Time-Varying Approach-Based Indian Stock Price Prediction Using ARIMA and Monte Carlo Simulation

T P Sameerapallavi

Dept. of CSE
Kakatiya University,
Warangal, Telangana, India.
Sameerapallavi33@gmail.com

Abstract:
The stock market is influenced by several factors such as economic and business policies, events that affect the economy of a country, natural disasters, climatic conditions, and also the psychological states of investors. As a result, the market is dynamic, volatile, and complex. The choice of predictor variables determines the interpretation of results and the level of forecast performance. There are generally four categories of information for stock market prediction: basic information, technical information, behavioural information, and information on the history of the stock return series. This paper proposes a stock prediction approach using commodities. Crude oil is considered to be a major commodity in India. In the proposed model, the Autoregressive Integrated Moving Average (ARIMA) model is used to forecast the stock price of companies based on international crude oil prices. The ARIMA parameters, such as the degree of integration, autoregressive order, and moving average order, have been tuned to predict the stock values. In order to tune these parameters, exploratory data analysis has been performed, and the data has been converted into stationary data. Monte-Carlo simulation is used to forecast the stock price based on the results of ARIMA. The experimental analysis proves that the proposed algorithm has predicted the stock values very efficiently.

Keywords: Autoregressive Integrated Moving Average, Monte-Carlo simulation, Stock price, Crude oil.

I. Introduction

The stock market is a place of exchange on which products and financial instruments (stocks, bonds, options, futures contracts, etc.) with the potential to make big profits [1]. However, empirical studies have shown that predicting the market condition is a very difficult task because of its complex, volatile, and non-linear nature. Moreover, several factors can affect the behaviour, including the economic and business conditions, political events, expectations of market participants, and their psychological state [2]. Therefore, the stock market is said to have a dynamic function whose data is volatile, irregular, and noisy; hence the enormous difficulty of modelling time series stock market predictions [3].

The wide fluctuations in stock market prices observed in recent years have led the financial community to pay renewed attention to the concept of stock market data prediction [4]. While the volatility of equity returns is frequently used by investors as a risk measurement instrument, its estimation nonetheless raises difficulties, and its interpretation calls for various precautions [5]. However, the analysis of various indicators available in this area tends to show that stock market volatility has recorded, since 1997, a marked upward trend.

However, in a very long-term perspective, the current level of market volatility does not appear to be unprecedented or extraordinarily high.

The prices of financial assets have suffered during recent years due to very strong fluctuations [6]. These spectacular movements have triggered the interest of researchers in the issue of stock market prediction. Price fluctuations are inherent in the very existence of markets [7]. All stakeholders must assume that they are exposed to the risk of loss. The study of volatility of the stocks does not reveal the complete information about the market impact suffered by financial players. Volatility is an approximate and biased indicator of risk, which is an empirical volatility calculated from past prices (historical volatility) [8]. In all cases, this volatility must be estimated because it is not directly observable.

In general, the modeling of these series requires data collection, analysis of characteristics and their relationships with the series in question, and finally the production of a system that facilitates prediction of future events or situations. [9] Two families of methods are used to model stock market time series: statistical methods and the methods of artificial intelligence (AI). Statistical methods are traditionally the most used; especially the Autoregressive Integrated Moving Average process (ARIMA) [10]. They are based on formalisms mathematics and offer the possibility of making statistical inference in addition to prediction.

From the perspective of statistical models, the design of ARIMA models have been made. Based on two perspectives, such as statistical and artificial intelligence techniques [11], prediction can be done generally. When compared to the widely used techniques of Ann's [1214], ARIMA models have shown efficient and robust outcomes in forecasting of financial time series, specifically in short-term prediction. In the field of finance and economics, these models have been utilized extensively. Other relevant statistics models include Generalized Autoregressive Conditional Heteroskedasticity (GARCH), exponential smoothing, and regression techniques. In [15-20], a few other related works are included that engage the ARIMA model for forecasting.

To predict the short-term stock price, an extensive process of building ARIMA models is demonstrated in this paper. The potential strength of ARIMA models is described by the obtained results from real-life data for providing short-term prediction details for investors that could be useful in decision-making in the investment process.

II. Proposed ARIMA and Monte Carlo simulation model

The proposed model is a combination of ARIMA and Monte Carlo simulation to predict stock market prices. The proposed algorithm is shown in Figure 1 below:

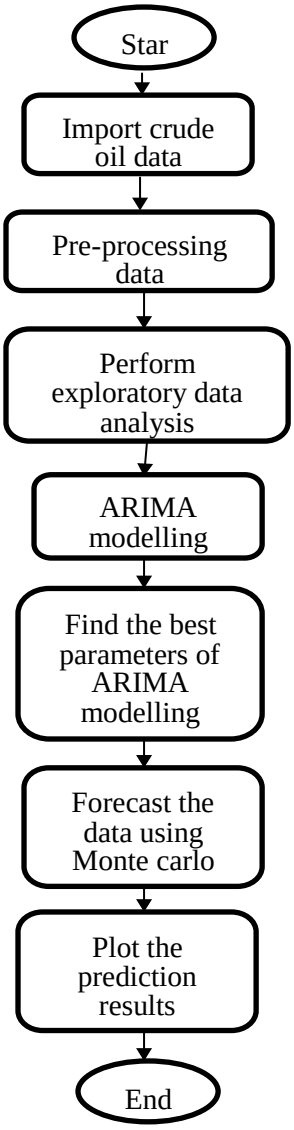


Figure1: Proposed ARIMA and Monte Carlo-based simulation

2.1 ARIMA model

2.1.1 Time series analysis

Time series analysis is a discipline that deals with the statistical analysis of time series. It serves, among other things, to make predictions for the future development of time series. The task is to find a suitable mathematical model for a given time series. To be able to judge its quality. In this work, the ARIMA model is described, which will be discussed below.

2.1.2 Time series

A time series describes a sequence of data over time. Examples are weather observations, economic trends, or fluctuating populations. Typically, the data are stochastic of a similar character or at least have stochastic properties. The goal of time series analysis is the precise mathematical description of existing data in the form of a time series. To describe the stochastic character of the data, a time series *X is used* as a set of random variables, understood:

$X = \{x_1, x_2, \ldots \ldots, x_n\} = \{x_t\}\ with\ t \in R\ and\ t \in \{t_1, t_2, \ldots \ldots, t_n\}$ (1)

The random variables *x_t* are time-dependent, with the points in time *t* at which the measurement takes place a dis- represent a concrete amount. The time intervals $t_i - t_{i+1}$ between two neighboring Random variables do not necessarily have to be equidistant. In this work, however, an equidistance is assumed between times t_i. Time series with x_t *are displayed* on the Ordinate and *t* on the abscissa. Usually the temporally adjacent points

are marked with lines connected to possible trends and possibly existing continuous patterns such as a signal that is overlaid by noise. The prerequisite for this, however, is that enough measurement data is available within a unit of time. There will now be some typical ones. Examples of simple time series are described.

White noise. A time series from independent normally distributed random variables is used in this Work called *white noise*. The white noise is noted as w_t and has, if not otherwise mentioned, the expected value $< \omega_t >= \mu_w = 0$ and the variance σ_w^2. It provides the foundation for many of the discussed stochastic models. In general, however, one understands by white noise a signal with a constant power density spectrum.

Moving average (MA). A *moving average* (MA) highlights the coarse oscillations and partially removes the small oscillations. An example is

1

$v_t = {}_3_ (w_{t-1} + w_t + x_{t+1})$ (2)

White noise *w t* can be replaced by a moving average that results in a smoothing of white noise leads.

Autoregression (AR). Is a time series of deterministic character or at least carries one deterministic component in itself, so the description by an *autoregression* is suitable. A Example is

$$x_t = x_{t-1} - 0.4x_{t-2+wt} \quad (3)$$

The value of the time series at time *t* depends on the previous two values. The stochastic properties come about through w_t.

Random walk with drift. *Random* walks are suitable for describing time series with a trend. For example, a random walk is carried out by described.

$x_t = \delta + x_{t-1} + w_t$ (4)

δ is called *drift* and parameterizes the slope of the trend. Eq. (4) can also be used as to be written.

$$x_t = \delta t + \sum_{j=1}^{t} wj \quad (5)$$

For $\delta = 0$, there is a simple random walk; its random movement alone is determined by w_t. Nevertheless, the variance of the random walk x_t *increases* with time.

2.1.3 ARIMA models

The ARIMA models are used to model time series of different characteristics. They include both purely deterministic time series and time series of purely stochastic character. They also contain models that have both deterministic and stochastic properties and also describe time series that show a trend, corresponding to $\delta \neq 0$ according to Eqs. (4) and (5). In the following, the partial models of the ARIMA model are presented so that they can then be so that the ARIMA model can be explained. Before doing this, it should be mentioned that a time series in general from a deterministic process *f* and a stochastic process *g* consists of :

$\Delta^d[time\ series\ X = f(x_{t-1},\dots, x_{t-p}) + w_t_ + g(w_{t-1}, \dots, w_{t-q}) + AR(p)\ w_t]___$ (6)

$MA(q) ARMA(p,q) ARIMA(p,d,q)$

2.1.3.1 Autoregressive models (AR)

The autoregressive models (AR models) are used to describe deterministic time series Character. The value x_t of the time series at time *t* should be taken from the previous time series values x_{t-j} can be calculated. An example of this has already been given in Eq. (3) delivered. The general Definition of an AR model is

$x_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \cdots + \emptyset_p x_{t-p} + w_t = \sum_{j=1}^{p} \emptyset_j x_{t-j} + w_t$ (7)

A model according to Eq. (7) from an *autoregressive model of order p* , shortened AR (*p*). The constants $\emptyset_j \neq 0$ are also called AR coefficients in the following. It will generally assumed that w_t is a white noise with the expected value $\mu_w = 0$ and the variance σ^2_w corresponds to. The expected value of Eq. (7) is $\mu_x = 0$. If the expected value $\mu_x \neq 0$, then Eq. (7) to be rewritten.

$x_t - \mu_x = \sum_{j=1}^{p} \emptyset_j (x_{t-j} - \mu_x) w_t$ (8)

The borderline case $w_t = 0$ leads to a completely deterministic time series without a random component.

2.1.3.2 Moving Average Models (MA)

In contrast to the AR models, the moving average models (MA models) assume that the time series at time *t* consists solely of white noise w_t at the current time *t* and is dependent on earlier noise terms w_{t-j} at times *t - j* . It follows the general definition.

$$xt = wt + \theta 1wt-1 + \theta 2wt-2 + \cdots + \theta qwt-q = \sum qk=1\theta kwt-k + wt \quad (9)$$

A model according to Eq. (9) from a *moving average model of order q*, abbreviated MA (*q*). The constants $\theta_k \neq 0$ are also called MA coefficients in the following. Also, here it is assumed that w_t is white noise. This enables the MA model can be used for time series of stochastic character. Since, by definition, the expected value of a white noise $\mu_w = 0$, it follows that the expected value of Eq. (2.8) $\mu_x = 0$.

2.1.3.3 Combination of AR and MA: ARMA models

ARMA models are a combination of AR (*p*) and MA (*q*) models. They are therefore suitable for Time series that have both a deterministic and a stochastic component. The Combination of equations (7) and (9) yields

$$x_t = \sum q_{k=1} \quad \theta_k w_{t-k} + \sum p_{j=1} \emptyset_j x_{t-j} + w_t \quad (10)$$

With $\emptyset \neq 0$ *and* $\theta \neq 0$. One speaks of an *ARMA (p, q) model of the AR order p and the MA Order q*. ARMA (*p, q*) models according to Eq. (2.9) have the expectation $\mu_x = 0$. However, if $\mu_x \neq 0$, can Eq. (10), analogous to Eq. (8), be modified? This will be

$$\alpha = \mu_x(1 - \sum p_{j=1} \emptyset_j) \quad (11)$$

Defined, with which Eq. (10)

$$x_t = \alpha + \sum q_{k=1} \quad \theta_k w_{t-k} + \sum p_{j=1} \emptyset_j x_{t-j} + w_t \quad (12)$$

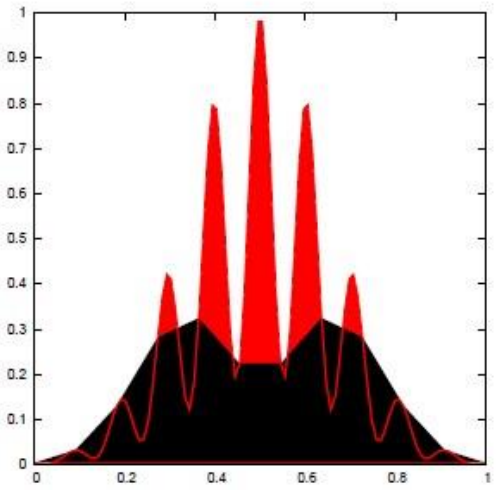
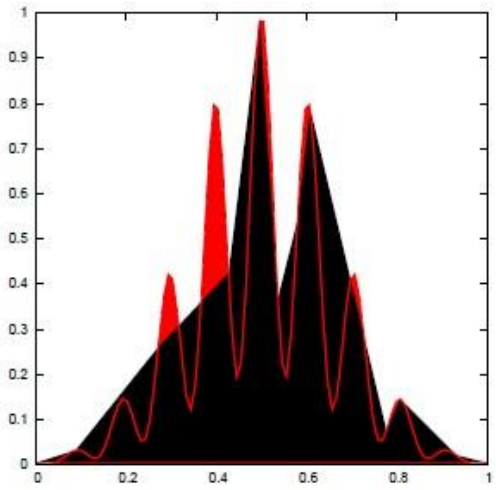
Can be written. The ARMA (*p, q*) models include the MA (*q*) models for $p = 0$ and, for $q = 0$, the AR (*p*) models. They form the basis for all ARIMA time series analyzes, as they are both AR and MA models. In the case of non-stationary time series, stationarity is generated to then be able to perform an ARMA (*p, q*) analysis.

2.2 Monte Carlo Simulation

If the random variable intervention is present, the probability of different outcomes is predicted by using a model known as Monte Carlo Simulation, which assists in demonstrating the risk and uncertainty impact in forecasting models. Monte Carlo Simulations are being exploited in different fields such as supply chain, engineering, finance, and science. To get multiple outcomes and obtain an estimated value through the averaging of results, multiple values are assigned to an uncertain variable in the Monte Carlo Simulation.

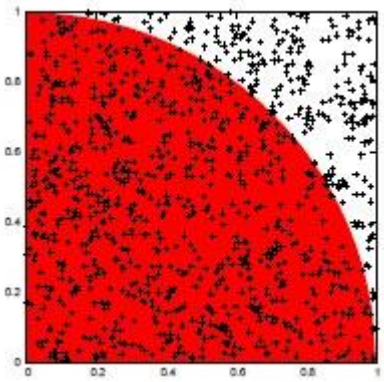
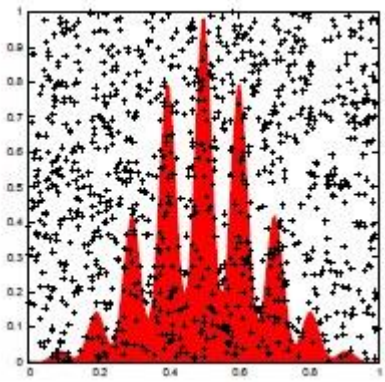
Numerous kinds of potential applications of Monte Carlo simulations are involved in the business and finance fields, as they include random variables. In large projects, they have been utilized for computing the cost overrun probability and the likelihood of an asset price moving in a certain way.

Monte Carlo is exactly the right method with a relatively small number of iterations; nevertheless, it is sufficient to get the exact result. However, there are some important things to consider. To do this, one first considers, starting from a strongly oscillating function, the method of numerical squaring over fixed resp. random support points to calculate the integral and thus the area. As you can see in Figure 2, the problem has remained the same.

(a) Fixed interval width points **(b) Random interval width Figure2: Influence of the support**

Man has random support points, but they can also be very bad. In addition, you also have to calculate all of these, which again becomes the problem of many iterations. That's why they like the way to deal with two-dimensional problems is helpful, but get started quickly with higher-dimensional calculations; their limits.
Another approach would be to choose no support points, but rather the radiation with randomly placed points. In the following Figure 3, see how one tries to determine the area under the function and, with it, the integral. If now look at both functions graphically you can see very quickly why the MC method is a relatively large error.

(a) Section of a circle **(b) More complicated Fct Figure3: Two different functions in comparison**

The area is determined by the number of Points below the function, calculated against the total points.
And from Figure 4, it can be seen that, although in the left Graph 3 there are many points below the function, the probability of the function being met is also very high. On the other hand, on the left the probability is much greater to miss the function and thus this is no longer adequately covered with points, which in turn leads to a large statistical error. If we look further at the left figure 4, results after analytical integration for the interval from 0 to 1 the value $f(x)|^1_0 =$
0.237343 but you get with the simple MC method $f(x)|^1_0 = 0.255 \pm 2.10^4$ at 1000 points.

2.2.1 Sampling

Now you can consider whether you might not be able to use the points more skilfully, since the outer areas contribute significantly less than the inner ones. And with this consideration now comes the so-called

weighting function comes into play, since now the points that are less often, are more important than the points, that one preferred.

From a mathematical point of view, this is a substitution of the shape

$$F|^1_0 = \int \quad f(w)\ dw = \int_w ___ f(x_x)\, w(x)dx \qquad (13)$$

In Figure 4 you can now see how the points after the function

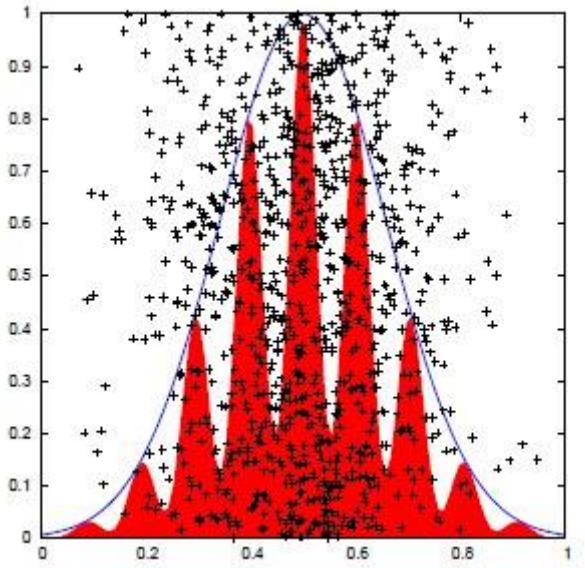
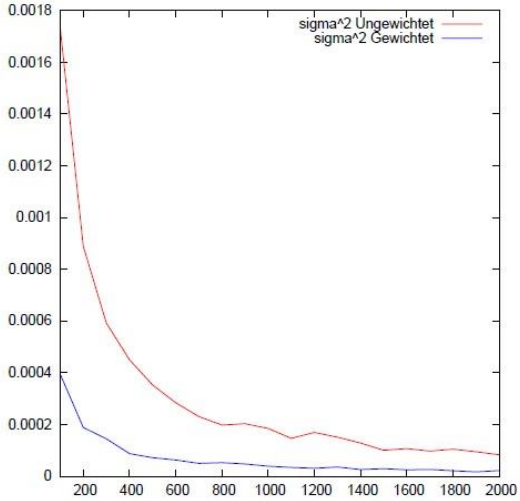
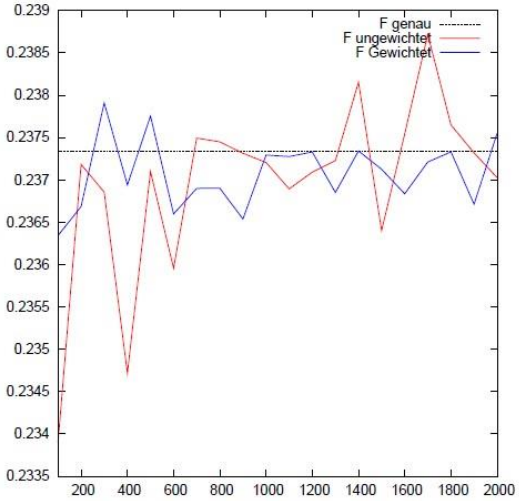


Figure 4: Weighted distribution

$$1\ 2w(x) = e^{-20(x^-{}_2)} \qquad (14)$$

Were weighted. We then get for the integral: $F(x)|^1_0 = 0.236635 \pm 2.10^5$ at 1000 points





(a) variances **(b) deviations Figure5: Both methods in comparison**

Figure 5 shows the variance of the methods plotted against the number of sample points. For this purpose, 300 runs were carried out for each number of points, and over that the variance and the integral are averaged. But since one is less concerned with values in statistical physics certain observables in certain places or states than for their Interested in mean values, one can use the Metropolis scheme since it for this it is not necessary to use the integral carry out operations.

III. Experimental Results

This section presents the experiments performed using the proposed algorithm. A sample of stock market data used for input is shown in table 1.

Table1: BPCL Stock mark data

Date	Open	High	Low	Close	Adj Close	Volume
27-02-2017	476	483.2	467.267	471.8	437.2352	3279614
28-02-2017	460	463.267	445.333	446.7	431.8217	7521292
01-03-2017	450	451.267	435.6	439.633	424.9901	5482936
02-03-2017	441.3	443.6	420.567	423.433	409.3297	8909256
03-03-2017	422.6	423.6	410.167	421.867	407.8158	11837120
06-03-2017	424.867	428.933	421.6	425.033	410.8763	4947909
07-03-2017	427.333	432.467	425.867	430.6	416.2579	3454140
08-03-2017	430.333	433.567	420.067	426	411.8112	3223611
09-03-2017	429.433	429.6	420.3	421.333	407.2996	3155532
10-03-2017	423.933	424.633	413.4	415.567	401.7256	3653210
14-03-2017	415.567	427.333	415.567	424.467	410.3292	4683009
15-03-2017	423.5	429.233	421.9	424.333	410.1997	2256408
16-03-2017	426	432.933	424.7	431.833	417.4499	2812221
17-03-2017	433.333	433.333	425.367	428.067	413.8093	3351987
20-03-2017	426.667	432	423.567	429.067	414.776	1791368
21-03-2017	428	432.667	422.533	427.933	413.6798	2790560
22-03-2017	424.433	432.333	418.7	430	415.6779	2946236
23-03-2017	436.667	440	432.067	436.7	422.1548	5743254
24-03-2017	433.333	434.667	430.067	433.133	418.7065	2752752

The data for the year 2017 is given as input to the ARIMA model. Figure 6 shows the input data and the ARIMA model output.



Figure6: Bharat Petroleum Stock Price in the year 2016

To determine whether an MA model is appropriate, the series with serial correlation and AR lags is identified.

Figure 7 represents the partial autocorrelation function.

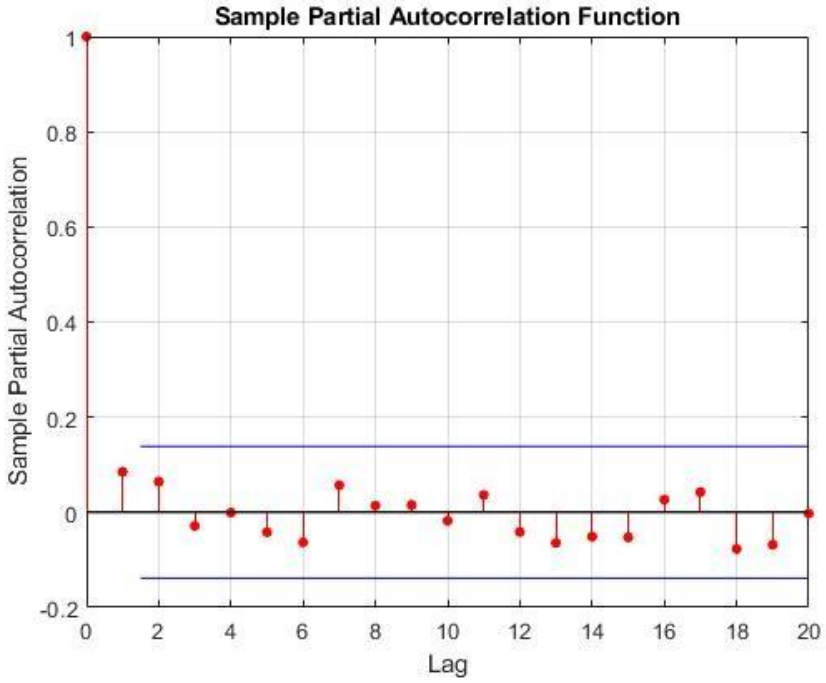


Figure7: Sample Partial auto correlation function

The result of the ARIMA model is given to the Monte Carlo Simulation to predict the future stock market data. The result is shown in figure 8.

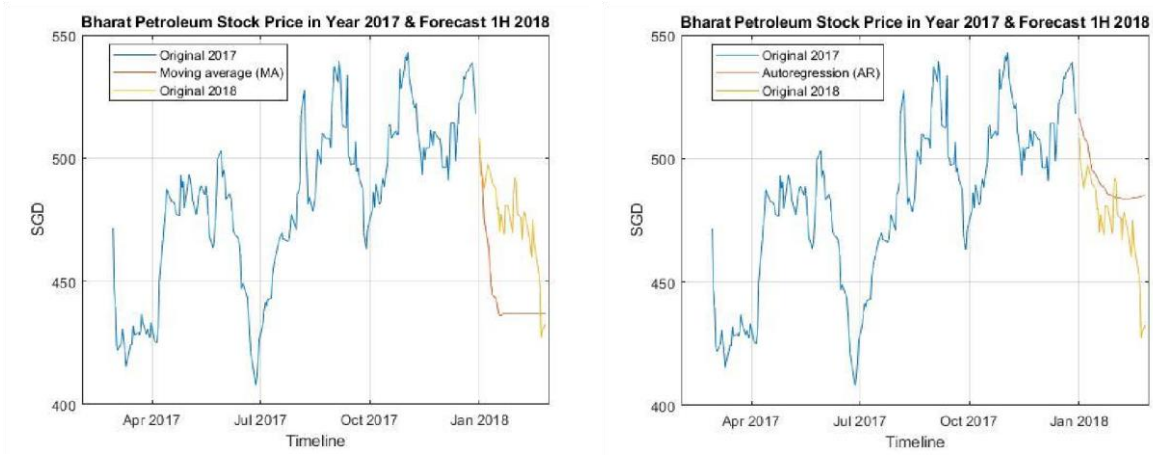


Figure8: Simulated prediction values.

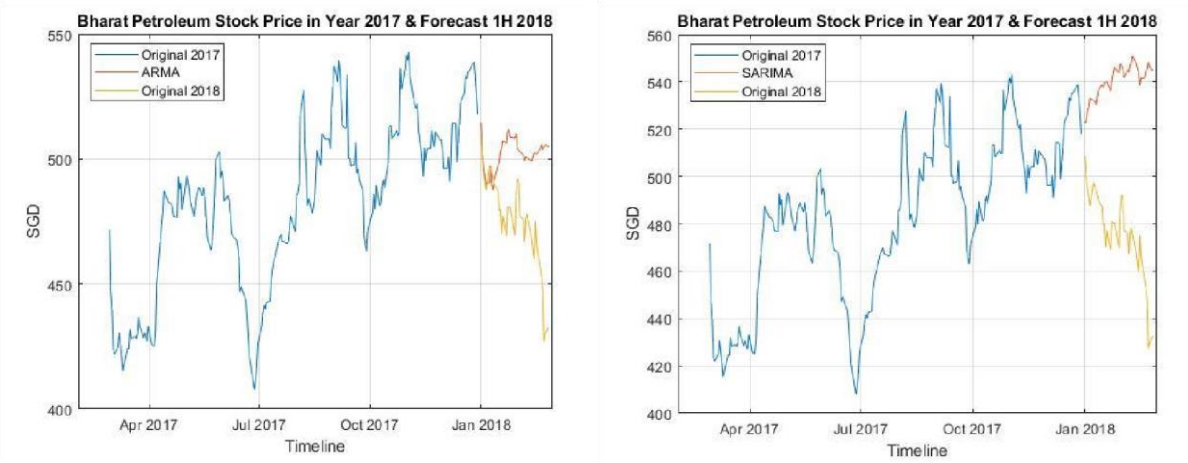
The proposed method is compared with the following models.

a. Moving Average
b. Auto-regression
c. ARMA
d. SARIMA

The proposed method produced better results when compared to these existing models. The experiments have also been performed on the stock data of India Oil Corporation, ONGC, and Reliance Industries. Figure 9 represents the comparative results. The error and accuracy values are tabulated in Table 2.



a. Moving Average b. Autoregression



c. ARMA d. SARIMA

Figure 9: Comparative Results

Table2: Accuracy comparison

	BPCL		IOC		ONGC			
	Mean Error	Accuracy (%)	Mean Error	Accuracy (%)	Mean Error	Accuracy (%)	Mean Error	Accuracy (%)
Moving Average	30.35	93.65	28.54	94.05	31.67	93.78	30.78	93.47
Auto regression	18.37	95.36	18.65	95.66	19.54	95.48	19.75	95.68
ARMA	26.70	94.21	25.48	94.58	27.85	94.94	26.54	94.58
SARIMA	64.23	88.28	66.98	87.05	65.98	88.65	64.65	88.31
ARIMA – Monte Carlo	16.262	96.78	17.96	96.44	16.58	97.65	18.45	96.85

Conclusion

Autoregressive Integrated Moving Average (ARIMA) model is used to forecast the stock price of companies based on international crude oil prices. The ARIMA parameters, such as the degree of integration, autoregressive order, and moving average order, have been tuned to predict the stock values. In order to tune these parameters, exploratory data analysis has been performed, and the data has been converted into stationary data. Monte-Carlo simulation is used to forecast the stock price based on the results of ARIMA. The experimental analysis proves that the proposed algorithm has predicted the stock values very efficiently.

REFERENCES:

1. Junttila, Juha, Juho Pesonen, and Juhani Raatikainen. "Commodity market-based hedging against stock market risk in times of financial crisis: The case of crude oil and gold." Journal of International Financial Markets, Institutions and Money 56 (2018): 255-280.
2. Jin, Xiaoye, and Ximeng An. "Global financial crisis and emerging stock market contagion: A volatility impulse response function approach." Research in International Business and Finance 36 (2016): 179-195.

3. Liu, Siyuan, Guangzhong Liao, and Yifan Ding. "Stock transaction prediction modeling and analysis based on LSTM." In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 2787-2790. IEEE, 2018.
4. Ding, Zhihua, Zhenhua Liu, Yuejun Zhang, and Ruyin Long. "The contagion effect of international crude oil price fluctuations on Chinese stock market investor sentiment." Applied Energy 187 (2017): 27-36.
5. Lin, Zhe. "Modelling and forecasting the stock market volatility of SSE Composite Index using GARCH models." Future Generation Computer Systems 79 (2018): 960-972.
6. Corbet, Shaen, Andrew Meegan, Charles Larkin, Brian Lucey, and Larisa Yarovaya. "Exploring the dynamic relationships between cryptocurrencies and other financial assets." Economics Letters 165 (2018): 28-34.
7. Zhang, Xi, Yunjia Zhang, Senzhang Wang, Yuntao Yao, Binxing Fang, and S. Yu Philip. "Improving stock market prediction via heterogeneous information fusion." Knowledge-Based Systems 143 (2018): 236-247.
8. Lahmiri, Salim. "Modeling and predicting historical volatility in exchange rate markets." Physica A: Statistical Mechanics and its Applications 471 (2017): 387395.
9. Vlasenko, Alexander, Olena Vynokurova, Nataliia Vlasenko, and Marta Peleshko. "A Hybrid Neuro-Fuzzy Model for Stock Market Time-Series Prediction." In 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), pp. 352-355. IEEE, 2018.
10. Bakar, Nashirah Abu, and Sofian Rosbi. "Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction." International Journal of Advanced Engineering Research and Science 4, no. 11 (2017): 237311.
11. J.J. Wang, J.Z. Wang, Z.G. Zhang and S.P. Guo, “Stock index forecasting based on a hybrid model”, Omega vol.40 pp.758-766, 2012.
12. L.C. Kyungjoo, Y. Sehwan and J. John, “Neural Network Model vs. SARIMA Model in Forecasting Korean Stock Price Index (KOSPI), Issues in Information System, vol. 8 no. 2, pp. 372-378, 2007.
13. N. Merh, V.P. Saxena, and K.R. Pardasani, “A Comparison Between Hybrid Approaches of ANN and ARIMA For Indian Stock Trend Forecasting”, Journal of Business Intelligence, vol. 3, no.2, pp. 23-43, 2010.
14. J. Sterba and Hilovska, “The Implementation of Hybrid ARIMA Neural Network Prediction Model for Aggregate Water Consumption Prediction”, Aplimat- Journal of Applied Mathematics, vol.3, no.3, pp.123-131, 2010.
15. C. Javier, E. Rosario, J.N. Francisco and J.C. Antonio, “ARIMA Models to Predict Next Electricity Price”, IEEE Transactions on Power Systems vol. 18 no.3, pp.1014-1020, 2003.
16. N. Rangan and N. Titida, “ARIMA Model for Forecasting Oil Palm Price”, Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, Universiti Sains Malaysia, 2006.
17. M. Khasel, M. Bijari, and G.A.R. Ardali, “Improvement of Auto Regressive Integrated Moving Average models using Fuzzy logic and pp. 956-967, 2009.
18. C. Lee, C. Ho, “Short-term load forecasting using lifting scheme and ARIMA model”, Expert System with Applications, vol.38, no.5, pp.5902-5911, 2011.
19. M. Khashei, M. Bijari, G. A. R. Ardal, “Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks”, Computers and Industrial Engineering, vol. 63, no.1, pp.37-45, 2012

20. C. Wang, “A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan Export”, Expert System with Applications, vol.38, no.8, pp.9296-9304, 2011.