Stock Market Forecasting Using ARIMA Model

- Tanmoy Das, Department of Data Science and Statistics, Christ University, Bangalore

Abstract

There are many analysis techniques from the perspective of time series analysis, which are nowadays implemented on a large scale in financial analysis. Various bonds and shares are traded in the world regularly. So, we get datasets from them by putting the values against time. Thus, it turns into a time series data. Some statistical approaches like the naïve method, simple exponential smoothing, ARMA, ARIMA, SARIMA models, Garch models, Neural Networking, simple moving average method, prophet forecasting, etc., can be used in predicting such time series data. Among them, one commonly used technique is ARIMA Model, as it is comprehended easily, and several characteristics are easily extracted from the dataset. Time series has a lot of predictive algorithms. In this paper, we will implement a simple ARIMA model for MasterCard Stock Data from 2006 to 2021.

Keywords

ARIMA Model, Time Series, Forecasting, Prediction

Introduction

Understanding what happens before it really happens has been a continuous concern of people around the world. And with the help of modern tools and technologies, it has become challenging to question traditional myths and find ways to predict the future. Predicting the future based on the past has been the core behind a forecasting algorithm.

The key factors for determining the state of the economy worldwide are named employment rate, overseas trade, and the commonly known stock market. The largest global network of the market is what simply defines the stock market. The amount of money that is evoked from trading in stock markets is a lot more than the cost of all goods and services provided to the entire world economy. While every different investment method comes with its own risk, investing in the stock market comes with higher risks as well as greater returns on the other hand. The concept is quite complicated and gets easily affected by external factors. But, as the market is rising rapidly in the world, several companies are nowadays more & more interested in analyzing the data generated by lakhs and lakhs of bond shares and stock exchanges that are getting generated. Therefore, several predicting models, statistical tools, and analyses are coming into play for forecasting to give several advantages, such as lower risk or building the portfolio to the stock market investors. In the world market, people who are involved in selling or buying stocks are subjected to share also their securities along with it that is the ETF (External Traded Funds) are also traded. As the market index exposes us to a glimpse of the performance of the world stock market, the idea of analysis and forecasting models is a good catch for investor companies.

MasterCard is one such company that is well known across the world for its leading global payments & technology that connects people and businesses. It is the second largest payment processing corporation worldwide which has a revenue of 18,884 US$ M and a share price of 354.83 US$ as of 2021.

In the different scenarios of time series prediction, ARIMA Model is well-fitted or generally used when there is evidence of non-stationarity in the data. In this paper, we utilize the use of ARIMA for the

prediction of stocks of MasterCard and evaluate the results and validate our results.

Literature Review

There are several methodologies for time series prediction available around the internet. Methods of estimating stock prices or predicting them have always been a gorgeous topic in the field of data scientists as well as business analysts and managers.

A neural network trained by second-order differencing has been used to predict the stock prices of the Taiwan stock exchange weighted stock index along with a non-linear ARIMA (1,2,1) model by Jung-Hua Wang and Jia-Yann Leu (2002).

Babu As and Reddy SK (2015) have used fuzzy neuron and neural network methods to investigate the behavior of daily exchange rates of the Indian Rupees (INR) with other currencies.

Brijesh Dhyani, Manish Kumar, Poonam Verma & Abhisheik Jain (2020) have used non-seasonal ARIMA to forecast the Nifty 50 values of India.

Objective

The primary objective is to predict the movement of the MasterCard stock market, which will help to follow the ongoing trend as of the current scenario.

Methodologies and Analysis

Here our data consists of 3872 rows which show the details of stock prices (volumes), opening price, closing price, highest on that day, and lowest on that day. We have tried here to take a plausible ARIMA model into consideration to predict the future stocks of MasterCard.

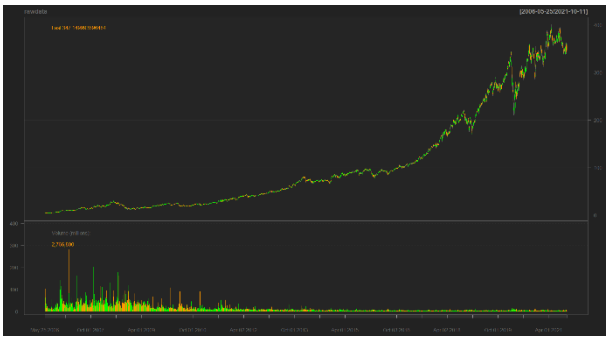

Fig. 1

Decomposition

Before going to any further analysis, it is always recommended to see the four major components of a time series because it tells a lot about the nature of the data at a glance.

Seasonal

Here we can see that the dataset is showing some repeated trends which might be caused by seasons.

Trend

As the name suggests, the trend in data shows us basically what the points are doing at a certain point in time. It can be of many types, upward trend, downward trend, or stationary.

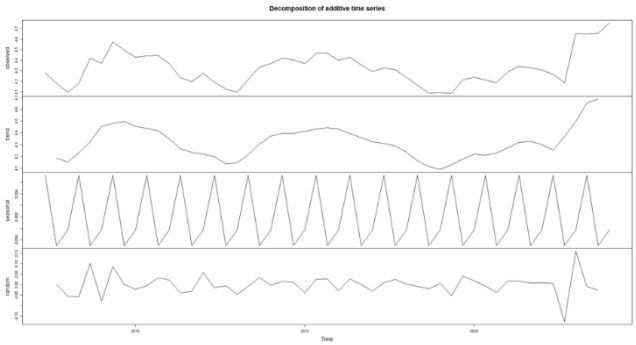

Fig. 2

Cyclical

If data is cyclical, it means that it has a trend with no particular repetition.

Random

Other than the above, there might be some random effects affecting the data, which do not follow any particular pattern or trend, which cannot be captured by the tools and are denoted as random components.

ADF

In ARIMA, stationarity is an important factor. Stationarity means constant means and variance in the data, which will eventually help in future prediction as the future values will also be with the same characteristics.

To check for stationarity in the data, the Augmented Dicky-Fuller test is preferred to be conducted. While performing with our data, it is found that the is a lag order. Hence, we can say that the data is now ready for ARIMA modelling.

Autocorrelation and the ACF, PACF plots

As the term "ARIMA' suggests, it consists of two major terms in it. AR represents Autoregressive, and MA represents the moving average method. To ensure of usage of both or one of them, autocorrelation graphs are plotted. This plot gives us the correlation of the concerned variables with themselves and also helps to find the required parameters for our ARIMA model.

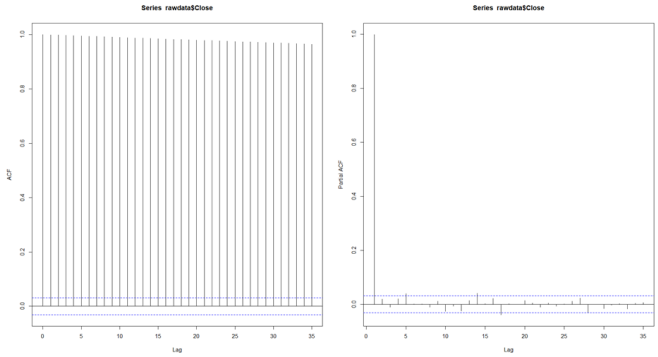

Fig. 3

ARIMA Model

ARIMA or Autoregressive Integrated Moving Average goes with the definition as Box and Jenkins claimed, any non-stationary data can be turned into stationary data by differencing the series of $Y_{t.}$
Which goes as,

$$Y_t = c + \varphi_1 y_{d_{t-1}} + \cdots + \varphi_p y_{d_{t-p}} + \cdots + \theta_1 e_{t-1} + \cdots + \theta_1 e_{t-1} + e_t$$

Or,

$$\left(1 - \sum_{i=1}^{p} \varphi_i B^i\right)(1-B)^d Y_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) e_t$$

Here, Y_t is the differenced time-series values, φ and θ are the unknown parameters and e is the error component. Y_t is expressed in terms of its past values and the current and past values of the error terms.

The ARIMA consists of three components as –

- Auto Regression (AR): The value we get by autocorrelation of the data to itself. The parameter "p" here shows the value.
- Differencing (I for integrated): The differencing of the time series to make it stationary. "d" is the parameter.
- Moving Average (MA): MA values are shown by the parameter "q" values which is the number of the lagged values of the error terms.

These three components help to build the model ARIMA (p, d, q).

We follow this method for our data and go for the first modelling and the forecasting.

Applying ARIMA $(1,1,1)$ we get the results,

$lambda$: 0.4480343

$Coefficients$:

	$ar1$	$ma1$	$drift$
	0.6409	-0.7049	0.0065
s.e.	0.1605	0.1490	0.0014

Table 1

Sigma^2	Log likelihood	AIC	AICc	BIC
0.009636	2794.72	-5581.44	-5581.43	-5557.29

Table 2

$Training\ set\ error\ measures$

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.01288764	0.9800469	0.6135914	-0.0372339494	1.399021	0.9901109	-0.01954681

Table 3

A quick residual check says a lot about the modelling before really going for the prediction.

Residual check

A residual is left of the data after fitting the model,

$$e_t = y_t - \hat{y}_t$$

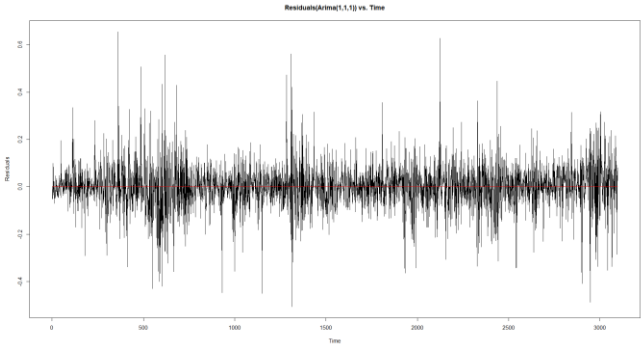

Fig 4.1

If we go for a normal curve

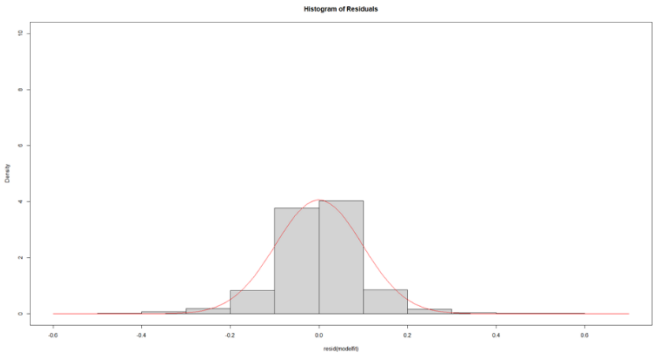

Fig 4.2

The plots of the residuals give us a decent normal curve, which motivates us for further progress with the analysis. The final residual plot is given us,

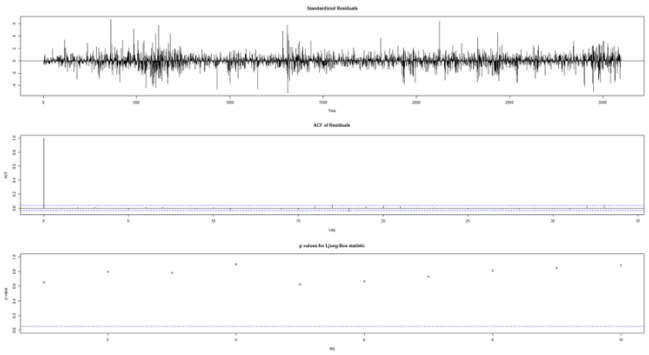

Fig 4.3

Ljung-Box Test

H_0: The dataset points are independently distributed.

A p-value greater than 0.05 does not reject the fact that data points are uncorrelated.

From the above graphs it is seen that no p-value go less than 0.05.

So, further analysis is not necessary and we are ready for forecasting.

Forecasting

First, we plot the model

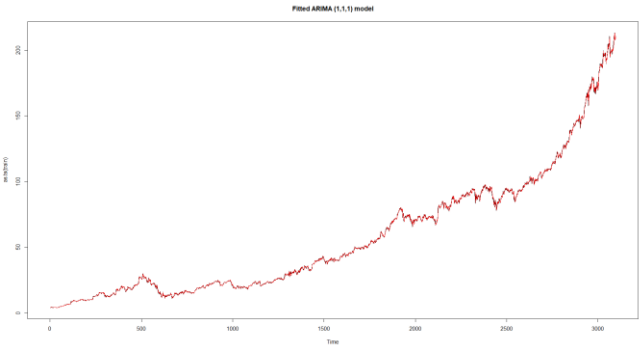

Fig 5

Here the red line over the black line shows the fitted model over the actual datasets respectively.

The below graph shows the forecast for a certain time period in future.

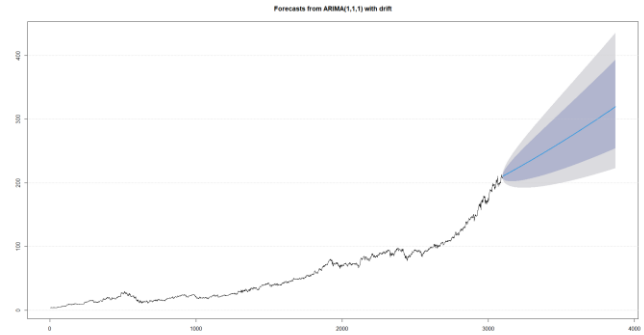

Fig 6

Here the forecasted or the predicted values are shown by the blue line and the bandwidths show the 80% and 95% confidence intervals respectively.

Results and Discussions

The prediction for MasterCard stock data for future is given below as a table:

Date	Predicted Values	Actual Values
20-09-2021	317.4764774	339.2613525
21-09-2021	317.6336947	336.9542542
22-09-2021	317.7909549	342.9766541
23-09-2021	317.9482581	353.6231995

24-09-2021	318.1056043	357.7080383
27-09-2021	318.2629934	356.3197937
28-09-2021	318.4204255	350.7368469
29-09-2021	318.5779006	352.7443237
30-09-2021	318.7354187	347.241272
01-10-2021	318.8929798	359.7254944
04-10-2021	319.0505838	343.7656555
05-10-2021	319.2082308	342.776886
06-10-2021	319.3659208	348.25
07-10-2021	319.5236538	353.9100037
08-10-2021	319.6814298	354.9599915

Table 4

As the predicted values are closer to the actual values, we say that our ARIMA model is a good model for prediction.

Conclusion

As per the objective of our study, we have successfully found a model to predict the future stock values of MasterCard. We followed the proper procedures, checked for stationarity in the data, decomposed it, fitted the model after determining appropriate parameters, forecasted the values, and compared them with the actual ones too. Another study with applied statistics has been proven to be of good use for the fields of science, business analytical fields, and the entire world economy.

Future Scope

As our model is a good fit for the data and for prediction, it is not an optimum model. Various other methods are available there in the world for the prediction of time series data and for the stock market. One can find other methods of forecasting and fit different models according to them. Then our results can come to the field of comparison where one can find the optimum model for forecasting the given set of data.

Generally, investment companies also can use this prediction to implement the results to their applications for the users to watch and get a grip on the does and do nots of the MasterCard stock.

One can carry the forecasting on, so maybe for the entire year 2022 and then 2023, as the recent time data is already available in the market. This can be done by a student of statistical background or of any other background.

References

[1] Dhyani. B, Kumar. M, Verma. P, Jain. A, (2020, March), "Stock Market Forecasting Technique using Arima Model", International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6, ISSN: 2277-3878, accessed 15 January, https://pdfs.semanticscholar.org/3530/ff5b8ba64747ceefeb094366121892c4cb42.pdf

[2] Babu AS, Reddy SK (2015) Exchange Rate Forecasting using ARIMA, Neural Network and Fuzzy Neuron. J Stock Forex Trad 4:155. doi:10.4172/2168-9458.1000155, accessed 15 January, https://kettletrading.com/FOREX.html

[3] Jung-Hua Wang and Jia-Yann Leu, "Stock market trend prediction using ARIMA-based neural networks," Proceedings of International Conference on Neural Networks (ICNN'96), *Washington, DC, USA, 1996, pp. 2160-2165 vol.4, doi: 10.1109/ICNN.1996.549236, accessed 15 January,* https://ieeexplore.ieee.org/abstract/document/549236/citations#citations

[4] Isenahd, Godknows M., Olubusoye, Olusanya E., Forecasting Nigerian stock market returns using ARIMA and artificial neural network models, CBN Journal of Applied Statistics [ISSN:] 2476-8472 [Volume:] 05 [Year:] 2014 [Issue:] 2 [Pages:] 25-48, accessed 15 January, https://www.econstor.eu/handle/10419/144781

[5] Bose. S, Saha. S, Forecasting of a time series (Stock Market) in R, accessed 13 January, https://www.student.unsw.edu.au/how-do-i-cite-electronic-sources

[6] Time Series Forecasting Using R, accessed 13 January, https://www.pluralsight.com/guides/time-series-forecasting-using-r

[7] Using R for Time Series Analysis, accessed 13 January, https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html

[8] Writing a literature review, accessed 15 January, https://psychology.ucsd.edu/undergraduate-

program/undergraduate-resources/academic-writing-resources/writing-research-papers/writing-lit-review.html#4.-Organize-your-notes-and-thou

[9] The Stock Market, accessed 15 January, https://www.bartleby.com/essay/The-Stock-Market-F3P6HVWKFTDX

[10] Mastercard, accessed 15 January, https://en.wikipedia.org/wiki/Mastercard

[11] Mastercard for you, accessed 15 January, https://www.mastercard.us/en-us.html

Appendix