

An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50

B. Uma Devi¹ D.Sundar² and Dr. P. Alli³

¹Assistant Professor, Department of Computer Science,
R.D.Govt. Arts College, Sivagangai

²Assistant Professor, MCA Department, Thiagarajar School of Management Madurai.

³Prof. & Head, CSE Department, Velammal College of Engg. & Technology, Madurai

ABSTRACT

The data mining and its tool has played a vital role in exploring the data from different ware houses. Using data mining tools and analytical technologies we do a quantifiable amount of research to explore new approach for the investment decisions .The market with huge volume of investor with good enough knowledge and have a prediction as well as control over their investments. The stock market some time fails to attract new investor. The reason states that non-aware and also people don't want to come forward to fall in to the risk. An approach with adequate expertise is designed to help investors to ascertain veiled patterns from the historic data that have feasible predictive ability in their investment decisions. In this paper the NSE – Nifty Midcap50 companies among them top 4 companies having max Midcap value has been selected for analysis. The historical data has a significant role in, helping the investing people to get an overview about the market behavior during the past decade. The stock data for the past five years has been collected and trained using ARIMA model with different parameters. The test criterions like Akaike Information Criterion Bayesian Information Criterion (AICBIC) are applied to predict the accuracy of the model. The performance of the trained model is analyzed and it also tested to find the trend and the market behavior for future forecast.

KEYWORDS

ARMA, ARIMA, ACF, PACF, AICBIC

1. INTRODUCTION

Data Mining is an approach where analytically designed to discover data (market related or business oriented) and in search of reliable patterns and/or logical relationships between variables. Appropriate validation is used to find the patterns with the detected patterns among the data set. It is also called as data discovery or knowledge discovery. Data mining enhance the revenue and reduce the cost incurred for the exploration of data.

The general research associated with stock market is highly focusing on neither buy nor sell. But it fails to address the dimensionality and expectancy of a naive investor [1]. The general trend towards stock market among the society is that it is highly risky for investment or not suitable for

trade. The seasonal variance and steady flow of any index will help both for existing and naïve investor to understand and make a decision to invest in the stock market.

To solve these types of problems, the time series analysis [11][15] will be the best tool for forecast and also to predict the trend. The trend chart will provide adequate guideline for the investor. Some time it may not address or forecast the variations or steady flow of the market. It may forecast only on particular season, but it is not adequate for long term decision making. The investors are very much interested to know the past trend or flow, seasonal growth, or variations of the stock [1]. A general view or expectation is that it must give a holistic view of the stock market.

Since, it is essential to identify a model to show the trend with adequate information for the investor to make a decision. It recommends that ARIMA is an algorithmic approach to transform the series is better than forecasting directly, and also it gives more accurate results [2]. This approach has not included any testing criteria for model estimation. Analysis of Price Causality and Forecasting in the Nifty Market futures employed to investigate the short-run [3]. This short run investigation will not bring any significance among the naïve investor as well market penetrator.

In this paper we focus on the real world problem in the stock market. The seasonal trend and flow is the highlight of the stock market. Eventually investors as well the stock broking company will also observe and capture the variations, constant growth of the index. This will aid new investor as well as existing people will make a strategic decision. It can be achieved by experience and the constant observations by the investors. In order to overcome the above said issues, we have suggested ARIMA algorithm in three steps,

Step 1: Model identification

Step 2: Model estimation

Step 3: Forecasting

2. PROBLEM DEFINITION ON INVESTMENT DECISIONS

The problems of stock forecasts are indispensable crops up from time to time. In all, the inevitable finish is that no issue what type of endeavor you are in, or what task you perform, there is a need for some kind of future estimate upon which to make a chart. Marketing society need forecasts to conclude to either enter or exit the business.

Finance professionals use forecasts to make financial plans. Investors invest their hard earned capital in stocks with the expectation of gaining from their investment through a positive payoff [4]. Since having an excellent knowledge about share price movement in the future serves the significance of fiscal professionals and investors. This familiarity about the future boosts their confidence by way of consulting and investing. But these movements predict the share prices without proper forecasting methods, only for the interest of the financial professional and investors. There are many forecasting methods in projecting price movement of stocks such as the Box Jenkins method, Black-Scholes model, and Binomial model.

3. OBJECTIVE OF THE STUDY

This study seeks to investigate [5] the best forecasting method under consideration, gives the minimum forecasting error by considering the following error metric.

The objectives are

- (a) *Model specification (or model identification)*;
- (b) *Model fitting (or model estimation)*;
- (c) *Model checking (or model verification or model criticism)*.

3.1. Forecasting Methods

The time series forecasting methods [6] generally classified into two groups based on statistical concepts and computational intelligence techniques such as neural networks (NN) or genetic algorithms (GA). Statistical time series forecasting methods are subdivided as:

- Exponential smoothing methods,
- Regression methods,
- Autoregressive integrated moving average (ARIMA) methods,
- Threshold methods,
- Generalized autoregressive conditionally heteroskedastic (GARCH) methods.

3.2 Proposed Approach

The Box-Jenkins methodology [6] is a five-step process for identifying, selecting, and assessing conditional mean models (for discrete, Univariate time series data [12]).

Phase 1: Identification

Step 1: Data Preparation

- Transform the data to stabilize the attributes.
- Find the difference if it is not stationary; successively difference
- Series to attain stationary.

Step 2: Model Selection

- Examine data, plot ACF and PACF to identify potential models

Phase 2: Estimation and Testing

Step 1: Estimation

- Estimate parameters in potential models
- Select best model using AICBIC criterion

Step 2: Diagnostics

- Check ACF/PACF of residuals
- Test the residuals
- Are the residuals are white noise

Phase 3: Forecast the application

- Forecasting the trend
- This model is used to forecast the future

3.3 ARIMA model

The ARIMA approach was popularized by Box and Jenkins[7][10], and ARIMA models are often referred to as Box-Jenkins models. The general transfer function model employed by the ARIMA procedure was discussed by Box and Tiao (1975).

ARIMA (**Auto Regressive Integrated Moving Average**) model [8] is a generalization of an autoregressive moving average (ARMA) model. An ARMA model expresses the conditional mean of y_t as a function of both past observations, y_{t-1}, \dots, y_{t-p} , and past innovations, $\epsilon_{t-1}, \dots, \epsilon_{t-q}$. The number of past observations that y_t depends on, p , is the AR degree. The number of past innovations that y_t depends on, is the MA degree.

In general, these models are denoted by ARMA (p, q). The form of the ARMA (p, q) model is, $y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$
Where ϵ_t is an uncorrelated innovation process with mean zero.

In this paper the results are obtained by Name-Value Arguments for Nonseasonal ARMA Models. The following parameters are used in the model,

C – Constant term

AR – Nonseasonal AR coefficients (ϕ_1, \dots, ϕ_p)

MA - Nonseasonal MA coefficients, ($\theta_1, \dots, \theta_q$)

ARLags – Lags corresponding to nonzero, nonseasonal AR coefficients

MALags – Lags corresponding to nonzero, nonseasonal MA coefficients

D - Degree of nonseasonal differencing, D [if D has value 0 meaning no nonseasonal integration]

Variance – Scalar variance of the innovation process (σ_ϵ^2)

Distribution - Distribution of the innovation process

4. EXPERIMENT AND RESULT ANALYSIS

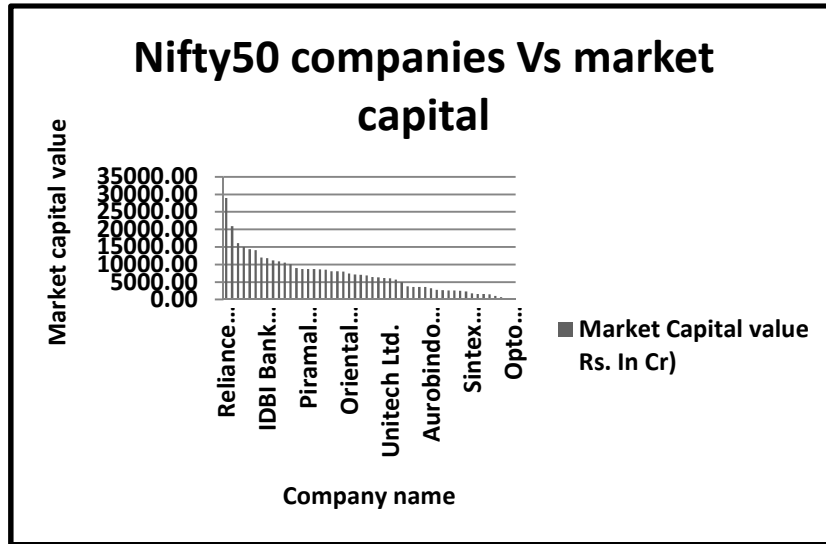


Figure 1. Niftymidcap50 companies Vs market capital

In Figure 1 the historical data has been collected from January 2007 to December 2011 from NSE India. The data preprocessing has been conducted based on the market capital value. The highest market capital values with the top four companies have been selected. Then the preprocessed data is transformed for the algorithm implementation. The Implementation works are done through MATLAB.

In phase1, the quarterly average of all closing point for the top four companies and Nifty-50 is [16] calculated and it is represented in the Table 1. The Figure 2 shows the graphical comparison for all the indices based on the quarterly average. In the above said figure all the series are *stationary and it is also non-linear*. The data appears to fluctuate around a constant mean, so differencing is not needed for all the indices. The Autocorrelation functions and Partial auto correlation functions are applied to identify the model i.e. AR, MA or ARMA.

Table 1. Niftymidcap50 & Top 4 indices - 3 Months Average.

Date (quarterly Months)	Nifty-Midcap50	Reliance	Ofss	Abb	Jsw-Steel
31-Jan-2007	2041.123	1350.292	1959.894	3626.511	444.3113
30-Apr-2007	2306.813	1612.465	2340.692	4112.518	583.603167
29-Jul-2007	2606.563	1911.517	2115.824	1145.562	673.7559
31-Oct-2007	3460.483	2733.4803	1561.786	1521.525	1023.01157
31-Jan-2008	2649.833	2528.6823	1115.875	1245.011	1047.55333
30-Apr-2008	2430.583	2441.604	1273.091	1045.722	952.017667
29-Jul-2008	2021.583	2121.8047	1237.654	853.0373	845.626267

31-Oct-2008	1246.267	1286.9073	508.0833	520.1437	255.050033
31-Jan-2009	1141.983	1295.0933	613.2047	423.4628	203.318367
30-Apr-2009	1818.933	1979.1053	1018.052	589.9227	461.967933
29-Jul-2009	2355.783	2007.211	1555.898	728.5094	689.952967
31-Oct-2009	2509	1699.1373	2133.343	765.5971	913.999333
31-Jan-2010	2582.6	1041.7717	2282.844	814.8283	1112.296
30-Apr-2010	2752.967	1053.1607	2158.268	822.6202	1141.246
29-Jul-2010	2872.05	1007.7467	2149.664	823.2794	1160.69067
31-Oct-2010	2974.85	1047.9233	2247.452	844.6569	1247.832
31-Jan-2011	2456.767	983.68733	2124.106	724.1128	940.469633
30-Apr-2011	2491.033	953.04773	2135.216	843.3934	925.7324
29-Jul-2011	2221.317	815.87697	2006.529	831.4897	731.599967
31-Oct-2011	1943.25	805.42053	2007.829	639.7541	587.729433

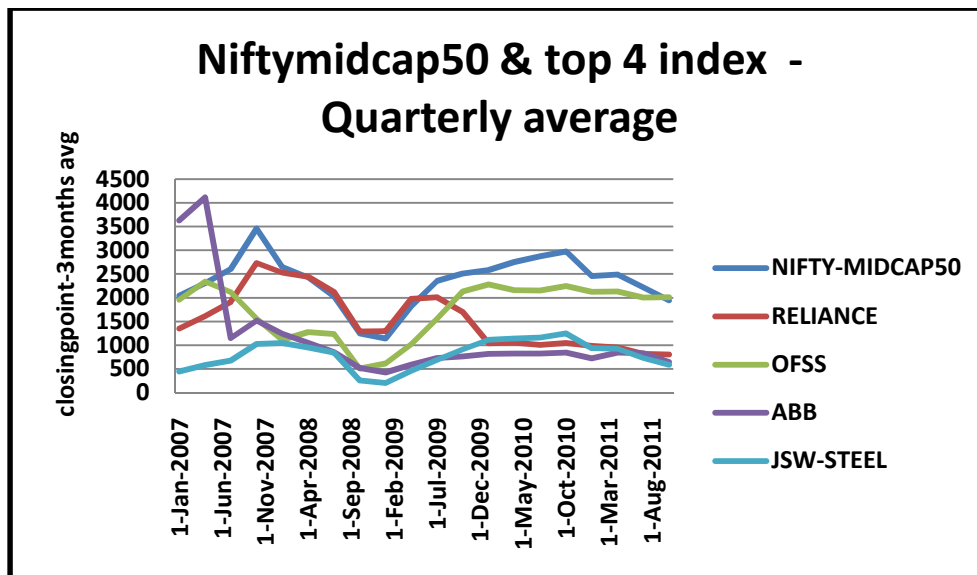


Figure 2. Comparison of Nifty50 & Top 4 indices - 3 months average

In phase-2, the parameters estimated through AR(Auto Regressive) and MA(Moving Average) are used the selection of potential models. The loglikelihood ratio test is accepted for each estimated models and the results is taken as input to the AICBIC. The AICBIC test is examined and the results are analysed to predict the best model. The row represents the AR(Auto Regressive) parameter and the column represents MA(parameter). The minimum value of the AICBIC test result is used to discover the selection of the best model. The Table 2 to 6 shows AICBIC test results. The comparison of the AICBIC test results are shown in Figures 3 to Figure 7.

AICBIC test for Best model Selection:

Table 2. NIFTY50: ARIMA (1, 0, 1)

AR/MR	MA(1)	MA(2)	MA(3)	MA(4)
AR(1)	302.68	317.63	307.80	307.63

Table 3. RELIANCE: ARIMA (1, 0, 1)

AR/MA	MA(1)	MA(2)
AR(1)	297.48	300.15

Table 4. OFSS: ARIMA (1, 0, 1)

AR/MA	MA(1)	MA(2)
AR(1)	285.13	288.05
AR(2)	288.10	306.73
AR(3)	290.16	311.61

Table 5. ABB: ARIMA (1, 0, 1)

AR/MA	MA(1)
AR(1)	318.9165

Table 6. JSWSTEEL: ARIMA (2, 0, 1)

AR/MA	MA(1)	MA(2)	MA(2)	MA(2)
AR(1)	274.26	277.86	275.04	275.06
AR(2)	269.38	271.65	271.31	272.27

Comparison of AICBIC test results for Best Model Selection:

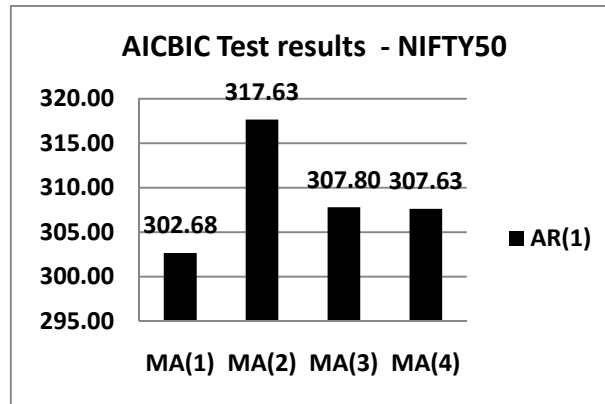


Figure 3. NIFTY-MIDCAP50

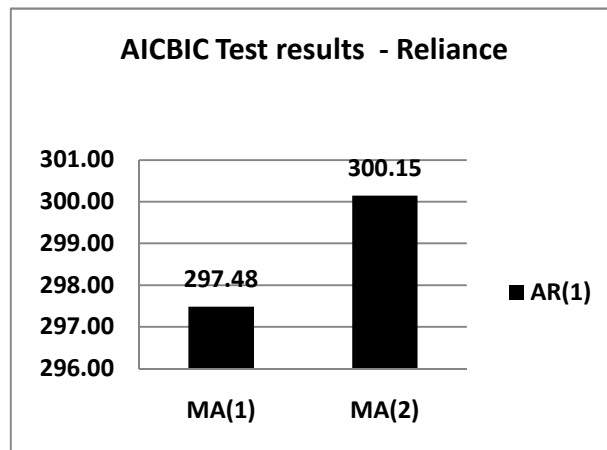


Figure 4. Reliance

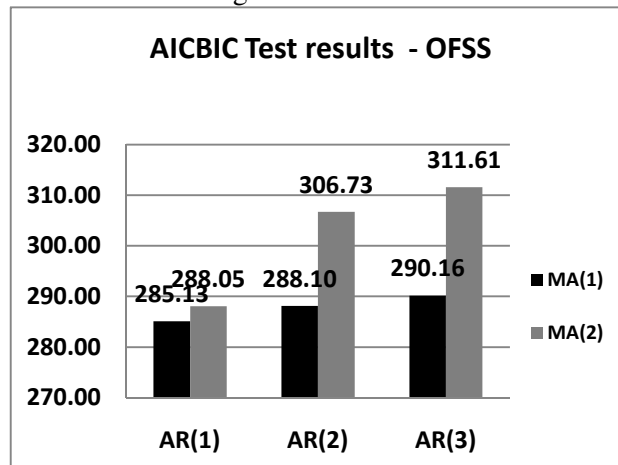


Figure 5. OFSS

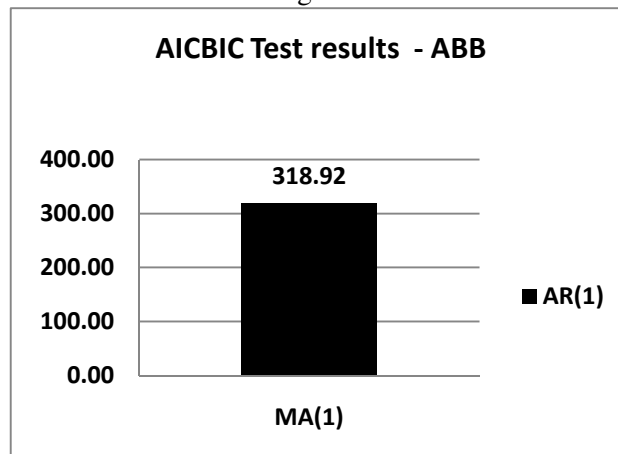


Figure 6. ABB

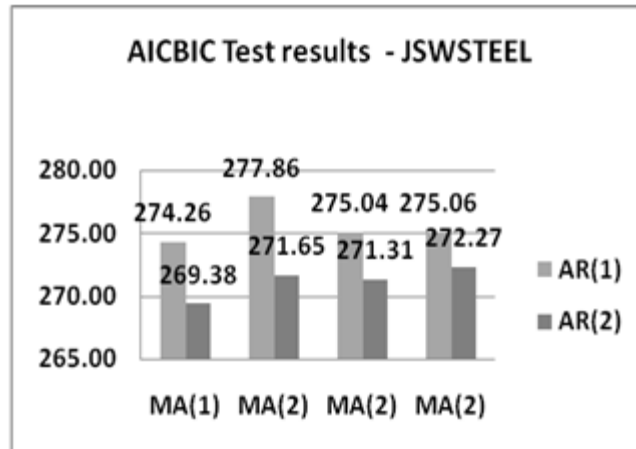


Figure 7. JSWSTEEL

The model selected is diagnosed subject to goodness of fit and infer the residuals for all indices. The goodness of fit will represent the distribution and the existence of the correlation among the data. In phase 3, the forecast is generated and find the approximation between the upper bound and lower bound. The model which is forecasted is used to predict the trend and seasonality for the future.

The equation given below shows best model which is forecasted,

$$y_{t(Nifty50)} = 1007.66 + 0.5585\phi_1 y_{t-1} + \epsilon_t + 0.3286\theta_1 \epsilon_{t-1}$$

$$y_{t(Reliance)} = 400.1841 + 0.7139\phi_1 y_{t-1} + \epsilon_t + 0.4257\theta_1 \epsilon_{t-1}$$

$$y_{t(OFSS)} = 605.9310 + 0.6125\phi_1 y_{t-1} + \epsilon_t + 1.0000\theta_1 \epsilon_{t-1}$$

$$y_{t(ABB)} = 93.8806 + 0.8225\phi_1 y_{t-1} + \epsilon_t - 0.7188\theta_1 \epsilon_{t-1}$$

$$y_{t(JSWSTEEL)} = 227.3470 + 1.5285\phi_1 y_{t-1} - 0.8200\phi_2 y_{t-2} + \epsilon_t - 0.6189\theta_1 \epsilon_{t-1}$$

5. PERFORMANCE MEASURES AND ACCURACY

The actual quarterly average data for the each index[14] is compared with the data generated by the model choosen. The three respective errors, such as MAPE, PMAD and % Error accuracy were arrived and is shown in Table 7. The bar chart given below (From Figures 8 to 10) represents the above mentioned performance measurements. The trend prediction interprets[13] that the investors are confidently recommend to invest in the index which has less percentage of error.

Table 7. Performance measures for NIFTY50 and TOP 4 indices

INDEX/ERROR	NIFTY-50	RELIANCE	OFSS	ABB	JSWSTEEL
MAPE	0.2108	0.3759	0.4073	0.3847	0.4798
PMAD	0.1792	0.3746	0.2902	0.5305	0.295
%Error-Accuracy	16.26%	31.40%	26.47%	38.12%	24.88%

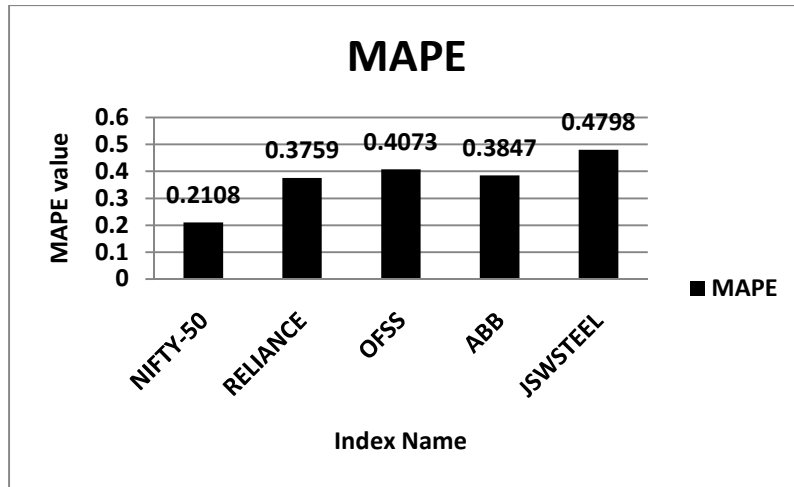


Figure 8. MAPE analysis between the Actual and Forecasted

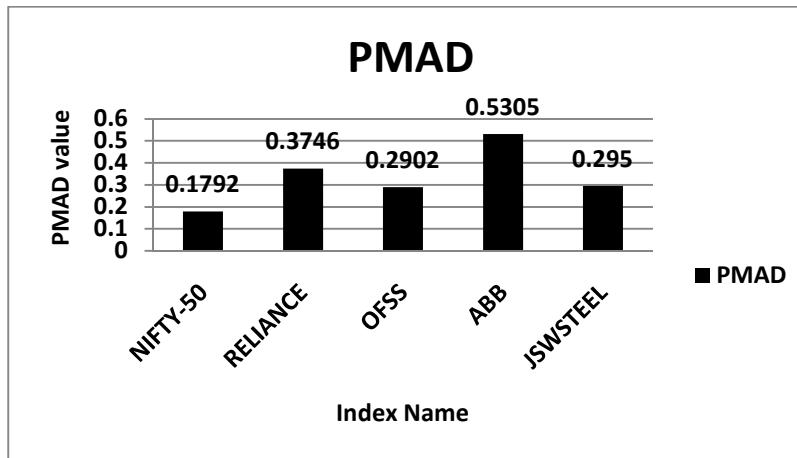


Figure 9. PMAD Analysis between the Actual and Forecasted

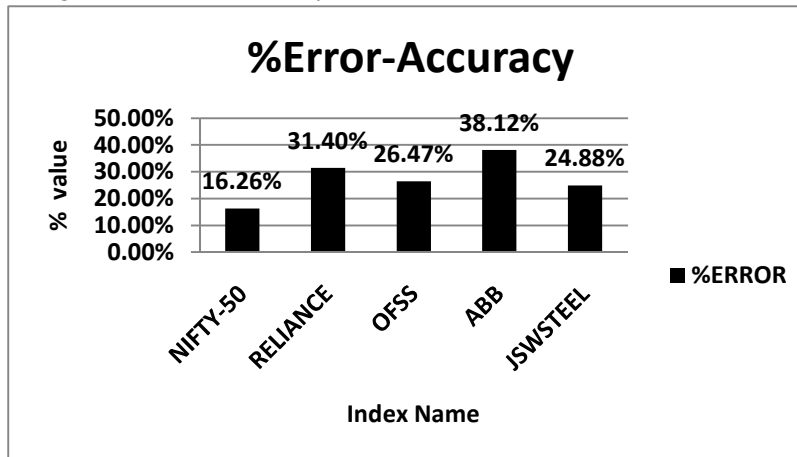


Figure 10. Error Accuracy analysis for all indices

6. CONCLUSION

In this paper the data has been collected from NSE.com. The historical data for the period of five year since 2007 to 2011 were taken in to account for analysis. The BoxJenkins methodology is used to identify the model. The AICBIC test criteria is applied against the data represented in the previous to select the best the model. The best model equation is derived for all indices. The MAPE, PMAD and % Error accuracy is applied to determine the discrimination between the actual historical data and the forecast data. This paper's inferences a new investment decisions or guidelines based on the minimum error percentage obtained through the above said performance measures. The future forecast of each index for next few years also highlighted in this paper. It is hoped that more innovative approaches will be conducted to bring the hidden information about the stock market.

REFERENCES

- [1] Fahim Faisal1, May 2012. Pp. 100 – 117, “Forecasting Bangladesh's Inflation Using Time Series ARIMA Models”, World Review of Business Research Vol. 2. No. 3.
- [2] Al Wadia, Mohd Tahir Ismail S, “Selecting Wavelet Transforms Model in Forecasting Financial Time Series Data Based on ARIMA Model”, Applied Mathematical Sciences, Vol. 5, 2011, no. 7, 315 – 326.
- [3] Kailash Chandra Pradhan, K. Sham Bhat, (2009), “An Empirical Analysis of Price Discovery, Causality and Forecasting in the Nifty Futures Markets”, International Research Journal of Finance and Economics, ISSN 1450-2887 Issue 26.
- [4] Kofi agyarko ababio, June 2012, “Comparative study of stock price forecasting using arima and arimax models”.
- [5] Chris Chatfield, “TIME-SERIES FORECASTING”.
- [6] Xiangyu Tang, Chunyu Yang, Jie Zho, 2009, “Stock Price Forecasting by Combining News Mining and Time Series Analysis”, International Conference on Web Intelligence and Intelligent Agent Technology– Workshops -IEEE/WIC/ACM.
- [7] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, 1994, “Time Series Analysis: Forecasting and Control”, UpperSaddleRiver”, NJ: Prentice-Hall, 3rd ed.
- [8] Hussein Ali Al-Zeaud European Journal of Economics, (2011), “Modeling and Forecasting Volatility Using ARIMA Model”, Finance and Administrative Sciences ISSN 1450-2275 Issue 35.
- [9] Daniel Traian PELE Vergil VOINEAGU, “Testing market efficiency via decomposition of stock return. Application to Romanian capital market”.
- [10] Box, George and Jenkins, Gwilym, San Francisco: Holden-Day, (1970), “Time series analysis: Forecasting and control “.
- [11] Brockwell, Peter J. and Davis, Richard A, (1987), “Time Series: Theory and Method”, Springer-Verlag.
- [12] Pankratz, Alan New York: John Wiley & Sons, (1983), “Forecasting with Univariate Box–Jenkins models: concepts and cases”.
- [13] S Abdulsalam Sulaiman Olaniyi, Adewole, Kayode S. Jimoh, R. G, JULY 2011, “Stock Trend Prediction Using Regression Analysis –A Data Mining Approach”, - ARPN Journal of Systems and Software- --Volume 1 No. 4, ISSN 2222-9833.
- [14] K.K.Sureshkumar, Dr.N.M.Elango, November 2011, “An Efficient Approach to Forecast Indian Stock Market Price and their Performance Analysis”, International Journal of Computer Applications (0975 – 8887), Volume 34– No.5.
- [15] K.K.Sureshkumar -Dr.N.M.Elango, “Exploiting data mining techniques for improving The efficiency of time series data Using spss-clementine”, journal of arts, science & commerce, e-issn 2229-4686 - ISSN 2231-4172.
- [16] Kailash Chandra Pradhan, (2009), “An Empirical Analysis of Price Discovery, Causality and Forecasting in the Nifty Futures Markets”, International Research Journal of Finance and Economics - ISSN 1450-2887 Issue 26.

[17] Yaron Klein, and Abraham Kandel, February 2001, “Knowledge Discovery in Time Series Databases Mark Last”, IEEE transactions on systems, man, and cybernetics—part b: cybernetics, vol. 31, no. 1.
 [18] Jesús Navarro-Moreno, JULY 2008, “ARMA Prediction of Widely Linear Systems by Using the Innovations Algorithm”, IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 56, NO. 7.
 [19] Wai-Keung Wong, Enjian Bai and Alice Wai-Ching Chu, December 2010, “Adaptive Time-Variant Models for Fuzzy-Time-Series Forecasting”, IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 40, No. 6.
 [20] www.wikipedia.com , www.mathworks.com

Annexure:

Table: Diagnostic Information for the selection of best model

Index name	No. of iterations	algorithm selected	P	D	Q	Best model ARIMA(P,D,Q)	Constant	AR{1} Coeff.	AR{2} Coeff.	MA{1} Coeff.	Variance	Distribution
NIFTY	57	sequential quadratic programming	1	0	1	ARIMA(1,0,1)	1007.6600	0.5585	*****	0.3286	139634.0000	Gaussian
RELIANCE	45	sequential quadratic programming	1	0	1	ARIMA(1,0,1)	400.1841	0.7139	*****	0.4257	107674.2606	Gaussian
OFSS	52	sequential quadratic programming	1	0	1	ARIMA(1,0,1)	605.9310	0.6125	*****	1.0000	58057.4000	Gaussian
ABB	70	sequential quadratic programming	1	0	1	ARIMA(1,0,1)	93.8806	0.8225	*****	-0.7188	314453.0000	Gaussian
JSWSTEEL	47	sequential quadratic programming	2	0	1	ARIMA(2,0,1)	227.3470	1.5285	-0.82	-0.6189	22738.2000	Gaussian

Note: ARIMA (1, 0, 1) = ARMA (1, 1) since D (Differencing) = 0.

Forecast results for NIFTY50 and TOP Four Indices:

Figure 11. NIFTY50-Actual Vs NIFTY50-Forecast

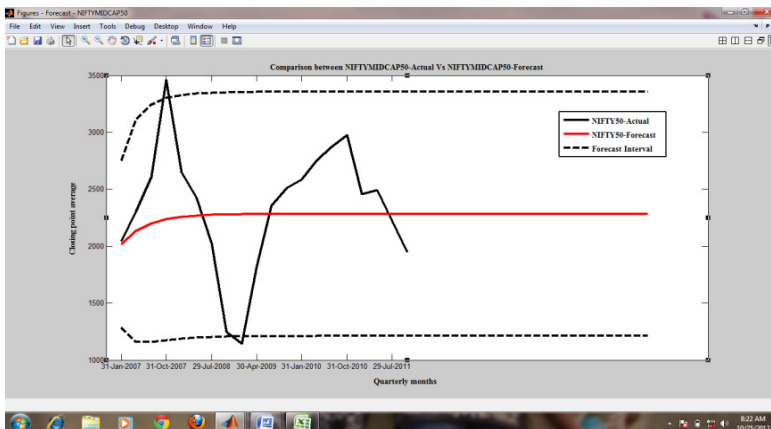


Figure 12. RELIANCE-Actual Vs RELIANCE-Forecast

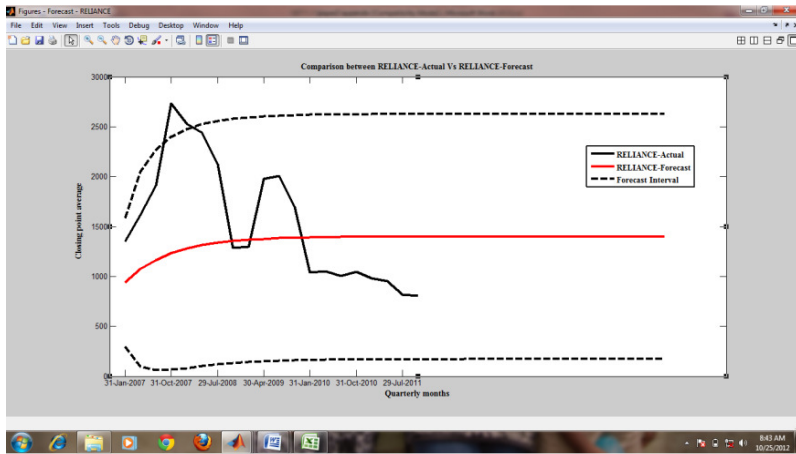


Figure 13. OFSS-Actual Vs OFSS-Forecast

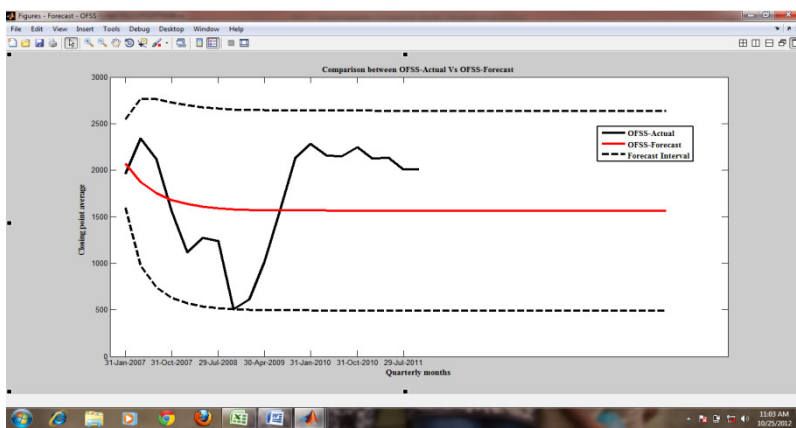


Figure 14. ABB-Actual Vs ABB-Forecast

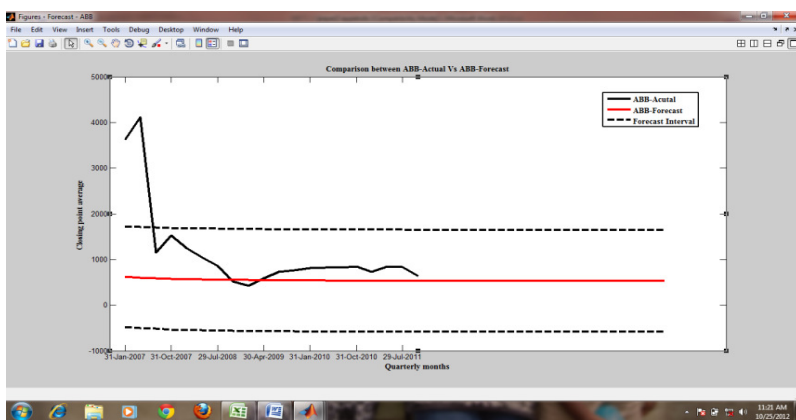


Figure 15. JSWSTEEL-Actual Vs JSWSTEEL-Forecast

