Global and Stochastic Analysis
Vol. 4 No. 1, January (2017), 77-81

MUK PUBLICATIONS
Open Access Publisher

FORECASTING NATIONAL STOCK PRICE USING ARIMA MODEL

MOHAMED ASHIK. A* AND SENTHAMARAI KANNAN. K

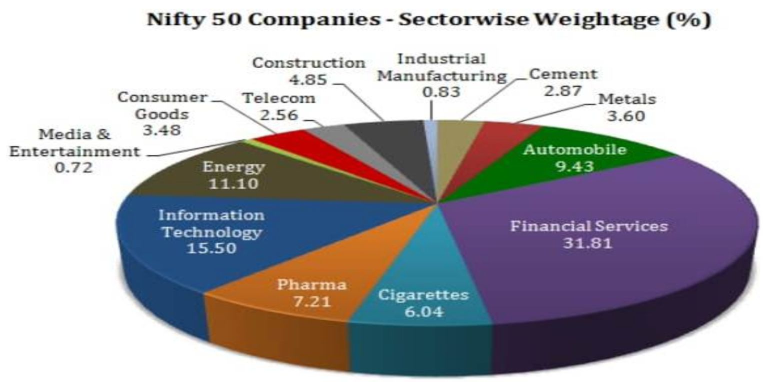
Abstract. Stock exchange deals in securities like shares or bonds issued by the companies or corporations in the private and public sectors. National Stock Exchange is widest and full automatic trading system in India. Nifty 50 is one of the stock price investor and 50 companies were invested in the traders. Autoregressive Integrated Moving Average model is one of the most accepted forecasting models and a vital area of the Box-Jenkins approach to time series modeling. In this paper, the Nifty 50 stock market prices were evaluated and predicted the trend of upcoming trading days stock market fluctuations using Box-Jenkins methodology. From the results, it can be observed that influence R-Square value is (94%) high and Mean Absolute Percentage Error is very small for the fitted model. Thus the prediction accuracy is more suitable of Nifty 50 closing stock price. It is concluded that closing stock price of Nifty 50 taken in the present study shows slow decreasing fluctuations trend for upcoming trading days.

1. Introduction

The National Stock Exchange (NSE) is a stock market in India, which is set up on November 1992. NSE was the first exchange in the country to provide a modern, fully automated screen-based electronic trading system which accessible easy trading facility to the investors widen across the length and breadth of the country. NSE was set up by a group of leading Indian financial institutions at the request of the government of India to bring precision to the Indian capital market. Based on the recommendation lay out in the government committee, It has been established with a diversified shareholding comprise domestic and universal investors. NSE was also instrumental in create the National Securities Depository Limited (NSDL) which allows investors to securely embrace and transfer their shares and bonds electronically. It also allows investor to hold and trade in as few as one share or bond. This not only made holding financial instrument convenient, but more importantly eliminated the need for paper certificates and really reduced the incidents of fake certificates and falsified transactions of Indian stock market. The NSDL combined with the transparency, lower transaction prices and effectiveness that NSE offered, significantly increased the attractiveness of the Indian stock market to overall investors.

2000 *Mathematics Subject Classification.* Primary 37M10,62M10; Secondary 60G10,60G25.
Key words and phrases. National Stock Exchange, Nifty 50, Autocorrelation, Bayesian Information Criteria, Autoregressive Integrated Moving Average Model.
* Corresponding Author.

Similarly, The Nifty 50 is an indicator of the top 50 major companies on the NSE. If the Nifty 50 goes upward, it means that most of the stocks in India went up during the period and the Nifty 50 goes downward, that the stock price of most of the major stocks on down.



A lot of method has been used for NSE, including autoregressive model (AR), autoregressive moving average model (ARMA), autoregressive integrated moving average model (ARIMA) and so on. In the steadiness of forecast and explanation, ARIMA is a widely used model. In this paper, the Nifty 50 data of 2015 is firstly examined and then ARIMA model is used to fit the data.
Stock market price are most concerned about the stock opening price, lowest price, highest price, closing price, adjusted closing price and volume. In technical analysis, the highest and lowest price represents the inclusive struggle among multi outer forces. The volume represents the market trade activity, and the closing price is on behalf of the balance from the multi contest, which can be seen as the opening price of the next trading day. In nature, a trading day closing price is not only associated with the previous trading day closing price. In this paper, the Nifty 50 closing stock price (in Rs.) data was used. Then the time period is January to December of 2015, with 245 observations. The Data is obtained from the financial part of Yahoo.com and the computations are done by using SPSS 20 software.

2. Review of Related Work

Naylor et al (1972) made more wide and detailed comparison of alternative methods and examined Box-Jenkins approach in compare to econometric model for the year 1963 through 1967. They observed that the accuracy of ARMA models of Box-Jenkins methodology was significantly better than the accuracy of econometric model. Nelson (1972) compared regression and ARMA methods for a longer time horizon. It concluded that the simple ARMA models are relatively stronger with respect to post sample predictions than the complex econometric models. If the mean square error is a fitting measure of loss an unweighted assessment clearly indicated that a decision maker will be better off relying simply on ARMA predictions in the post sample period in the forecasting phase. Leseps and Morell (1977) in their study establish that the stock price follows a long-term trend with short-term fluctuation and forecast the exchange rate. Jawahar Farook and Senthamarai Kannan (2014) used Stochastic Modeling to forecast Carbon Dioxide Emissions for the upcoming months. Banerjee, D. (2014) applied ARIMA model to forecast in Indian Stock Exchange the future stock indices. Paulo Rotela Junior et al. (2014) described ARIMA model to obtain short-term forecasts for the next month in order to minimize prediction errors for the Bovespa Stock Index. Renhao Jin et al. (2015) used ARIMA model to predict in Shanghai Composite Stock Price Index and they are consider to closing stock price.

3. Objective of Study

To forecast the National Stock Exchange closing stock price of Nifty 50 using ARIMA model in Time Series Analysis.

4. Methodology

4.1. Box-Jenkins (ARIMA). In time series analysis, an ARIMA model is a generalization of an ARMA model. These models are fitted to time series data either to better identify with the data or to predict future points in the series. They are applied in many cases where data illustrate evidence of non-stationarity, where as differencing step can be applied to reduce the non-stationarity. Non-seasonal ARIMA models are generally denoted ARIMA (p, d, q) where parameters are non-negative integers then p, d, q refer to the autoregressive, differencing, and moving average terms for the non-seasonal component of the ARIMA model. Seasonal ARIMA models are usually denoted ARIMA (p, d, q) (P, D, Q)$_m$, where m refers to the number of periods in each season, and P,D,Q refer to the autoregressive, differencing, and moving average terms for the seasonal component of the ARIMA model. ARIMA models form an important area of the Box – Jenkins approach to time-series modeling. It is also known as Box-Jenkins method. A non-seasonal stationary can be modeled as a combination of the past values and the errors which can be denoted as ARIMA (p, d, q) are can be expressed as

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

The Box-Jenkins (ARIMA) methodology for analyzing and modeling time series is characterized by four steps:

- Identification
- Estimation
- Diagnostic checking
- Forecast

4.1.1. *Identification.* The identification stage, finding the time series data is stationary or not and compare the estimated Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to find a match. We choose, as a tentative model, the ARMA process whose theoretical ACF and PACF best match the estimated ACF and PACF.

4.1.2. *Estimation.* Estimating the parameters for Box – Jenkins models is a rather complicated non – linear estimation problem. The main approaches for fitting Box - Jenkins models are non linear least squares and maximum likelihood estimation. Parameter estimates are usually obtained by maximum likelihood, which is asymptotically correct for time series. Estimators are always sufficient, efficient, and consistent for Gaussian distribution and which are asymptotically normal with efficient for several non-Gaussian distribution.

4.1.3. *Diagnostic Checking.* The diagnostic checking is necessary to test the appropriateness of the selected model. Model selection can be made based on the values of certain criteria like log likelihood, Akaike Information Criteria (AIC)/ Bayesian Information Criteria (BIC)/ Schwarz-Bayesian Information Criteria (SBC).

$$AIC = \left\{n\left(1+\log 2\pi\right)+n\log\sigma^2+2m\right\}$$

$$BIC = -2\log(L) + k\,\log\left(n\right)$$

$$SBC = \log\sigma^2 + \left(m\log n\right)/n$$

If the model selection is done, it is necessary to verify the satisfactoriness of the estimated model. This is done by studying the pattern among the residuals, if there is any. The estimated residuals can be computed as
$\hat{e} = Y_t - \hat{Y}_t$; Where $\hat{Y}_t$ is the estimated observation at time t.
The values of $\hat{e}_t$, which are either less than -3 or greater than 3, indicate that the corresponding residuals are outliers. The values of ACF may be studied to verify whether the series of residuals is white-noise. After tentative model has been fitted to the data, it is important to perform diagnostic checks to test the satisfactoriness of the model. It has been found that it is effective to measure the overall adequacy of the chosen model by examining a quantity Q known as Ljung-Box statistic whose approximate distribution is chi-square and computed as follows:

$$Q = n\left(n+2\right)\sum_{p=1}^{h}\left(n-k\right)^{-1}r_p^2$$

The Ljung-Box (Q) statistic is compared to critical values from chi-square distribution. While the diagnostic checking is fulfilled effectively and the model is found adequate, the fitted model can be used for forecasting purpose.

4.1.4. *Forecasting.* Forecasting is the prediction of values of a variable based on identified past values of that variable or other associated variables. Forecasting also may be based on expert judgments, which in turn are based on chronological data and experience. In analysis part, the appropriate model is found satisfactory, and the fitted model can be used for forecasting purpose.

5. Empirical Result

5.1. Descriptive Statistics. During the period of Nifty 50 in NSE, there are no outliers observations (Fig 1). The minimum and maximum of stock closing price is 7558.80 and 8996.25 respectively. The Nifty 50 stock closing prices are observed at an average of 8281.52 among a standard deviation of 346.77.

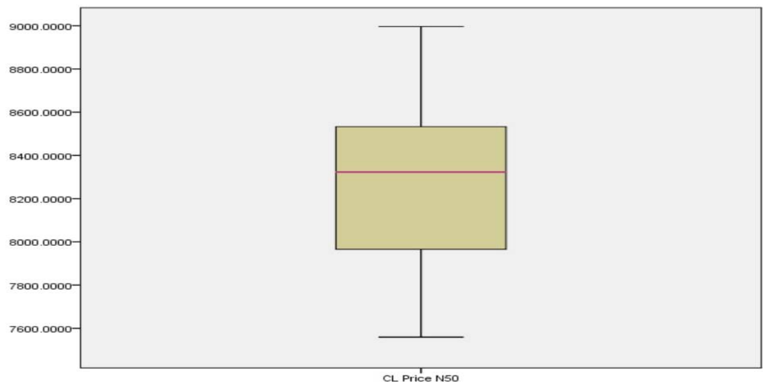


FIGURE 1. Whisker-Box diagram plot of Nifty 50 closing stock price

TABLE 1. Descriptive Statistics of Nifty 50 Closing Stock Price

			Statistic	**Std.Error**
CL Price N50	N		245	
	Mean		8281.52	22.155
	95% Confidence	Lower Bound	8237.87	
	Interval for Mean	Upper Bound	8325.15	
	5% Trimmed Mean		8280.69	
	Median		8323	
	Variance		120255.471	
	Std.Deviation		346.77	
	Minimum		7558.80	
	Maximum		8996.25	
	Range		1437.45	
	Sum		2028971.35	
	Skewness		-.034	.156
	Kurtosis		-.974	.310
Missing Value: Nil				

5.2. Stationary Sequence. A timeplot diagram is firstly used all the NSE Nifty50 data of 2015 based on closing price. As shown in Figure 2, a clear U-shape trend can be found, which is corresponding the Indian economics. The Nifty 50 changes from around 8400 points in the beginning to around 8000 points in the end of year 2015. This fluctuation trend breaks the hypotheses of weaker stationary. In many application cases, the weaker stationary is used instead of strongly stationary.

A weaker form of stationarity commonly engaged in time series is known as second-order stationarity, which only require that 1^{st} moment and auto-covariance do not be different among respect to time. So, for a continuous-time stochastic process x(t), it has the following properties: the mean function E{x(t)} must be constant and the covariance function depends only on the difference between t_1 and $t_2$only desires to be indexed by one variable rather than two variables.

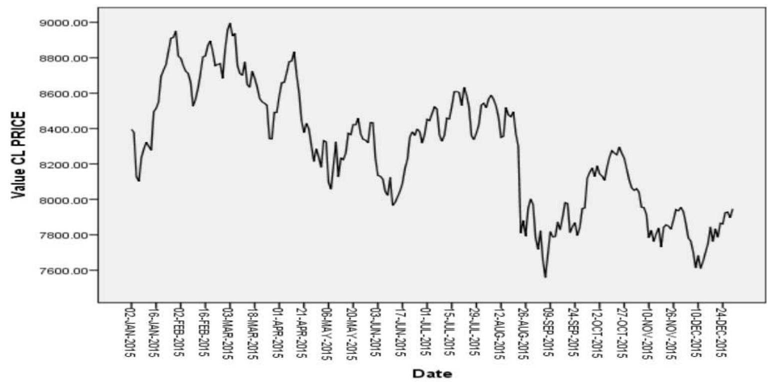


FIGURE 2. Time plot of Nifty 50 closing stock price

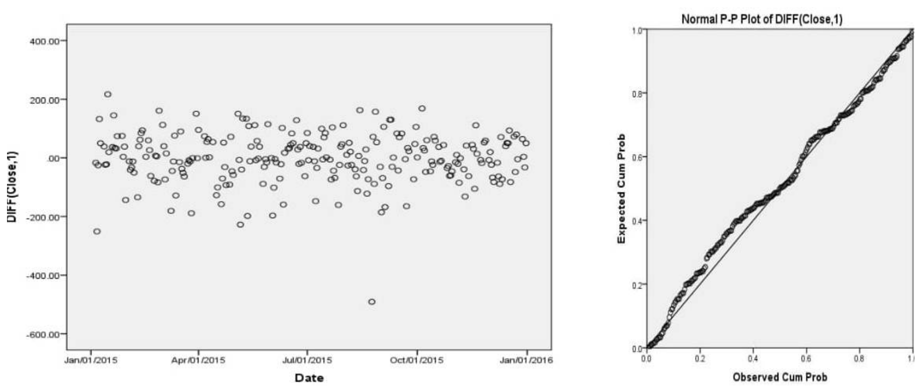


FIGURE 3. Time plot and P-P plot of the first order differencing of Nifty50 closing stock price

Following ARIMA model, a First order differencing is computed for the data after that time plot and P-P plot of the differencing data is shown in Figure 3. The differencing data shows a stationary pattern, and an Autocorrelation (Figure 4) is also done on the differencing data, which displays a short-term autocorrelation and confirms the stationary of the differencing data. To construct an precise inference of the data, autocorrelation check for white noise is done on the differencing data.

5.3. ARIMA Modeling. The basic idea of ARIMA model is to view the data sequence as formed by a Stochastic Process on time. Once the model has been identified, the model can be used to estimate the future value based on the the past and present value of the time series. Modern statistical methods and Econometric models have been able to aid companies predict the future in certain process.
According to the identification rules on time series, the corresponding model can be established. If a partial correlation function of a stationary sequence is truncated, and auto-correlation function is tailed, it can be concluded the sequences for AR

model; if partial correlation function of a stationary sequence is tailed, and the auto-correlation function is truncated, it can be strong that the MA model can be fitted for the sequence. If the partial correlation function of a stationary sequence and the autocorrelation function are tailed, then the ARMA model is appropriate for the sequence. Based on the results from an ARIMA model can be fitted to the original Nifty 50 data of 2015, also the parameters in ARIMA (p,1,q) need to be determined. From the Figure 4, the Autocorrelation and Partial Autocorrelation are safe to the concluded.

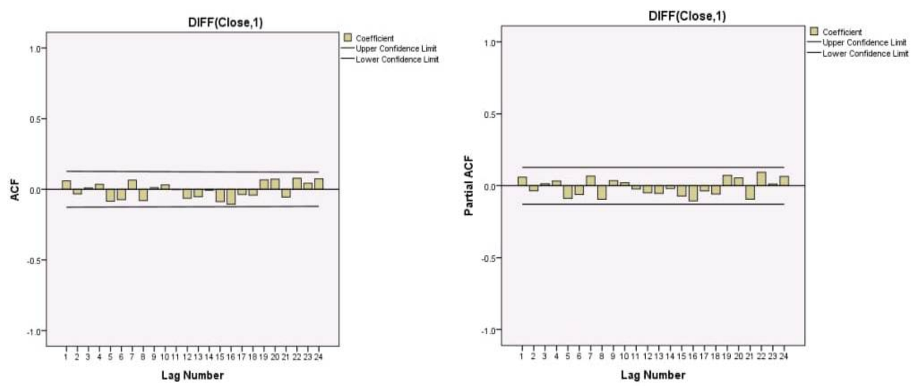


FIGURE 4. ACF and PACF diagram of the first order differencing of Nifty 50 closing stock price

5.4. Fitted ARIMA Model. Plotting of ACF and PACF of (Figure 4) show that the order of p and q can almost be 1. We entertained Nine interim ARIMA models and choose that model which has minimum Normalized BIC. The models and corresponding Normalized BIC values are presented in Table 2. From the Table 2, the most suitable model is ARIMA (0, 1, 1) due to having the lowest Normalized BIC values.

TABLE 2. Normalized BIC value of Nifty 50 closing stock price

ARIMA (p,d,q)	Normalized BIC
ARIMA (0,1,0)	9.453
ARIMA (1,1,0)	9.341
ARIMA (0,1,1)	**8.979**
ARIMA (1,1,1)	9.001
ARIMA (2,1,0)	9.244
ARIMA (2,1,1)	9.026
ARIMA (2,1,2)	9.054
ARIMA (1,1,2)	9.028
ARIMA (0,1,2)	9.019

The model verification is concerned with checking the residuals of the model to observe if they contain any systematic pattern which still can be removed to get better on the chosen ARIMA. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For the various correlations up to 24 lags are computed and the same along with their significance which is tested by Box-Ljung (Q) test are provided in Table 3.

TABLE 3. ACF value of first order differencing of Nifty 50 closing stock price

Lag	**Autocorrelation**	**Std.Error**	**Box-Ljung Statistic**		
			Value	**DF**	**Sig**
1	.059	.064	.857	1	.355
2	-.032	.063	1.107	2	.575
3	.009	.063	1.126	3	.771
4	.035	.063	1.424	4	.840
5	-.084	.063	3.197	5	.670
6	-.073	.063	4.556	6	.602
7	.064	.063	5.604	7	.587
8	-.080	.063	7.232	8	.512
9	.012	.063	7.269	9	.609
10	.030	.062	7.504	10	.677
11	-.003	.062	7.506	11	.757
12	-.063	.062	8.522	12	.743
13	-.052	.062	9.230	13	.755
14	-.006	.062	9.241	14	.815
15	-.087	.062	11.223	15	.737
16	-.106	.062	14.161	16	.587
17	-.036	.061	14.501	17	.631
18	-.041	.061	14.952	18	.665
19	.066	.061	16.132	19	.648
20	.072	.061	17.528	20	.618
21	-.055	.061	18.331	21	.628
22	.078	.061	19.990	22	.584
23	.043	.061	20.500	23	.612
24	.074	.061	21.995	24	.580
Std.Error - Process assumed is independence (white noise)					
Sig - Based on the asymptotic chi-square approximation					

As the results indicate, none of these correlations is notably different from zero at a reasonable level. This proves that the chosen ARIMA model is an appropriate model. The ACF and PACF of the residuals (Figure 5) also indicate 'good fit' of the model. Therefore the fitted ARIMA model for Nifty 50 closing stock price is

$$\hat{Y}_t = C + Y_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$\hat{Y}_t = -1.827 + Y_{t-1} + 0.063\varepsilon_{t-1} + \varepsilon_t$$

The graphical comparison of the actual values and the predict values of Nifty50 is presented. These measures show that the forecasting inaccuracy is low and the forecast graph of Nifty50 closing stock price is given in Figure 6.

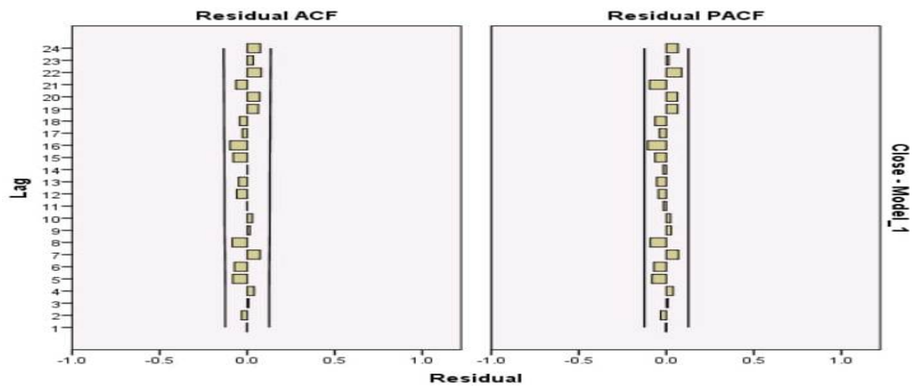


FIGURE 5. Residual ACF and PACF diagram of actual Nifty 50

TABLE 4. Model Parameters of Nifty 50 closing stock price

		Estimate	SE	T	Sig
CL Price N50 **No transformation**	Constant	-1.827	5.783	-.316	.752
	Difference	1			
	MA Lag 1	-.063	.064	-.982	.327

TABLE 5. Model Statistics of Nifty 50 closing stock price

Model	Model Fit				Ljung-Box			No.of
	R-squared	RMSE	MAPE	MAE	Statistics	DF	Sig	Outliers
CL Price N50	.940	82.998	.491	60.627	13.724	17	.687	0

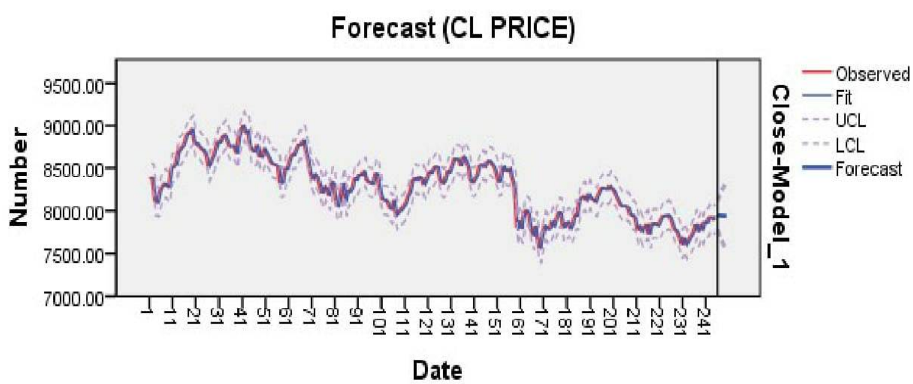


FIGURE 6. Forecast graph of Nifty 50 closing stock price

The forecast value is around 7820 and with a relative small standard error, which is also reflected in the 95% confidence limits. The predicted value is a little different from actual value, which is adequate for financial practice. The fluctuation of

Nifty 50 stock data is can be caused by various factors, such as India financial, RBI policy, International events and policy's. Finally (Figure 7), the comparison of forecast and actual closing stock price of Nifty 50.

TABLE 6. Predicted value of Nifty 50 closing stock price

Obs	Forecast	LCL	UCL	|Std.Error|
246	7814.92	7780.49	8115.34	23.63
247	7816.09	7701.73	8190.45	31.45
248	7778.26	7641.95	8246.57	37.26
249	7740.44	7591.62	8293.25	47.15
250	7726.61	7547.23	8333.99	52.26

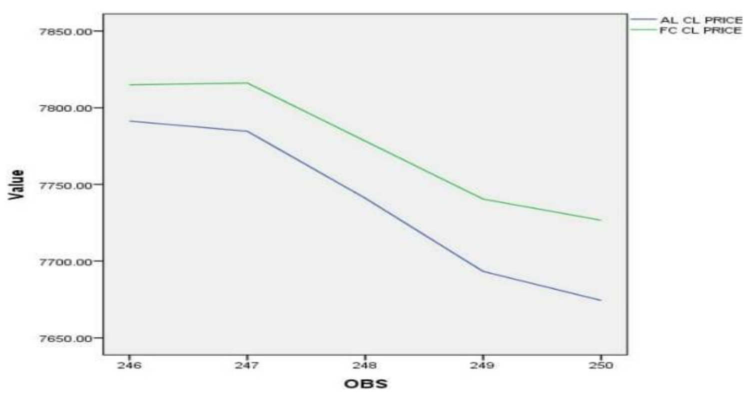


FIGURE 7. Actual and Forecast graph of Nifty 50 closing stock price

6. Conclusion

This paper does a study on 2015 NSE Nifty 50. ARIMA model offer an excellent technique for forecasting the importance of any variable. It is strength deceit in the verity that the method is fitting for any time series with any pattern of change. In the process of model building, the original Nifty 50 data is found to be Non-Stationary, but the first order differencing of original Nifty 50 data is stationary. In this study, ARIMA (0, 1, 1) model is developed for analyzing and forecasting Nifty 50 closing stock price among all of various tentative ARIMA models as it has lowest BIC values. From the results, it can be observed that influence R-Square value is (94%) high and Mean Absolute Percentage Error is very small for the fitted model. Thus the prediction accuracy is more fitting of Nifty 50. It is concluded that closing stock price of Nifty 50 taken in the present study shows slow decreasing fluctuations trend for upcoming trading days.

Acknowledgment: The First author express their sincere gradate to Basic Scientific Research Fellowship from University Grant Commission, New Delhi. The authors express their thanks to UGC-SAP (DRS-I) for providing infrastructure facilities.

References

1. Anderson, B. and Ledolder, J.: *Statistical Methods for Forecasting*, John Wiley & Sons, New York, 1983.
2. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31(3)** (1986) 309-317.
3. Box, G. E. P. and Jenkins, G. M.: *Time Series analysis:Forecasting and Control*, Holden Day, San Francisco, 1970.
4. Box, G. E. P. and Pierce, D.: Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models, *Journal of the American Statistical Association* **65** (1970) 1509-1526.
5. Brockwell, P. J. and Davis, R. A.: *Introduction to time series and forecasting, 2^{nd} edition*, Springer-Verlag, New York, 2002.
6. Javier, C., Rosario E. francisco and Antonio, J.C.: ARIMA Model to predict Next Electricity Price, *IEEE Transactions on Power Systems* **18(3)** (2003) 1014-1020.
7. Jawahar Farook, A. and Senthamarai Kannan, K.: Stochastic Modeling for Carbon Dioxide Emissions, *Journal of Statistics & Management Systems* **17(1)** (2014) 92-117.
8. Jun Zhang, Rui Shan and Wenfang.: Applying Time Series Analysis Builds Stock Price Forecast Model, *Modern Applied Science* **3(5)** (2009) 152-157.
9. Nail, P.E. and Momani, M.: Time Series Analysis for Rainfall Data in Jordan:Case Study for Using Time Series Analysis, *American Journal of Environmental Science* **5(5)** (2009) 599-604.
10. Renhao Jin, Sha Wang, Fang Yan and Jie Zhu.: The Application of ARIMA Model in 2014 Shanghai Composite Stock Price Index, *Journal of Applied Mathematics and Statistics* **3(4)** (2015) 199-203.
11. Walder Enders.: *Applied Econometric Time Series, 4^{th} edition,* Wiley, 2014.
12. Young H. Kim, Edward L. Davis and Charles T. Moses.: An ARIMA Model Approach to the Behaviour of Weekly Stock Prices of Fortune 500 Firms and S&P Small Cap 600 Firms, Oxford Journal:An International Journal of Business & Economics **10(1)** (2015) 22-47.

Mohamed Ashik: Research Scholar, Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu - 627 012, India.
E-mail address: ashikstat@gmail.com

Senthamarai Kannan: Professor, Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu - 627 012, India.
E-mail address: senkannan2002@gmail.com