Journal of Computer and Mathematical Sciences, *Vol.9(12),1915-1930 December 2018* *ISSN 0976-5727 (Print)*
(An International Research Journal), *www.compmath-journal.org* *ISSN 2319-8133 (Online)*

Application of Time Series Analysis in the Daily Stock Exchange Data

Mohankumari C[1], Vishukumar M[2] and Nagaraja Rao Chillale[3]

[1]Assistant Professor, Department of Statistics,
School of Applied Sciences, REVA University, INDIA.
[2]Professor, Department of Mathematics,
School of Applied Sciences, REVA University, INDIA.
[3]Associate Professor, Department of Statistics,
Bangalore University, Bangalore, INDIA.

(Received on: November 29, 2018)

ABSTRACT

Forecasting is a necessity of human life and a common problem in all branches of learning, financial and economic problems are domains in which forecasting is of major importance. In the field of stock exchange, the basic goal of market participants is to predict the future trends of stock price and determine the best time to execute transactions in order to optimize investment decisions.

A stock, also known as equity of share is a portion of the ownership in a corporate by an individual. Hence, a stock of a company entitles its holder a share in its profit. Only by issuing shares a corporate company can mobilize huge capitals. The stock market is a field of financial game and it can fetch bigger financial benefits compared to fixed deposits with banks. The stability as well as the inflation of the economy of a country is swiftly and better reflected by the trend in the stock market. So the study of the fluctuations in the stock market becomes important. There are many approaches to know the depth of an analysis of stock price variation. So we have arrived to propose forecasting methods such as **Time Series Analysis** to provide better accuracy in forecasting as compared to traditional methods. Published stock data obtained from National Stock Exchange (NSE) are used with stock price predictive model developed. The results obtained revealed that the ARIMA model has a strong potential for short-term prediction and can compete favourably with existing prediction techniques for stock price prediction.

Keywords: Time Series Analysis, ARIMA models.

1. INTRODUCTION

Stock market analysts have adopted many statistical techniques likes Auto Regressive Moving Average (ARMA) , Auto Regressive Integrated Moving Average (ARIMA), Auto Regressive Conditional Heteroscedasticity(ARCH),Generalized Auto Regressive Conditional Heteroscadasticity(GARCH), ARMA-EGARCH , Box and Jenkins approach along with various soft computing and evolutionary computing methods.

What is the most I can lose on this investment?

This is the question that almost every Investor who has invested or is considering for making investment in a risky asset at some point of time. The Value at risk measures the potential loss in value of a risky asset or portfolio over a defined period for a given confidence interval.

2. REVIEW OF LITERATURES

- Chan *et al*. (2014): They examine the intraday relationship between returns and returns volatility in the stock index and stock index futures markets. There results indicate a strong inter market dependence in the volatility of the cash and futures returns. Price innovations that originate in either the stock or futures markets can predict the future volatility in the other market. They show that this relationship persists even during periods in which the dependence in the returns themselves appears to weaken. The findings are robust to controlling for potential market frictions such as asynchronous trading in the stock index. There results have implications for understanding the pattern of information flows between the two markets.
- Puspanjali *et al*. (2012): Authors presents a scheme using Differential Evolution based Functional Link Artificial Neural Network (FLANN) to predict the Indian Stock Market Indices. The Model uses Back-Propagation (BP) algorithm and Differential Evolution(DE) algorithm respectively for predicting the Stock Price Indices for one day, one week, two weeks and one month in advance. The Indian stock prices i.e. BSE (Bombay Stock Exchange), NSE, INFY etc. with few technical indicators are considered as input for the experimental data. In all the cases, DE outperforms the BP algorithm. The Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are calculated for performance evaluation. The MAPE and RMSE in case of DE are found to be very less in comparison to BP method. The simulation study has been done using Java-6 and Net Beans within this.

Accurate stock prediction is always a very challenging task. They proposed FLANN model trained with back propagation is giving good result as per the recorded RMSE, and MAPE values during testing for one day, one week, two weeks and one month ahead respectively. The DE optimized FLANN is proving it's superiority as far as RMSE and MAPE are concerned. Further they predict for further work on performance of DE is to be compared with Particle Swarm Optimization (PSO).

- Preethi *et al*. (2012): Author predicted surveys on recent literature in the area of Neural Network, Data Mining, Hidden Markov Model and Neuro-Fuzzy system used to predict the stock market fluctuation. Neural Networks and Neuro-Fuzzy systems are identified to be the leading machine learning techniques in stock market index prediction area. The Traditional techniques are not cover all the possible relation of the stock price fluctuations. There are new approaches to known in-depth of an analysis of stock price variations. NN and Markov Model can be used exclusively in the finance markets and forecasting of stock price. In this paper, they propose a forecasting method to provide better an accuracy rather traditional method.

Further, authors predicted the future work as Neural Network and Markov model can also explore for other applications and comparative study with other time series analysis and forecasting models.

- Men *et al*. (2008): In their work the authors predicted the fluctuations of stock prices and trade volumes are investigated by the method of Zipf plot, where Zipf plot technique is frequently used in physics science. Author discuss the statistical properties of fat tails phenomena and the power law distributions for the daily stocks prices and trade volumes. In the second part, they consider the fat tails phenomena and the power law distributions of Shanghai Stock Exchange Index and Shenzhen Stock Exchange Index during the years 2001-2006 by Zipf plot method.
- Wang *et al*. (2008): Here the authors predicted the data of Chinese stock markets is analyzed by the statistical methods and computer sciences. The fluctuations of stock prices and trade volumes are investigated by the method of Zipf plot, where Zipf plot technique is frequently used in physics science. The objective of this research is to investigate the power law behavior and the fat tails phenomena of Chinese stock markets. Some research work has been done for Chinese stock markets. In this paper, they (Jun Wang, Bingli Fan and Dongping) continue the research work by Zipf plot method. The work they have done during the period between 2002-2007.

For our proposed research, we got the required data from website: nseindia.com.

3.1. DATA: The data series we study are daily frequency observation on closing prices of company and are going to take from nseindia.com, the National Stock Exchange of India. We consider the data for company: ONGC. The observations run from 1st January 2013 to 31st May 2016 for ONGC.

3.2. DATA COLLECTION: The analysis was conducted based on closing prices of the data from 1st January 2013 to 31st May 2016 expressed in Rupees and did not include dividends. Table 1 presents the names of the companies and sample of dataset. The required data was obtained from nseindia.com, the National Stock Exchange of India Limited or S & P CNX NIFTY (NSE), is a Mumbai-based stock exchange. It is the largest stock exchange in India in terms of the daily turnover and the volume of equities and derivatives method.

TABLE of sample of our data set:

DATE	ONGC
02-May-16	217.75
03-May-16	215.3
04-May-16	210.65
05-May-16	210.5
06-May-16	208.45
09-May-16	211.6
10-May-16	207.9
11-May-16	205.25
12-May-16	206.3
13-May-16	203.75
16-May-16	202.6
17-May-16	210.25
18-May-16	214.25
19-May-16	210.05
20-May-16	213.1
23-May-16	208.15
24-May-16	206.65
25-May-16	210.15
26-May-16	216.05
27-May-16	213
30-May-16	213.45
31-May-16	210.75

The following TABLE is sample of computed relative returns

DATE	ONGC
3-May-16	-0.0216
4-May-16	-0.00071
05-May-16	-0.00974
06-May-16	0.01511
09-May-16	-0.01749
10-May-16	-0.01275
11-May-16	0.005116
12-May-16	-0.01236
13-May-16	-0.00564
16-May-16	0.037759
17-May-16	0.019025
18-May-16	-0.0196
19-May-16	0.01452
20-May-16	-0.02323
23-May-16	-0.00721
24-May-16	0.01693
25-May-16	0.02807
26-May-16	-0.01412
27-May-16	0.002113
30-May-16	-0.01265

3.3. DATA PREPARATION: Relative return is calculated for the data i.e., the return that an asset achieves over a period of time compared to a benchmark. The relative return is the difference between the absolute return achieved by the asset and the return achieved by the benchmark.

The daily return on the portfolio was calculated using the formula:

$R_t = \frac{S_t - S_{t-1}}{S_{t-1}}$, Where S_t : Value of stock at the close of day t & S_{t-1}: Value of portfolio at the close of day t-1

Why is relative return so important?

Because it is a way to measure the performance of actively managed funds, which should get a return greater than that of the market. Relative return can also be used within a context smaller than the entire market.

The bottom line is that absolute return does not say much on its own.

You need to look at the relative return to see how an investment's return compares to other similar investments. Once you have a comparable benchmark in which to measure your investment's return, you can then make a decision of whether your investment is doing well or poorly and act accordingly.

4. STATISTICAL TOOLS

4.1. TIME SERIES ANALYSIS: A time series is a sequence of observation taken sequentially in time.

In statistics, a **time series** is a sequence of data points, measured typically at successive times, spaced at time intervals. **Time series *analysis*** comprises methods that attempt to understand such time series, often either to understand the underlying context of the data or to make forecasts (predictions). **Time series *forecasting*** is the use of a model to forecast future events based on known past events: to forecast future data points before they are measured.

A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values in a series for a given time will be expressed as deriving in some way from past values, rather than from future values. As shown by Box and Jenkins[1], models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the *autoregressive* (AR) models, the *integrated* (I) models, and the *moving average* (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce *autoregressive moving average* (ARMA) and *autoregressive integrated moving average* (ARIMA) models.

4.2. Autoregressive process: The AR (p) model is given by

$$x_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \ldots\ldots.. + \phi_p X_{t-p} + \varepsilon_t$$

Where: μ is a constant, $\phi_1, \phi_2, \phi_3, \ldots\ldots\ldots\ldots \phi_p$ are the autoregressive parameters.

Assumption: Here the term ε_t is the source of randomness and is called white noise. It is assumed to have the following characteristics:

1. $E[\varepsilon_t] = 0$ 2. $E[\varepsilon_t^2] = \sigma^2$ 3. $E[\varepsilon_t \varepsilon_s] = 0 \quad \forall t \neq s$

4.3. Moving average process: Independent from the autoregressive process, each element in the series can also be affected by the past error (or random shock) that cannot be accounted for by the autoregressive component, that is:

$x_t = \mu + \varepsilon_t - \theta_1 * \varepsilon_{t-1} - \theta_2 * \varepsilon_{t-2} - \theta_3 * \varepsilon_{t-3} - ... - \theta_q * \varepsilon_{t-q}$

Where: μ is a constant, $\theta_1, \theta_2, \theta_3, \ldots.. \theta_q$ are the moving average parameters.

Assumption: Here the term ε_t is the source of randomness and is called white noise. It is assumed to have the following characteristics:

1. $E[\varepsilon_t] = 0$ 2. $E[\varepsilon_t^2] = \sigma^2$ 3. $E[\varepsilon_t \varepsilon_s] = 0 \quad \forall t \neq s$

4.4. Autoregressive Moving average model: The general model introduced by Box and Jenkins (1976) includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. Specifically, the three types of parameters in the model are: the autoregressive parameters (*p*), the number of differencing passes (*d*), and moving average parameters (*q*). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (*p*, *d*, *q*); so, for example, a model described as (0, 1, 2) means that it contains 0 (zero) autoregressive (*p*) parameters and 2 moving average (*q*) parameters which were computed for the series after it was differenced once.
The ARIMA (p, d, q) model is given by

$x_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + ... + \phi_p X_{t-p} + \varepsilon_t - \theta_1 * \varepsilon_{t-1} - \theta_2 * \varepsilon_{t-2} - \theta_3 * \varepsilon_{t-3} - ... - \theta_q * \varepsilon_{t-q}$

Where: μ is a constant, $\phi_1, \phi_2, \phi_3, ... \phi_p$ are the autoregressive parameters and $\theta_1, \theta_2, \theta_3, \ldots.. \theta_q$ are the moving average parameters.

4.5. 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.

4.6. Proposed Approach: The Box-Jenkins methodology[6] is a three-step process for identifying (Identification), selecting (Estimation and Testing), and assessing(Forecast the application) conditional mean models (for discrete, Univariate time series data[12])

5. METHODOLOGY

The method used in this study to develop ARIMA model for stock price forecasting is explained in detail in subsections below. The tool used for implementation is [R-Software and Eviews software version 8.1]. Stock data used in this research work are historical daily stock prices obtained from ONGC company stock exchanged. The data composed of four elements, namely: open price, low price, high price and close price respectively. In this research the closing price is chosen to represent the price of the index to be predicted. Closing price is chosen because it reflects all the activities of the index in a trading day.

To determine the best ARIMA model among several experiments performed, the following criteria are used in this study for each stock index.

(i) Relatively small of AIC (Akaike Information Criterion) or BIC (Bayesian or Schwarz Information Criterion)
(ii) Relatively small standard error of regression (S.E. of regression)
(iii) Relatively high of adjusted R^2.
(iv) Q-statistics and Correlogram show that there is no significant pattern left in the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) of the residuals, it means the residual of the selected model are white noise.

The section below described the processes of ARIMA model-development.

A. An ARIMA (p, d, q) Model for Stock Index:

NSE stock data used in this study covers the period from 1st January 2013 to 31st May 2016 having a total number of 891 observations. Table 1 depicts the summary statistics of company ONGC. A common rule-of-thumb tests for normality is to run descriptive statistics to get Skewness and Kurtosis. **Skewness** is "the tilt (or lack of it) in a distribution". The more common type is right skew, where the smaller tail points to the right. Less common is left skew, where the smaller tail is points left. Skew should be within +1 to -1 range when the data are normally distributed. **Kurtosis** is "the peakedness of a distribution" (i.e., kurtosis should be within +3 to -3 range when the data are normally distributed).

Table-1: Descriptive STATISTICS of company ONGC

	ONGC_1
Observations	891
Median	307.2500
Maximum	465.6500
Minimum	191.5000
Std. Dev.	61.37813
Skewness	0.308752
Kurtosis	2.517890
Jarque-Bera	22.78510
Probability	0.000011

We observed that Skewness for the daily returns of all the stocks are within +1 to -1 , which is an indication that the data are normally distributed.

From the table below We observe that kurtosis of ONGC has normal kurtosis which is an indication that data are not normally distributed. But we assume in the long run the variables are normally distributed.

Figure 1 depicts the original pattern of the series to have general overview whether the time series is stationary or not. From the graph below the time series have random walk pattern.

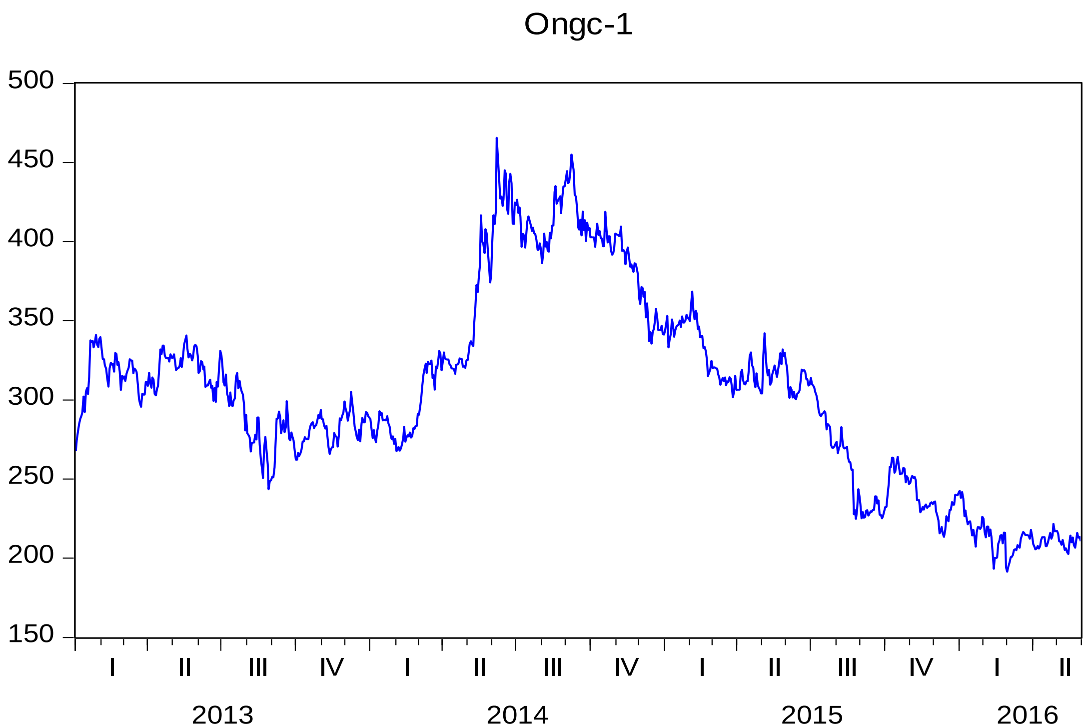

Figure 1: Graphical representation of the Stock closing price of ONGC

Figure 2 is the correlograms of ONGC. From the graph, the ACF dies down extremely slowly which simply means that the time series is nonstationary. If the series is not stationary, it is converted to a stationary series by differencing[lag]. After the first difference, the series “DONGC” becomes stationary as shown in figures 3 and 4 of the line graph and correlogram respectively.

Date: 11/08/18 Time: 18:26

Sample: 1/01/2013 5/31/2016

Included observations: 891

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|*******	.|*******	1	0.993	0.993	881.14	0.000
.|*******	.| |	2	0.986	0.020	1751.1	0.000
.|*******	.|* |	3	0.980	0.078	2612.0	0.000
.|*******	.| |	4	0.975	0.045	3464.9	0.000
.|*******	.| |	5	0.970	-0.026	4309.1	0.000
.|*******	.| |	6	0.964	-0.020	5144.0	0.000
.|*******	.| |	7	0.958	0.024	5970.5	0.000
.|*******	.| |	8	0.953	0.037	6789.6	0.000
.|*******	.| |	9	0.949	0.015	7601.7	0.000
.|*******	.| |	10	0.944	0.032	8407.3	0.000
.|*******	.| |	11	0.939	-0.047	9205.2	0.000
.|*******	.| |	12	0.934	-0.031	9994.9	0.000
.|*******	.| |	13	0.929	0.007	10777.	0.000
.|*******	.| |	14	0.925	0.071	11553.	0.000
.|*******	.| |	15	0.920	-0.027	12322.	0.000
.|*******	.| |	16	0.916	0.009	13084.	0.000
.|*******	*| |	17	0.910	-0.069	13838.	0.000
.|*******	.| |	18	0.905	-0.006	14584.	0.000
.|******|	.| |	19	0.900	0.004	15323.	0.000
.|******|	.| |	20	0.894	-0.010	16053.	0.000
.|******|	.| |	21	0.889	-0.037	16775.	0.000
.|******|	.| |	22	0.883	0.014	17489.	0.000
.|******|	.| |	23	0.878	0.022	18195.	0.000
.|******|	.| |	24	0.872	-0.018	18894.	0.000
.|******|	.| |	25	0.868	0.025	19585.	0.000
.|******|	.| |	26	0.863	0.021	20270.	0.000
.|******|	.| |	27	0.858	-0.033	20948.	0.000
.|******|	.| |	28	0.853	0.011	21619.	0.000
.|******|	.| |	29	0.848	0.041	22283.	0.000
.|******|	.| |	30	0.844	-0.008	22942.	0.000
.|******|	.| |	31	0.840	0.001	23594.	0.000
.|******|	.| |	32	0.835	-0.040	24239.	0.000
.|******|	.| |	33	0.829	-0.021	24877.	0.000
.|******|	.| |	34	0.824	0.014	25508.	0.000
.|******|	.| |	35	0.819	-0.031	26132.	0.000
.|******|	.| |	36	0.814	-0.014	26748.	0.000

Figure 2 : The Correlogram of ONGC

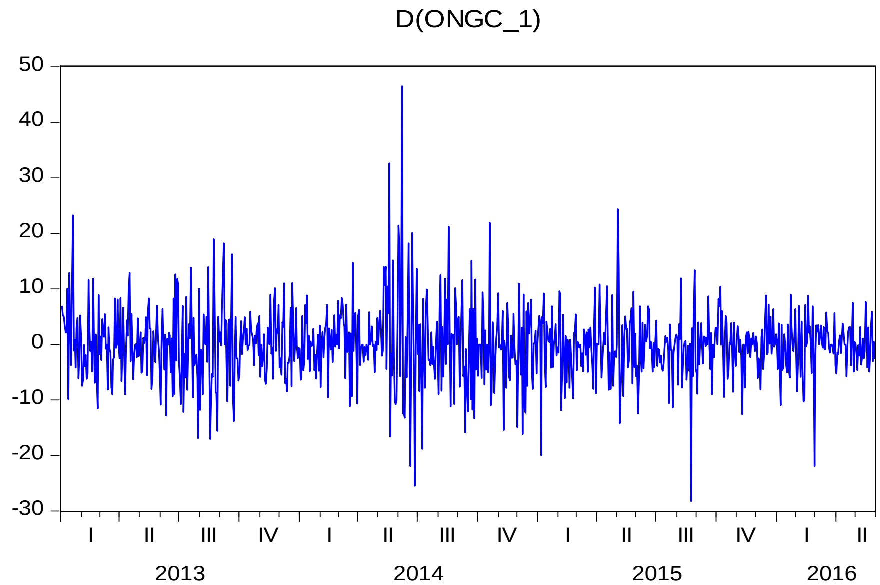

Figure 3: Graphical representation of the Stock closing price of DONGC

Date: 11/08/18 Time: 18:31
Sample: 1/01/2013 5/31/2016
Included observations: 890

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>.| |</td><td>.| |</td><td>1</td><td>-0.017</td><td>-0.017</td><td>0.2522</td><td>0.616</td></tr>
<tr><td>*| |</td><td>*| |</td><td>2</td><td>-0.099</td><td>-0.099</td><td>8.9389</td><td>0.011</td></tr>
<tr><td>.| |</td><td>.| |</td><td>3</td><td>-0.042</td><td>-0.046</td><td>10.550</td><td>0.014</td></tr>
<tr><td>.| |</td><td>.| |</td><td>4</td><td>0.029</td><td>0.018</td><td>11.299</td><td>0.023</td></tr>
<tr><td>.| |</td><td>.| |</td><td>5</td><td>0.022</td><td>0.015</td><td>11.744</td><td>0.038</td></tr>
<tr><td>.| |</td><td>.| |</td><td>6</td><td>-0.033</td><td>-0.030</td><td>12.704</td><td>0.048</td></tr>
<tr><td>.| |</td><td>.| |</td><td>7</td><td>-0.033</td><td>-0.030</td><td>13.709</td><td>0.057</td></tr>
<tr><td>.| |</td><td>.| |</td><td>8</td><td>-0.032</td><td>-0.039</td><td>14.627</td><td>0.067</td></tr>
<tr><td>.| |</td><td>.| |</td><td>9</td><td>-0.007</td><td>-0.018</td><td>14.668</td><td>0.100</td></tr>
<tr><td>.| |</td><td>.| |</td><td>10</td><td>0.056</td><td>0.048</td><td>17.503</td><td>0.064</td></tr>
<tr><td>.| |</td><td>.| |</td><td>11</td><td>0.013</td><td>0.012</td><td>17.646</td><td>0.090</td></tr>
<tr><td>.| |</td><td>.| |</td><td>12</td><td>-0.009</td><td>0.002</td><td>17.719</td><td>0.124</td></tr>
<tr><td>*| |</td><td>*| |</td><td>13</td><td>-0.072</td><td>-0.067</td><td>22.430</td><td>0.049</td></tr>
<tr><td>.| |</td><td>.| |</td><td>14</td><td>0.037</td><td>0.030</td><td>23.677</td><td>0.050</td></tr>
<tr><td>.| |</td><td>.| |</td><td>15</td><td>0.015</td><td>-0.002</td><td>23.887</td><td>0.067</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>16</td><td>0.090</td><td>0.094</td><td>31.167</td><td>0.013</td></tr>
<tr><td>.| |</td><td>.| |</td><td>17</td><td>-0.020</td><td>-0.006</td><td>31.524</td><td>0.017</td></tr>
<tr><td>.| |</td><td>.| |</td><td>18</td><td>-0.015</td><td>0.007</td><td>31.732</td><td>0.024</td></tr>
<tr><td>.| |</td><td>.| |</td><td>19</td><td>0.004</td><td>0.003</td><td>31.746</td><td>0.033</td></tr>
<tr><td>.| |</td><td>.| |</td><td>20</td><td>0.065</td><td>0.057</td><td>35.645</td><td>0.017</td></tr>
</table>

.| |	.| |	21	-0.008	-0.010	35.711	0.024
.| |	.| |	22	-0.040	-0.021	37.200	0.022
.| |	.| |	23	-0.007	0.007	37.249	0.031
.| |	.| |	24	-0.020	-0.028	37.630	0.038
.| |	.| |	25	-0.044	-0.051	39.410	0.033
.| |	.| |	26	0.042	0.028	41.024	0.031
.| |	.| |	27	-0.009	-0.011	41.107	0.040
*| |	*| |	28	-0.072	-0.071	45.945	0.018
.| |	.| |	29	-0.019	-0.010	46.287	0.022
.| |	.| |	30	0.026	-0.007	46.926	0.025
.| |	.| |	31	0.062	0.049	50.447	0.015
.| |	.| |	32	0.025	0.026	51.010	0.018
.| |	.| |	33	-0.039	-0.017	52.455	0.017
.| |	.| |	34	0.035	0.035	53.588	0.018
.| |	.| |	35	0.027	0.020	54.256	0.020
.| |	.| |	36	-0.015	-0.027	54.478	0.025

Figure 4 : The Correlogram of DONGC(after first differncing)

In figure 5 the model checking was done with Augmented Dickey Fuller (ADF) unit root test on "DONGC" of NSE index. The result confirms that the series becomes stationary after the first-difference of the series.

Null Hypothesis: D(ONGC_1) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=20)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-23.46046	0.0000
Test critical values:	1% level	-3.437491	
	5% level	-2.864582	
	10% level	-2.568443	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ONGC_1,2)
Method: Least Squares
Date: 11/08/18 Time: 22:24
Sample (adjusted): 1/04/2013 5/31/2016
Included observations: 888 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ONGC_1(-1))	-1.118378	0.047671	-23.46046	0.0000
D(ONGC_1(-1),2)	0.098917	0.033418	2.959967	0.0032
C	-0.085187	0.219051	-0.388890	0.6975

R-squared	0.513850	Mean dependent var	-0.008896
Adjusted R-squared	0.512751	S.D. dependent var	9.350446
S.E. of regression	6.526910	Akaike info criterion	6.593117
Sum squared resid	37701.49	Schwarz criterion	6.609296
Log likelihood	-2924.344	Hannan-Quinn criter.	6.599302
F-statistic	467.7128	Durbin-Watson stat	2.009761
Prob(F-statistic)	0.000000		

Figure 5: Augmented Dickey Fuller(ADF) Unit Root tests for differncing ONGC of NSE index

Table 2 shows the different parameters of autoregressive (*p*) and moving average (*q*) among the several ARIMA model experimented upon . ARIMA(1,1,1) is considered the best for ONGC, as shown in figure 6. The model returned the smallest Akaike information criterion 6.5907 and relatively smallest standard error of regression 6.5190 as shown in figure 6.

Dependent Variable: D(ONGC_1)
Method: Least Squares
Date: 11/12/18 Time: 16:20
Sample (adjusted): 1/03/2013 5/31/2016
Included observations: 889 after adjustments
Convergence achieved after 12 iterations
MA Backcast: 1/02/2013

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.122226	0.159915	-0.764316	0.4449
AR(1)	0.827155	0.077058	10.73420	0.0000
MA(1)	-0.874930	0.066938	-13.07078	0.0000

R-squared	0.012103	Mean dependent var	-0.071822
Adjusted R-squared	0.009873	S.D. dependent var	6.551479
S.E. of regression	6.519056	Akaike info criterion	6.590705
Sum squared resid	37653.31	Schwarz criterion	6.606870
Log likelihood	-2926.568	Hannan-Quinn criter.	6.596884
F-statistic	5.427539	Durbin-Watson stat	1.963713
Prob(F-statistic)	0.004541		

Inverted AR Roots	.83
Inverted MA Roots	.87

Figure6: ARIMA(1,1,1) estimation output with DONGC of NSE index

Figure 7 is the correlogram of residuals of the series. If the model is good, the residuals (difference between actual and predicted values) of the model are series of random errors. Since there are no significant spikes of ACFs and PACFs, it means that the residual of the selected ARIMA model are white noise, no other significant patterns left in the time series. Therefore, there is no need to consider any AR(p) and MA(q) further.

Date: 11/12/18 Time: 14:58
Sample: 1/01/2013 5/31/2016
Included observations: 890

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|*******	.|*******	1	0.995	0.995	884.41	0.000
.|*******	.|* |	2	0.992	0.141	1763.7	0.000
.|*******	.| |	3	0.989	0.046	2638.3	0.000
.|*******	.| |	4	0.985	-0.002	3508.1	0.000
.|*******	.| |	5	0.982	-0.013	4372.8	0.000
.|*******	.| |	6	0.978	-0.007	5232.5	0.000
.|*******	.| |	7	0.975	0.019	6087.6	0.000
.|*******	.| |	8	0.972	-0.029	6937.4	0.000
.|*******	.| |	9	0.969	0.032	7782.8	0.000
.|*******	.| |	10	0.965	-0.010	8623.3	0.000
.|*******	.| |	11	0.962	-0.032	9458.6	0.000
.|*******	.| |	12	0.958	0.021	10289.	0.000
.|*******	.| |	13	0.955	0.015	11115.	0.000
.|*******	.| |	14	0.952	-0.003	11937.	0.000
.|*******	.| |	15	0.949	-0.014	12753.	0.000
.|*******	.| |	16	0.945	0.002	13565.	0.000
.|*******	.| |	17	0.942	-0.030	14372.	0.000
.|*******	.| |	18	0.939	0.019	15174.	0.000
.|*******	.| |	19	0.935	-0.013	15971.	0.000
.|*******	.| |	20	0.932	0.017	16763.	0.000
.|*******	.| |	21	0.929	-0.013	17551.	0.000
.|*******	.| |	22	0.925	-0.007	18333.	0.000
.|*******	.| |	23	0.922	-0.007	19111.	0.000
.|*******	.| |	24	0.918	0.002	19884.	0.000
.|*******	.| |	25	0.915	-0.014	20652.	0.000
.|*******	.| |	26	0.911	0.018	21416.	0.000
.|*******	.| |	27	0.908	-0.011	22174.	0.000
.|*******	.| |	28	0.904	-0.020	22928.	0.000
.|******|	.| |	29	0.901	0.013	23676.	0.000
.|******|	.| |	30	0.898	0.022	24421.	0.000
.|******|	.| |	31	0.895	-0.001	25160.	0.000
.|******|	.| |	32	0.891	-0.011	25895.	0.000
.|******|	.| |	33	0.888	-0.017	26626.	0.000
.|******|	.| |	34	0.885	0.009	27351.	0.000
.|******|	.| |	35	0.881	0.009	28072.	0.000
.|******|	.| |	36	0.878	-0.025	28788.	0.000

Figure 7: Correlogram of Residuals of ONGC

In forecasting form, the best model selected can be expressed as follows:

$$x_t = \mu + \phi_1 X_{t-1} + \varepsilon_t - \theta_1 * \varepsilon_{t-1}$$

where, $\varepsilon_t = Y_t - Y_t^{\wedge}$ (i.e., the difference between the actual value of the series and the forecast value), μ is a constant, ϕ_1 is the first autoregressive parameter and θ_1 is the first moving average parameter. In table 2 the bold row represents the best ARIMA model among the several experiments.

Table-2 : Statistical results of different ARIMA parameters for ONGC, NSE index

ARIMA	AIC	SIC	Adjusted R^2	S.E of Regression
(1,0,0)	6.5989	6.6097	0.9886	6.5496
(1,0,1)	6.6009	6.6170	0.9886	6.5525
(2,0,0)	7.2737	7.2845	0.9776	9.1781
(0,0,1)	9.8426	9.8534	0.7081	33.1580
(0,0,2)	9.9130	9.9237	0.6868	34.3451
(1,1,0)	6.6003	6.6111	-0.0008	6.5542
(0,1,0)	6.5985	6.6039	0.0000	6.5520
(0,1,1)	6.6004	6.6112	-0.0007	6.5545
(1,1,2)	6.5927	6.6088	0.0078	6.5256
(2,1,0)	6.5912	6.6020	0.0086	6.5244
(2,1,2)	6.5934	6.6096	0.0075	6.5280
(1,1,1)	**6.5907**	**6.6068**	**0.0098**	**6.5190**
(3,1,1)	6.6013	6.6175	0.0001	6.5537

6. RESULTS AND DISCUSSIONS

The experimental results of each of stock index are discussed in the below sections.

A. Result of ARIMA Model for NSE Price Prediction of companies ONGC.

Figure 8 gives graphical illustration of the level accuracy of the predicted price against actual stock price to see the performance of the ARIMA model selected. Table 3 is the result of actual and fitted values of ARIMA(1,1,1) considered the best model for ONGC, NSE index.

From the graph, it is obvious that the performance is satisfactory. From the graph of forecast of ONGC for ARIMA(1,1,1) we get RMSE as 70.3723, MAE as 48.1688 and MAPE as 13.7205.

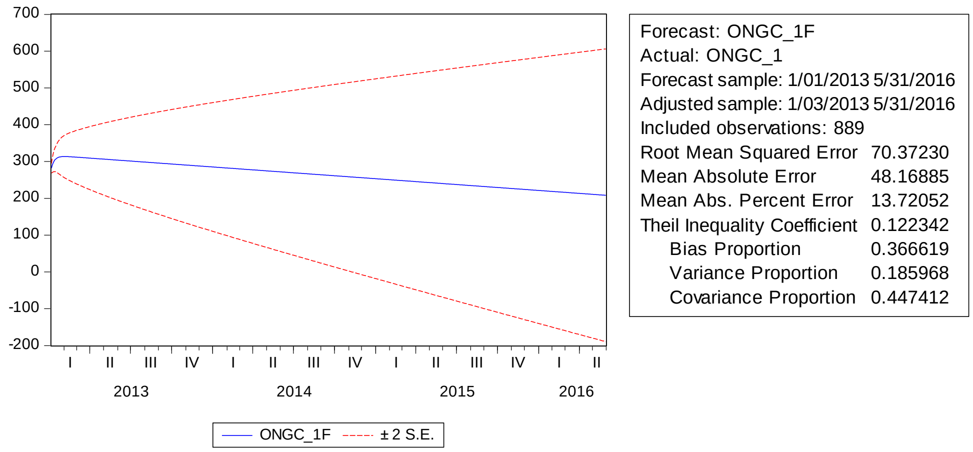

Figure 8: Graph of forecast of ONGC

Table3: Sample of Emperical results of ARIMA(1,1,1) of ONGC at NSE index

obs	Actual	Fitted	Residual	Residual Plot
1/03/2013	5.20000	7.19691	-1.99691	.*. |
1/04/2013	5.00000	6.02724	-1.02724	.*. |
1/07/2013	3.20000	5.01341	-1.81341	.*. |
1/08/2013	2.10000	4.21237	-2.11237	.*. |
1/09/2013	2.10000	3.56408	-1.46408	.*. |
1/10/2013	10.0500	2.99686	7.05314	.|* |
1/11/2013	-9.90000	2.12078	-12.0208	* .|. |
1/14/2013	12.9000	2.30738	10.5926	.|.* |
1/15/2013	2.30000	1.38138	0.91862	.*. |
1/16/2013	-3.75000	1.07760	-4.82760	*|. |
1/17/2013	10.6500	1.10085	9.54915	.|.* |
1/18/2013	23.2500	0.43324	22.8168	.|. * |
1/21/2013	-1.15000	-0.75283	-0.39717	.*. |
1/22/2013	0.85000	-0.62486	1.47486	.*. |
1/23/2013	-4.20000	-0.60844	-3.59156	*|. |
1/24/2013	3.15000	-0.35281	3.50281	.|* |
1/25/2013	4.75000	-0.48030	5.23030	.|* |
1/28/2013	-6.15000	-0.66829	-5.48171	*|. |
1/29/2013	-1.65000	-0.31202	-1.33798	.*. |

B. CONCULSION

This paper presents extensive process of building ARIMA model for stock price prediction. The experimental results obtained with best ARIMA model demonstrated the potential of ARIMA models to predict stock prices satisfactory on short-term basis. This could guide investors in stock market to make profitable investment decisions.

REFERENCES

1. Time Series Analysis- Forecasting and Control (Third Edition): George E.P.Box, Gwilym M. Jenkins, Gregory C. Riensel.
2. Applied Multivariate Statistical Analysis (Third and Fifth Edition): Richard A.Johnson, Dean W. Wichern.
3. Dongping Men, Jun Wang ; Jiguang Shao: The Statistical Analysis of Stock Prices and Trading Volumes for the Chinese Stock Markets, *ISECS International Colloquium*, pp. 37-41 (2008).
4. G. Preethi, B. Santhi: Stock Market Forecasting Techniques: A Survey *Journal of Theoretical and Applied Information Technology*. Vol. 46 No.1 (15th December 2012).
5. Xiangyu Tang, Chunyu Yang, Jie Zho, "Stock Price Forecasting by Combining News Mining and Time Series Analysis", International Conference on Web Intelligence and Intelligent Agent Technology- Workshops -IEEE/ WIC/ACM. (2009).
6. Introduction to Mathematical Finance: Ross S.M (2004).
7. Introduction to Time Series And Forecasting (Second Edition- Indian Print)SPRINGER: Brockwell, P.J and Davis, R.A (2002).

8. Introduction to Futures and Options, Markets(Second Edition): John C.Hull.
9. Jun Wang, Bingli Fan and Dongping: Data Analysis and Statistical Behaviors of Stock Market Fluctuations, *Journal of Computers*, vol. 3(10) (2008).
10. Kalok Chan , K. C. Chan and G. Andrew Karolyi : Intraday Volatility in the Stock Index and Stock Index Futures Markets, *Oxford Journals*, Volume 4, Issue 4, pp.657-684 (2014).
11. Pankratz, Alan New York: John Wiley & Sons, “Forecasting with Univariate Box–Jenkins models: concepts and cases” (1983).
12. Puspanjali Mohapatra (Asst. Prof), Alok Raj , Tapas Kumar Patra (Reader): Indian Stock Market Prediction Using Differential Evolutionary Neural Network Model *International Journal of Electronics Communication and Computer Technology (IJECCT)* Volume 2, Issue 4 (July 2012).
13. Time Series Application to Finance: Ngai Hang Chan.
14. An Introduction To Multivariate Analysis (Third Edition): Anderson, T.W, John Wiley, New York (2004).
15. Ayodele A. Adebiyi O. Adewumi, Charles K. Ayo, "Stock price prediction using the ARIMA Model" AMSS 16th International conference on computer modelling and simulation, UK Sim. (2014).
16. B. Uma Devi, D. Sundar and Dr. P. Ali, "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMaA Model for NIfty Midcap-50" *International Journal of Data Management Process (IJDKP)* Vol.3, No.1, January (2013).