International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-05, Issue-12, Mar 2020

Survey on Classical Time Series Forecasting Methods

*Ms. Ruchi Singh, #Dr. Jyoti Kharade

*Student, #Professor, Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, India. *ruchi191096@gmail.com, #kharadejyoti09@gmail.com

Abstract— **Stock market or stock price have always been the talk of the town. Be it a small vendor to big level CEO they are always interested in the stock market and making investment in them to earn profit. Stock market plays an important role for anyone who invests money for short-term duration or long-term duration. A Stock market is a structure where buying and selling of stocks are carried out. The investor goes end to end of investigation regarding "In which company to invest" or "Suitable time to invest". The investor may have knowledge about where to invest his or her money but the main problem arises whether the investment in shares will give a good amount of returns in terms of dividend. To solve this circumstance statistical time series forecasting methods such as Moving average (MA), ARIMA, SARIMA were introduced which will help the investor in making investments. This paper attempts to analyze the closing price of data as the profit and loss calculation is usually done on it.**

Keywords—ARIMA, Moving Average, SARIMA, Stock Prediction, Stock Price, Time Series Forecasting.

I. INTRODUCTION

Time is an important factor when it comes to prediction of any trend in the stock market. This should not be neglected.

Time series is defined as the number of data points in subsequent form. The independent variable in time series is always the time which is helpful for forecasting. It traces the movement of the data points.

Time series analysis can be useful to see how the chosen data points change over the period of time interval. In this paper we are analyzing the closing price of the stock of about five years. We would obtain the entire list of the closing price from each day. Using the time series analysis, we can obtain information about the stock whether it's going up and down.

Time series forecasting uses information and knowledge about the historical data and the correlated pattern which is helpful to predict the future.

There are two ways of analysis: Fundamental analysis as well as Technical analysis. Analyzing the company's future benefit on the basis of its present financial growth and environment comes under the fundamental analysis, whereas reading the plots and graphs to identify the trends comes under Technical analysis.

The objectives of the research paper are:
1)To Study the Moving Average method (MA)
2)To Study the ARIMA method
3)To Study SARIMA method
4)Compare MA, Auto Arima, SARIMA method.

The purpose of carrying out this study is to understand the statistical method of time series forecasting (Moving Average, ARIMA, SARIMA), their performance on the stock data and which method among them generates reasonable results. This will help in having a better approach for the traditional methods for forecasting stock prices.

In this study, Section I contains the introduction of the study. Section II contains related work of the study that describes the previous research works. Section III discusses the three different methods in detail. Section IV contains comparison of the results between the three approaches. Section V contains the findings related to the model. Section VI contains conclusion about the paper.

II. RELATED WORK

S. Gour [1] helps the investor with the obtained historical data for buying and selling of the stock. With the analysis of the data, it helps to predict the long run results. In this paper he used the model of decision tree algorithm which is one of the best and effective data mining techniques. Using the decision tree model better results were obtained for the forecasting.

Y. X. Lu, T. Zhao [2] they studied about prediction of changes of balance. The methods used are first clustering and then predicting. This paper made classification using user's information.

D. Banerjee [3] used the monthly stock data for the prediction using ARIMA. As ARIMA is a notable model

 DOI : 10.35291/2454-9150.2020.0202

for the time series performance the results obtained were too satisfactory.

Tiwari, S Bharadwaj, A. and Gupta [4] used the Holt Winters neural network as well as ARIMA method for the prediction of opening price of Nifty50 along with the sentiment analysis.

A. Bhattacharjee, J. Kharade [5] used the month wise stock price value and used those data on the algorithm like Logistic Regression which did not predict expected results but the Cluster-Then-Predict algorithm provided better results.

Debaditya Raychaudhuri [6] have taken previous 10 years Sensex closing data and generated a predictive model using ARIMA time series forecasting method and concluded that over the time there is an upward trend and can be used to make prediction of closing data.

Sreeraksha M S, Bhargavi M S [7] used three distinct machine learning models such as ARIMA, Support Vector Regression, LSTM Neural Network and analyzed the accuracy of the closing price by comparing the accuracy result of all the three models.

Gaurav Priyadarshi, Avneet Ranjan, Sharath Kumar, Bipul Mohanta [8] presented the paper which is a comparative study of the time series analysis model (ARIMA) and effective as well as powerful business intelligence tool (TABLEAU) to predict the closing price of Google Inc.

III. METHODOLOGY

3.1) Moving Average (MA)

Moving average is one of the classical statistical time series forecasting methods. To get the idea about the trends moving average method is useful. It is an average of any subdivision of data or numbers. For each consequent step the values predicted are taken for review and the earliest examined values are removed.

Moving average is also called the rolling average or running average. As Moving average is a technical analysis, it helps to filter out the noise.

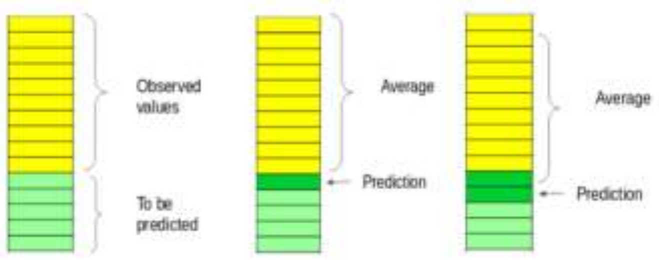

Figure.3.1.1. Moving Average technique

Formula for moving average are:

Simple Moving Average

$$SMA = \frac{A1+A2+\cdots+An}{n} \qquad (1)$$

Where

A= average in period n

n= number of time periods.

Exponential Moving Average

$$EMAt = [Vt * \left(\frac{s}{1+d}\right)] + EMAy * [1 - \frac{s}{1+d}] \qquad (2)$$

Where

$EMAt$ =EMA today

Vt = Value today

$EMAy$ = EMA yesterday

s = smoothing

d= number of days.

For predicting or forecasting long term trends, moving average is very much useful. It can be calculated for any period of time.

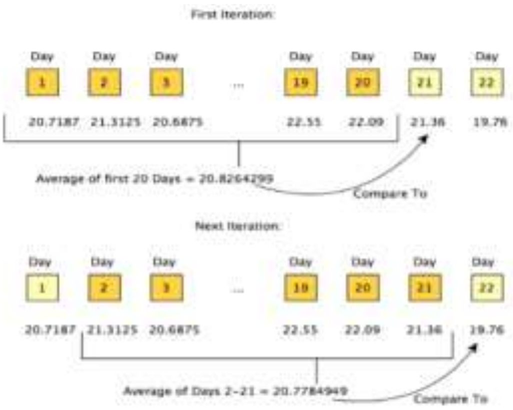

Figure.3.1.2. Moving Average procedure

The flexible nature of moving average allows the investor to analyze the trends at different time intervals and help them in making any decision regarding putting money in the stock market.

3.2) ARIMA (Nonseasonal)

In this paper, the ARIMA model would be trained using the Auto Arima method. For this we will first understand the ARIMA method.

Arima is a popular statistical time series model. It stands for Auto-Regressive Integrated Moving Average.

ARIMA works on a principle like data should be stationary which means the mean and the variance of the data should not vary with time. The series can be made stationary by using the differencing method. Also, the data must be a univariate series. ARIMA has three components- AR (Auto regressive), I (differencing), MA (Moving Average).

AR refers to the values in the past which are used for forecasting purposes. Parameter ‘p’ is for AR. MA refers to the number of forecasted errors in the past to predict the values in future. Parameter ‘q’ is for MA. The most important parameter is differencing ‘d’ which is used to make the data stationary.

Implementation step of ARIMA:

1. Loading of data
2. Preprocessing of data such as making the series univariate.

 DOI : 10.35291/2454-9150.2020.0202

ISSN : 2454-9150 Vol-05, Issue-12, Mar 2020

3. Making the series stationary.
4. Determining the value of d for making the series stationary.
5. To determine the input parameters, ACF and PACF plots are created.
6. From the ACF and PACF plots, the input values of p and q are obtained.
7. Collecting all the processed data as well as the input value, the ARIMA model is fitted.
8. Further process is predicting the future.
9. And lastly, calculation of RMSE is done for checking the performance of the model.

Before implementing the ARIMA model, multiple tasks have to be carried out like making data stationary, determining the value of p and q.

Auto ARIMA makes the task simpler by eliminating the step 3 to 6 from the above explained step.

Implementation step of Auto ARIMA:

1. Loading the data
2. Preprocessing of the data
3. Fitting the Auto ARIMA model
4. Prediction based on validation data
5. To check how the model performed, RMSE is calculated.

Auto ARIMA automatically selects the best combination of (p, q, d) which is best suited for the model and which results in the least error.

Auto Arima selects the leading parameters by taking the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) value into consideration.

AIC helps in selecting the model, where it assesses the quality of each model.

The equation of AIC:

$$AIC = -2ln\,(L)\; + 2k \tag{3}$$

Where

L = likelihood

k= number of parameters.

The model with minimum or low AIC is considered a good model.

BIC is another criterion for the selection of models. The model which has the lowest BIC is accepted.

The equation of BIC:

$$BIC = ln\,(n)k\; - 2\,ln\,\left(\hat{L}\right) \tag{4}$$

Where

$\hat{L}$ = maximized value of likelihood function

n = number of observations

k= number of parameters estimated by model.

3.3) SARIMA:

SARIMA or Seasonal Autoregressive Integrated Moving Average is an extension of the ARIMA model.

SARIMA supports the time series data which are univariate and which contain seasonal components.

Drawback of the ARIMA model is that it does not support the seasonal data which varies from time to time.

For the configuration of SARIMA models, the trend as well as seasonal elements should be considered.

The SARIMA model can be denoted as:

$$SARIMA\;(p, d, q)(P, D, Q)m \tag{5}$$

The trend elements are same as the ARIMA model which are considered as non-seasonal part of model:
p = autoregressive order
d = differencing order
q = moving average order

The newer elements that should be configured are the four seasonal elements:
P = Seasonal Autoregressive order value
D = Seasonal Differencing order value
Q = Seasonal Moving Average order value
m= number of observations per year.

IV. COMPARISON BETWEEN MOVING AVERAGE, ARIMA, SARIMA

The Stockpredict.csv dataset consists of the daily stock prediction data that has been obtained from QUANDL, a source for financial, economic dataset. Its parent organization is Nasdaq. The dataset consists of 8 variables and 1235 observations as shown in figure 4.1 and it is loaded in Python.

```
In [5]: print(df.tail())
            Date    Open    High  ...   Close  Total Trade Quantity  Turnover (Lacs)
1230  2013-10-14  160.85  161.45  ...  159.45             1281419.0          2039.09
1231  2013-10-11  161.15  163.45  ...  160.05             1880046.0          3030.76
1232  2013-10-10  156.00  160.80  ...  160.15             3124853.0          4978.80
1233  2013-10-09  155.70  158.20  ...  155.55             2049580.0          3204.49
1234  2013-10-08  157.00  157.80  ...  155.80             1720413.0          2688.94

[5 rows x 8 columns]

In [6]: print('\n Shape of the data:')
   ...: print(df.shape)

 Shape of the data:
(1235, 8)
```

Figure.4.1. Stock Price Dataset

From the dataset, for the stock prediction the major focus will be on the closing price data as the profit and loss calculation are performed on it and another is the Date which is useful as the closing price varies from time to time. To check the performance of the time series model, the RMSE (Root Mean Square Error) value is taken into consideration.

RMSE is the variance of the residuals (prediction errors). Residuals are the measure how far the regression line are from the data points, whereas RMSE is a measure of how spread out these residuals are. It is a standard way to measure the error.

 DOI : 10.35291/2454-9150.2020.0202

4.1) Evaluation of Moving Average (MA)

```
In [11]: print('\n Shape of training set:')
    ...: print(train.shape)

 Shape of training set:
(987, 2)

In [12]: print('\n Shape of validation set:')
    ...: print(valid.shape)

 Shape of validation set:
(248, 2)

In [13]: preds = []
    ...: for i in range(0,valid.shape[0]):
    ...:     a = train['Close'][len(train)-248+i:].sum() + sum(preds)
    ...:     b = a/248
    ...:     preds.append(b)

In [14]: rms=np.sqrt(np.mean(np.power((np.array(valid['Close'])-preds),2)))
    ...: print('\n RMSE value on validation set:')
    ...: print(rms)

 RMSE value on validation set:
104.51415465984348
```

Figure.4.1.1. RMSE value of Moving Average (MA)

```
In [18]: print(valid[['Close', 'Predictions']])
       Close  Predictions
987    208.3   152.306452
988   208.45   152.310711
989    209.4   152.313376
990      212   152.331373
991   210.25   152.327871
...      ...          ...
1230   230.9   163.132287
1231   227.6   162.960239
1232   218.2   162.784877
1233   209.2   162.614857
1234  215.15   162.426005

[248 rows x 2 columns]
```

Figure.4.1.2. Prediction value using Moving Average (MA)

Using the dataset, a Moving Average Model is created and it is observed that the RMSE value is 104.51 which is close to 105 which is not very promising.

4.2) Evaluation of ARIMA(Nonseasonal)

```
In [22]: model = auto_arima(training, start_p=1, start_q=1,max_p=3, max_q=3,
d=0,trace=True,error_action='ignore',suppress_warnings=True)
Performing stepwise search to minimize aic
Fit ARIMA: (1, 0, 1)x(0, 0, 0, 0) (constant=True); AIC=4830.493, BIC=4850.072,
Time=0.346 seconds
Fit ARIMA: (0, 0, 0)x(0, 0, 0, 0) (constant=True); AIC=8567.690, BIC=8577.479,
Time=0.019 seconds
Fit ARIMA: (1, 0, 0)x(0, 0, 0, 0) (constant=True); AIC=4828.502, BIC=4843.186,
Time=0.188 seconds

Near non-invertible roots for order (3, 0, 0)(0, 0, 0, 0); setting score to inf (at
least one inverse root too close to the border of the unit circle: 0.993)
Fit ARIMA: (3, 0, 2)x(0, 0, 0, 0) (constant=True); AIC=4825.653, BIC=4859.915,
Time=1.955 seconds
Total fit time: 7.362 seconds

In [23]: model.fit(training)
Out[23]:
ARIMA(maxiter=50, method='lbfgs', order=(2, 0, 1), out_of_sample_size=0,
      scoring='mse', scoring_args=None, seasonal_order=(0, 0, 0, 0),
      start_params=None, suppress_warnings=True, trend=None,
      with_intercept=True)
```

Figure.4.2.1. Model training of ARIMA(Nonseasonal)

```
In [26]: rms=np.sqrt(np.mean(np.power((np.array(valid['Close'])-
np.array(forecast['Prediction'])),2)))
    ...: print('\n RMSE value on validation set:')
    ...: print(rms)

 RMSE value on validation set:
88.87244002802066
```

Figure.4.2.2. RMSE value of ARIMA

```
In [29]: print([valid['Close'],forecast['Prediction']])
[Date
2017-10-09    208.30
2017-10-10    208.45
2017-10-11    209.40
2017-10-12    212.00
2017-10-13    210.25
               ...
2018-10-01    230.90
2018-10-03    227.60
2018-10-04    218.20
2018-10-05    209.20
2018-10-08    215.15
Name: Close, Length: 248, dtype: float64, Date

2017-10-09    208.548943
2017-10-10    208.722613
2017-10-11    207.800777
2017-10-12    207.869447
2017-10-13    207.053110
                 ...
2018-10-01    162.648857
2018-10-03    162.580187
2018-10-04    162.512017
2018-10-05    162.444343
2018-10-08    162.377161
Name: Prediction, Length: 248, dtype: float64]
```

Figure.4.2.3. Prediction value using ARIMA(Nonseasonal)

The RMSE value obtained using the ARIMA model is 88.87 which is an improvement but the predictions are still not close to real values.

4.3) Evaluation of SARIMA(Seasonal)

```
In [33]: model = auto_arima(training, start_p=1, start_q=1,max_p=3, max_q=3,
m=12,start_P=0, seasonal=True,d=1, D=1,
trace=True,error_action='ignore',suppress_warnings=True)
Performing stepwise search to minimize aic
Fit ARIMA: (1, 1, 1)x(0, 1, 1, 12) (constant=True); AIC=4820.561, BIC=4844.968,
Time=5.277 seconds
Fit ARIMA: (0, 1, 0)x(0, 1, 0, 12) (constant=True); AIC=5479.245, BIC=5489.008,
Time=0.067 seconds
Near non-invertible roots for order (2, 1, 0)(0, 1, 1, 12); setting score to inf
(at least one inverse root too close to the border of the unit circle: 1.000)
Fit ARIMA: (2, 1, 2)x(0, 1, 1, 12) (constant=True); AIC=4816.660, BIC=4850.830,
Time=8.470 seconds
Near non-invertible roots for order (2, 1, 2)(0, 1, 1, 12); setting score to inf
(at least one inverse root too close to the border of the unit circle: 1.000)
Total fit time: 101.054 seconds

In [34]: model.fit(training)
Out[34]:
ARIMA(maxiter=50, method='lbfgs', order=(0, 1, 0), out_of_sample_size=0,
      scoring='mse', scoring_args=None, seasonal_order=(0, 1, 1, 12),
      start_params=None, suppress_warnings=True, trend=None,
      with_intercept=True)
```

Figure.4.3.1. Model training of SARIMA(Seasonal)

```
In [36]: rms=np.sqrt(np.mean(np.power((np.array(valid['Close'])-
np.array(forecast['Prediction'])),2)))
    ...: print('\n RMSE value on validation set:')
    ...: print(rms)

 RMSE value on validation set:
44.93656370997757
```

Figure.4.3.2. RMSE value of SARIMA

```
In [37]: print([valid['Close'],forecast['Prediction']])
[Date
2017-10-09    208.30
2017-10-10    208.45
2017-10-11    209.40
2017-10-12    212.00
2017-10-13    210.25
               ...
2018-10-01    230.90
2018-10-03    227.60
2018-10-04    218.20
2018-10-05    209.20
2018-10-08    215.15
Name: Close, Length: 248, dtype: float64, Date
```

 DOI : 10.35291/2454-9150.2020.0202

ISSN : 2454-9150 Vol-05, Issue-12, Mar 2020

```
2017-10-09     210.108915
2017-10-10     209.733647
2017-10-11     210.301063
2017-10-12     210.770876
2017-10-13     210.914529
                  ...
2018-10-01     292.554615
2018-10-03     292.809512
2018-10-04     293.471089
2018-10-05     294.087560
2018-10-08     294.393044
Name: Prediction, Length: 248, dtype: float64]
```

Figure.4.3.3. Prediction value using SARIMA(Seasonal)

The RMSE value obtained using SARIMA model is 44.93 which is a modest improvement from ARIMA. Also, the prediction value has some improvement too.

V. FINDINGS

The performance of the Moving Average for the stock prediction is very poor as it gives very high RMSE value. The predicted value and the testing value have a huge difference between them.

If the 2017-10-09 closing price is 208.3, then the predicted value of that same day is around 152.30 which is not even close to real value.

In the ARIMA model where AIC and BIC are used for the selection of the best model, the minimum AIC value obtained is 4825.653 and the best order for fitting the model is (2,0,1) where p=2, d=0, q=1.

Also, if the 2017-10-09 closing price is 208.30 then the predicted value of that day is nearby around 208.54.

But if the closing price of 2018-10-01 is 230.90 then the predicted value is 162.64. Therefore, the performance of the ARIMA model is average.

The SARIMA model takes seasonal value into consideration where the stock data have predictable changes over the time.

The minimum AIC obtained is 4816.660.

The best order for the non-seasonal elements is (0,1,0) where p=0, d=1, q=0. The order for seasonal elements is (0,1,1,12) where P=0, D=1, Q=1 and m=12.

The RMSE value obtained is pretty good than the Moving Average and ARIMA. If a closing price of 2017-10-09 is 208.30 then the predicted value is 210.10, which is a very good prediction.

The closing price of 2018-10-01 is 230.90 then the predicted value is 292.55 which shows there is an upward trend in the closing data.

VI. CONCLUSION

In this paper, three different time series forecasting methods had helped understanding the trends in the stock market. The model is analyzed and fitted for training. After applying the various time series models the Moving Average model has a very poor performance. So, moving average cannot be applied for stock prediction. The next model, ARIMA(Nonseasonal) has better performance than the moving average but the drawback of this model is, it does not consider the seasonal value. The last model, SARIMA, takes seasonal value into the model and has way better performance than MA and ARIMA. The main motive behind using this model is to make the investor feel better that he can have valuable returns.

REFERENCES

[1] S. Gour, "*Developing Decision Model by Mining Historical Prices Data of Infosys for Stock Market Prediction*", International Journal of Computer Sciences and Engineering, Vol.**4**, Issue.**10**, pp. **92-97, 2016**.

[2] Y. X. Lu, T. Zhao, "*Research on time series data prediction based on clustering algorithm*", In the Proceedings of the 2017 American Institute of Physics Conference, United States, pp. **1864-020152, 2017.**

[3] D. Banerjee, "*Forecasting of Indian stock market using time series ARIMA model*", In the proceedings of 2nd IEEE International Conference on Business and Information Management (ICBIM), pp. **131-135, January 2014.**

[4] Tiwari, S., Bharadwaj, A. and Gupta, S., "*Stock price prediction using data analytics". In* International Conference on Advances in Computing, Communication and Control (ICAC3) (pp.**1-5**) IEEE,**2017**

[5] A. Bhattacharjee, J. Kharade, "*Cluster-Then-Predict and Predictive Algorithms (Logistic Regression)*", International Journal of Computer Sciences and Engineering, Vol.**6**, Issue.**2**, pp.**249-252, 2018.**

[6] Debaditya Raychaudhuri, "*BSE Sensex Closing Index Data Analysis and Forecasting using the ARIMA Model*", International Journal of Computer Sciences and Engineering, Vol.**7**, Issue.**6**, pp.**379-389, 2019.**

[7] Sreeraksha M S, Bhargavi M S, "*A Comparative Study of Machine Learning Models for Stock Market Rate Prediction*", International Journal of Computer Sciences and Engineering, Vol.**7**, Issue.**6**, pp.**985-990, 2019**.

[8] Gaurav Priyadarshi, Avneet Ranjan, Sharath Kumar, Bipul Mohanta, "*Stock Price Prediction Using Time Series Analysis and Business Intelligence*", International Journal of Computer Sciences and Engineering, Vol.**07**, Special Issue.**14**, pp.**28-31, 2019.**

 DOI : 10.35291/2454-9150.2020.0202