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 warehouses. 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, $y_{t-1},...y_{t-p}$, and past innovations, $\epsilon_{t-1},...\epsilon_{t-q}$. 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,

$$y_t = C + \varnothing_1 y_{t-1} + \varnothing_2 y_{t-2} + \ldots + \varnothing_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

Where $\epsilon_t$ 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 ($\varnothing_1,\ldots,\varnothing_p$)
- MA - Nonseasonal MA coefficients, ($\theta_1,\ldots,\theta_q$)
- 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 ($\sigma^2_{\epsilon}$)
- Distribution - Distribution of the innovation process
4. **EXPERIMENT AND RESULT ANALYSIS**

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 phase 1, the quarterly average of all closing point for the top four companies and Nifty-50 is 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 autocorrelation functions are applied to identify the model i.e. AR, MA or ARMA.

![Nifty50 companies Vs market capital](image)

**Figure 1.** Niftymidcap50 companies Vs market capital

<table>
<thead>
<tr>
<th>Date (quarterly Months)</th>
<th>Nifty-Midcap50</th>
<th>Reliance</th>
<th>Ofss</th>
<th>Abb</th>
<th>Jsw-Steel</th>
</tr>
</thead>
<tbody>
<tr>
<td>31-Jan-2007</td>
<td>2041.123</td>
<td>1350.292</td>
<td>1959.894</td>
<td>3626.511</td>
<td>444.3113</td>
</tr>
<tr>
<td>30-Apr-2007</td>
<td>2306.813</td>
<td>1612.465</td>
<td>2340.692</td>
<td>4112.518</td>
<td>583.603167</td>
</tr>
<tr>
<td>29-Jul-2007</td>
<td>2606.563</td>
<td>1911.517</td>
<td>2115.824</td>
<td>1145.562</td>
<td>10 23.01157</td>
</tr>
<tr>
<td>31-Oct-2007</td>
<td>3460.483</td>
<td>2733.4803</td>
<td>1561.786</td>
<td>1521.525</td>
<td>1023.01157</td>
</tr>
<tr>
<td>31-Jan-2008</td>
<td>2649.833</td>
<td>2528.6823</td>
<td>1115.875</td>
<td>1245.011</td>
<td>1047.55333</td>
</tr>
<tr>
<td>30-Apr-2008</td>
<td>2430.583</td>
<td>2441.604</td>
<td>1273.091</td>
<td>1045.722</td>
<td>952.017667</td>
</tr>
<tr>
<td>29-Jul-2008</td>
<td>2021.583</td>
<td>2121.8047</td>
<td>1237.654</td>
<td>853.0373</td>
<td>845.626267</td>
</tr>
</tbody>
</table>
In phase-2, the parameters estimated through AR(Auto Regressive) and MA(Moving Average) are used to select potential models. The loglikelihood ratio test is accepted for each estimated model and the results are 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)

<table>
<thead>
<tr>
<th>AR/MR</th>
<th>MA(1)</th>
<th>MA(2)</th>
<th>MA(3)</th>
<th>MA(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>302.68</td>
<td>317.63</td>
<td>307.80</td>
<td>307.63</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>AR/MA</th>
<th>MA(1)</th>
<th>MA(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>297.48</td>
<td>300.15</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>AR/MA</th>
<th>MA(1)</th>
<th>MA(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>285.13</td>
<td>288.05</td>
</tr>
<tr>
<td>AR(2)</td>
<td>288.10</td>
<td>306.73</td>
</tr>
<tr>
<td>AR(3)</td>
<td>290.16</td>
<td>311.61</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>AR/MA</th>
<th>MA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>318.9165</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>AR/MA</th>
<th>MA(1)</th>
<th>MA(2)</th>
<th>MA(3)</th>
<th>MA(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>274.26</td>
<td>277.86</td>
<td>275.04</td>
<td>275.06</td>
</tr>
<tr>
<td>AR(2)</td>
<td>269.38</td>
<td>271.65</td>
<td>271.31</td>
<td>272.27</td>
</tr>
</tbody>
</table>

Comparison of AICBIC test results for Best Model Selection:

Figure 3. NIFTY-MIDCAP50
Figure 4. Reliance

Figure 5. OFSS

Figure 6. ABB
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,

\[
\begin{align*}
\gamma_t(\text{Nifty50}) &= 1007.66 + 0.5585 \phi_1 y_{t-1} + \varepsilon_t + 0.3286 \theta_1 \varepsilon_{t-1} \\
\gamma_t(\text{Reliance}) &= 400.1841 + 0.7139 \phi_1 y_{t-1} + \varepsilon_t + 0.4257 \theta_1 \varepsilon_{t-1} \\
\gamma_t(\text{OFSS}) &= 605.9310 + 0.6125 \phi_1 y_{t-1} + \varepsilon_t + 1.0000 \theta_1 \varepsilon_{t-1} \\
\gamma_t(\text{ABB}) &= 93.8806 + 0.8225 \phi_1 y_{t-1} + \varepsilon_t - 0.7188 \theta_1 \varepsilon_{t-1} \\
\gamma_t(\text{JSWSTEEL}) &= 227.3470 + 1.5285 \phi_1 y_{t-1} - 0.8200 \phi_2 y_{t-2} + \varepsilon_t - 0.6189 \theta_1 \varepsilon_{t-1}
\end{align*}
\]

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>
<thead>
<tr>
<th>INDEX/ERROR</th>
<th>NIFTY-50</th>
<th>RELIANCE</th>
<th>OFSS</th>
<th>ABB</th>
<th>JSWSTEEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>0.2108</td>
<td>0.3759</td>
<td>0.4073</td>
<td>0.3847</td>
<td>0.4798</td>
</tr>
<tr>
<td>PMAD</td>
<td>0.1792</td>
<td>0.3746</td>
<td>0.2902</td>
<td>0.5305</td>
<td>0.295</td>
</tr>
<tr>
<td>% Error Accuracy</td>
<td>16.26%</td>
<td>31.40%</td>
<td>26.47%</td>
<td>38.12%</td>
<td>24.88%</td>
</tr>
</tbody>
</table>
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 papers 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

Appendix:

Table: Diagnostic Information for the selection of best model

<table>
<thead>
<tr>
<th>Index name</th>
<th>No. of iterations</th>
<th>algorithm selected</th>
<th>P</th>
<th>Q</th>
<th>ARIMA(P,D,Q)</th>
<th>Constant</th>
<th>AR(1) Coeff.</th>
<th>AR(2) Coeff.</th>
<th>MA(1) Coeff.</th>
<th>Variance</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIFTY</td>
<td>57</td>
<td>sequential quadratic programming</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>ARIMA(1,0,1)</td>
<td>1097.6600</td>
<td>0.5585</td>
<td>0.1286</td>
<td>135063.0000</td>
<td>Gaussian</td>
</tr>
<tr>
<td>RELIANCE</td>
<td>45</td>
<td>sequential quadratic programming</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>ARIMA(1,0,1)</td>
<td>400.1341</td>
<td>0.7139</td>
<td>0.4257</td>
<td>107674.2650</td>
<td>Gaussian</td>
</tr>
<tr>
<td>OIL</td>
<td>52</td>
<td>sequential quadratic programming</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>ARIMA(1,0,1)</td>
<td>650.9310</td>
<td>0.6125</td>
<td>1.0000</td>
<td>59857.4000</td>
<td>Gaussian</td>
</tr>
<tr>
<td>ASB</td>
<td>70</td>
<td>sequential quadratic programming</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>ARIMA(1,0,1)</td>
<td>93.8806</td>
<td>0.8225</td>
<td>-0.7188</td>
<td>314453.0000</td>
<td>Gaussian</td>
</tr>
<tr>
<td>ISWiSTEEL</td>
<td>47</td>
<td>sequential quadratic programming</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>ARIMA(2,0,1)</td>
<td>227.3470</td>
<td>1.5285</td>
<td>-0.82</td>
<td>-0.6189</td>
<td>22738.2000</td>
</tr>
</tbody>
</table>

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