) is relatively the best model. The model gives a very small AIC of -5.05864, small BIC/SIC of -5.04137, and comparatively negligible value of S.E. of regression of 0.019237. It is also observed from the model selection criteria table (Table 10) that out of nine models verified, ARMA(0,0)(0,0) is found to be the best model as its LogL, AIC, BIC, and HQ coefficients are smaller than the remaining eight models.

Conclusion

The main objective of this paper is to study the stationarity of the indices of BSE and NSE and to forecast using the ARIMA model. For this purpose, the weekly closing indices of BSE and NSE are obtained from the website yahoofinance.com for the period from June 6, 2014 to June 3, 2018. The ADF test is administered to check for the presence of unit root to confirm the stationarity of index series. The results of the test confirm the presence of unit root and show non-stationarity. The ADF test has confirmed that the given time series are stationary at first difference.

For the present work, ARIMA (0,1,0) model is chosen as the top model from nine different models because it gratifies all the norms of goodness of fit statistics as the other eight models have not satisfied such criteria. This best candidate model is selected for making predictions of *BSE_CLOSE* and *NSE_CLOSE* for the period ranging from January 7, 2018 to June 3, 2018 using the weekly data ranging from January 6, 2014 to December 31, 2017. The study also makes a comparison between predicted and actual performance of *BSE_CLOSE* and *NSE_CLOSE* during the sample period. The results of the best fitted model highlight the strength of ARIMA model to forecast the *BSE_CLOSE* and *NSE_CLOSE* satisfactorily on a short-term basis and would guide the individuals to select gainful investment options.

Research Implications

The findings of the study have the following implications for investors, researchers, and the academic fraternity.

(1) The study has elucidated the procedure for testing the stationarity in time series data using the Augmented Dicky - Fuller test and correlogram analysis. The study has enlightened the criterion and modus operandi for selection of the best ARIMA model and the methodology for forecasting *BSE_CLOSE* and *NSE_CLOSE*. This will aid the researchers and academicians to carry out further research.

(2) The forecasting of market indices (*BSE_CLOSE* and *NSE_CLOSE*) and comparision of forecast and actual performance will assist the investors to know the market trends, risk analysis, and to take investment decisions.

Limitations of the Study and Scope for Further Research

The ARIMA model has few constraints regarding the exactness of forecasting because of its wide usage for short-run predicting the values in the time series to notice the minor variations in the data. In case of erratic variations in the data set (too large variations) due to change in government policies or the structural breaks in economy (economic instability) etc., it turns out to be intricate to capture the accurate trend. Hence, this model turns out to be useless to predict long-run changes. Moreover, the forecasting using the ARIMA model would depend upon the hypothesis of linearity in historical data, however, there is no confirmation that *BSE_CLOSE* or *NSE_CLOSE* are linear in nature.

Forecasting of *BSE_CLOSE* and *NSE_CLOSE* using ARIMA model was made with the fundamental supposition that the given series follow an absolutely linear model. Non linear prediction methods using latest softwares may also be considered with less error (white noise) term. Further, the study may be extended to multivariate time series forecasting, that is, predicting a dependent variable using more than one independent varaibles. In future, similar studies may be conducted for forecasting various economic variables, that is, gold and silver prices, curency exchange rates, individual stock prices, production from agriculture and industry, electricity consumption, export performance of various industries, etc.

References

Abdullah, L. (2012). ARIMA model for gold bullion coin selling price forecasting. *International Journal of Advances in Applied Sciences, 1*(4), 153 - 158.

Adebiyi, A.A., & Adewumi, O.A. (2014). *Stock price prediction using the ARIMA model.* 2014 UK Sim - AMSS 16th International Conference on Computer Modelling and Simulation (pp.105 - 111). Durban, South Africa: IEEE.

Banerjee, D. (2014). *Forecasting of Indian stock market using time-series ARIMA model.* 2014 2nd International Conference on Business and Information Management (ICBIM). Retrieved from https://ieeexplore.ieee.org/abstract/document/6970973/

Box, G.E.P., & Jenkins, G.M. (1976). *Time series analysis, forecasting and control.* San Francisco : Holden - Day.

Brooks, C. (2008). *Introductory econometrics for finance.* New York, USA : Cambridge University Press.

Contreras, J., Espinola, R., Nogales, F.J., & Conejo, A. J. (2003). ARIMA models to predict next day electricity prices. *IEEE Transaction Power Systems, 18*(3), 1014 - 1020.

Dikshita, & Singh, H. (2019). Estimating and forecasting volatility using ARIMA model: A study on NSE, India. *Indian Journal of Finance, 13*(5), 37 - 51. doi:10.17010/ijf/2019/v13i5/144184

Draxler, T. C. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Argument against avoiding RMSE in the literature. *Geoscientific Model Development, 7*(3), 1247 - 1250.

Fama, F. E. (1965). Random walks in stock market prices. *Financial Analysts Journal, 21*(5), 55 - 59.

Gerakos, J. J., & Gramacy, R. (2013). *Regression - based earnings forecasts.* DOI : http://dx.doi.org/10.2139/ssrn.2112137

Gideon, S. (1978). Estimation of dimensions of a model. *Annals of Statistics, 6*(2), 461 - 464.

Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Science, 4*(2), 117 - 121.

Gujarati, D. N. (2004). *Basic econometrics* (4th ed.). New Delhi : The McGraw Hill Company.

Hanke, E. J., & Wichern, W. D. (2005). *Business forecasting* (8th ed.). USA: Pearson Education Inc.

Jadhav, S., Kakade, S., Utpat, K., & Deshpande, H. (2015). Indian share market forecasting with ARIMA model. *International Journal of Advanced Research in Computer and Communication Engineering, 4*(11), 334 - 336.

Jarrett, J. E., & Kyper, E. (2011). ARIMA modeling with intervention to forecast and analyze Chinese stock prices. *International Journal of Engineering, Business and Management, 3*(3), 53 - 58.

Merh, N., Saxena, V. P., & Pardasani, K. R. (2010). A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting. *Journal of Business Intelligence, 3*(2), 23 - 43.

Meyler, A., Kenny, G., & Quinn, T. (1998). *Forecasting Irish inflation using ARIMA models (Technical Paper No. 3/RT/98).* Retrieved from https://mpra.ub.uni-muenchen.de/11359/

Nochai, R., & Tidia, N. (2006). ARIMA model for forecasting oil palm price. *Proceedings of the 2nd IMT- GT Regional Conference on Mathematics, Statistics and Applications* (pp. 1-7). Universiti Sains Malaysia, Penang.

Pai, P. - F., & Lin, C. - S. (2005). A hybrid ARIMA and support vector machines model in stock price prediction. *Omega, 33*, 497 - 505.

Rao, N. K., & Mukherjee, K. (1971). Random walk hypothesis: An empirical study. *Arthaniti, 4*(1), 53 - 59.

Savadatti, P. (2017). Forecasting of area, production, and productivity of food grains in India: Application of ARIMA model. *Arthashastra Indian Journal of Economics and Research, 6*(6), 7 - 22. doi: 10.17010/aijer/2017/v6i6/120114

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461 - 464. doi:10.1214/aos/1176344136

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of Operations Research Society, 66*(8), 1352 - 1362.

Tsay, S. R. (2005). *Analysis of financial time series* (2nd ed.). Wiley series in probability and statistics. Chicago : John Wiley & Sons, Inc.

About the Author

Dr. C. Viswanatha Reddy is the Head & Chairman, BoS, Department of Business Management, Rayalaseema University, Kurnool, Andhra Pradesh. He has 38 research papers, 7 books to his credit, and has organized two national level seminars and one capacity building programme.