TIME SERIES ANALYSIS ON STOCKS

DATA 621 FINAL PROJECT

Critical Thinking Group 4

Rajwant Mishra,
Priya Shaji,
Debabrata Kabiraj,
Isabel Ramesar,
Sin Ying Wong and
Fan Xu

ABSTRACT

- In this project, 5 years of live historical dataset (2015 to 2020) of stocks has been explored and analyzed with time series models.
- Time series models used are AR(Auto Regression) and MA(Moving Average) .
- Time series forecasting process has been executed using ACF() function and ACF plots.
- Lastly, we have evaluated all data models comparing its prediction score to analyze which model has performed better.

I. KEY WORDS

- Lag : A “lag” is a fixed amount of passing time; One set of observations in a time series is plotted (lagged) against a second, later set of data. The kth lag is the time period that happened “k” time points before time i. The most commonly used lag is 1, called a first-order lag plot.
- Seasonality : In **time series** data, **seasonality** is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly, or quarterly.
- Stationary : Stationary graphs are relevant to time series analysis, where we seek to understand the changes of a graph over time. With time series analysis, it is expected for data to vary over time, however, it is difficult to figure out the exact pattern by which a graph will change over time.
- Random Walk : A random walk, on the other hand, does not have this same tendency to centralize towards the mean due to the individual points along the walk being dependent on the previous points. This adds variance the more points are included in the walk, which can cause the path of the walk to deviate very far away from the mean.
- White Noise : With a white noise graph, we know that the distribution of the points will be normal and centered around zero with the same variance because the points are independent, so the tendency over time will be towards the mean
- AR (Auto regressive) :
- MA (Moving Average) :

II. KEY WORDS

- Autoregressive integrated moving average : in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model
- AR (Auto regressive) : In this **regression model,** the response variable in the previous time period has become the predictor and the errors have our usual assumptions about errors in a simple linear regression model. The **order** of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time. So, the preceding model is a first-order autoregression, written as AR(1).
- MA (Moving Average) **:Moving averages** are a simple and common type of smoothing used in **time series** analysis and **time series** forecasting. Calculating a **moving average** involves creating a new **series** where the values are comprised of the **average** of raw observations in the original **time series.**
- In time series analysis, the moving-average model (MA model), also known as moving-average process, is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

INTRODUCTION

- We collected data from NYSE from Lat Five year using the API in R
- We also scrapped data of Sectors and Stock so that we can understand the trend by Sector from our data.
- We Analyzed 5 year data by Sector and choose one of the stocks from Healthcare sector.
- We partitioned our data in data before 2020 and after 2020.
- We Build AR and MA model on data before 2020 and predicted stock value for Year 2020
- We were able to check accuracy of the model by Model comparison and graph .

I. LITERATURE REVIEW

- One of the interesting works in stocks analysis is “using data mining with time series data in short-term stocks prediction” which explores methodologies similar to our project.
- Their approach uses data mining with time series data using examples related with short-term stocks prediction which is proved to be important to a better understanding of the field
- Specific challenges: developers focus on the issue of representing time series data in order to effectively and efficiently apply data mining.

II. LITERATURE REVIEW

- Another interesting issue was to find out if different time series or parts of time series have similar behavior.
- This issue can be approached through the use of similarity measures or indexing techniques.
- Over-fitting is a common problem across data mining applications.
- Achievements: A new concept, named as “median strings” is presented as a simple and at the same time powerful representation of time series data.

III. LITERATURE REVIEW

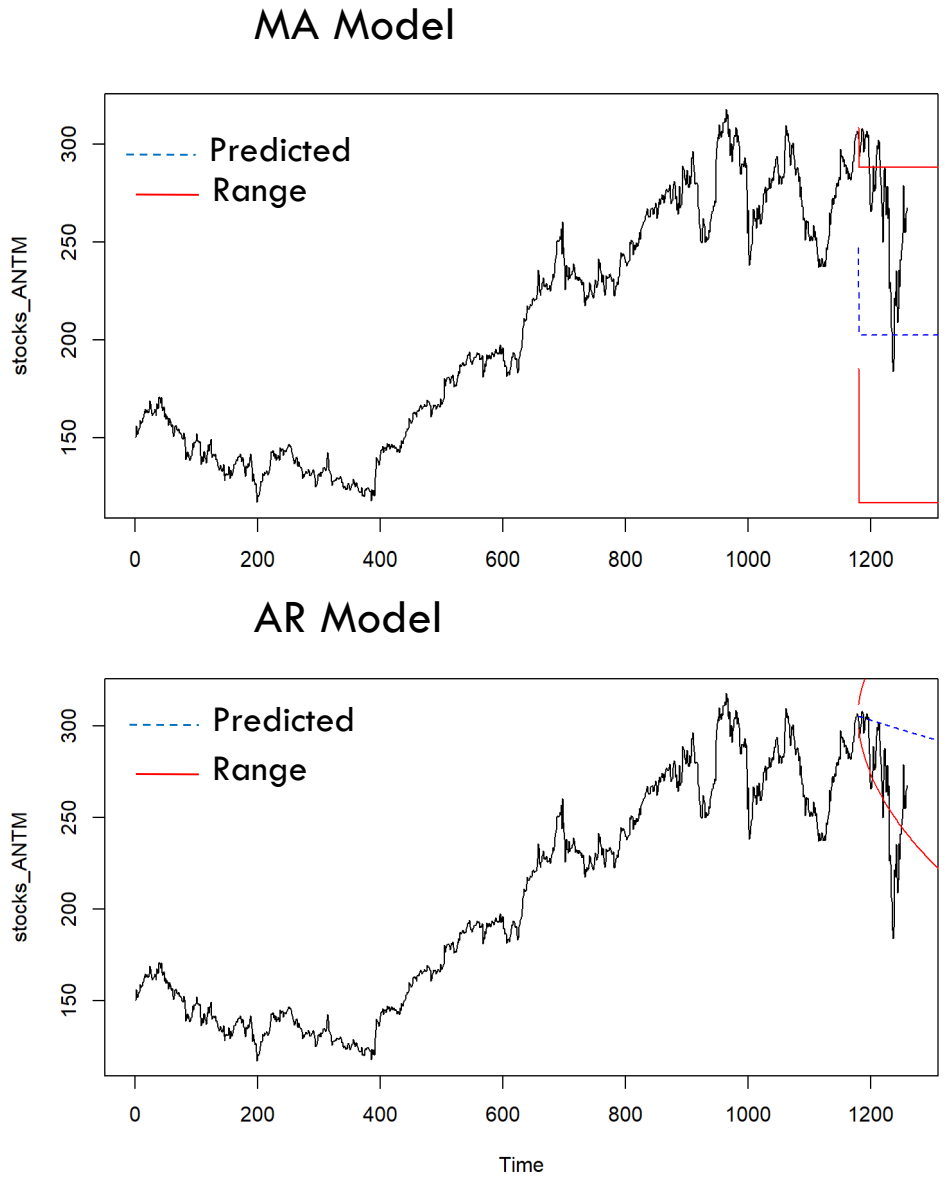
- Our work/investigation on our project is different from their's by our approach used to solve a similar issue
- Time series models used by us are AR and MA model, as both these models perform with better predictions specifically when market is not stable, which hold true for current covid-19 scenario.
- Link to book we used to learn how other researchers have solved similar issue: Data mining with time series data

METHODOLOGY

- We have used Auto Regressive Integrated Moving Average Model with AR and MA model.
- Together with the autoregressive (AR) model, the moving-average model is a special case and key component of the more general ARMA and ARIMA models of time series, which have a more complicated stochastic structure.
- A time series is a sequence of measurements of the same variable(s) made over time. Usually the measurements are made at evenly spaced times - for example, monthly or yearly. Let us first consider the problem in which we have a y-variable measured as a time series. As an example, we might have y a measure of global temperature, with measurements observed each year. To emphasize that we have measured values over time, we use "t" as a subscript rather than the usual "i," i.e., yt means y measured in time period t.
- An autoregressive model is when a value from a time series is regressed on previous values from that same time series. for example, yt on y_{t-1}: $y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$.
- In this regression model, the response variable in the previous time period has become the predictor and the errors have our usual assumptions about errors in a simple linear regression model. The order of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time. So, the preceding model is a first-order autoregression, written as AR(1).

EXPERIMENT RESULT

- As we can very clearly see the from the figure in right :
- MA Model has predicted best in case of Year 2020 from the model which was fitted on data until Year Dec-2019.
- AR model prediction is way off from actual value, even though its close to negative range of the prediction.



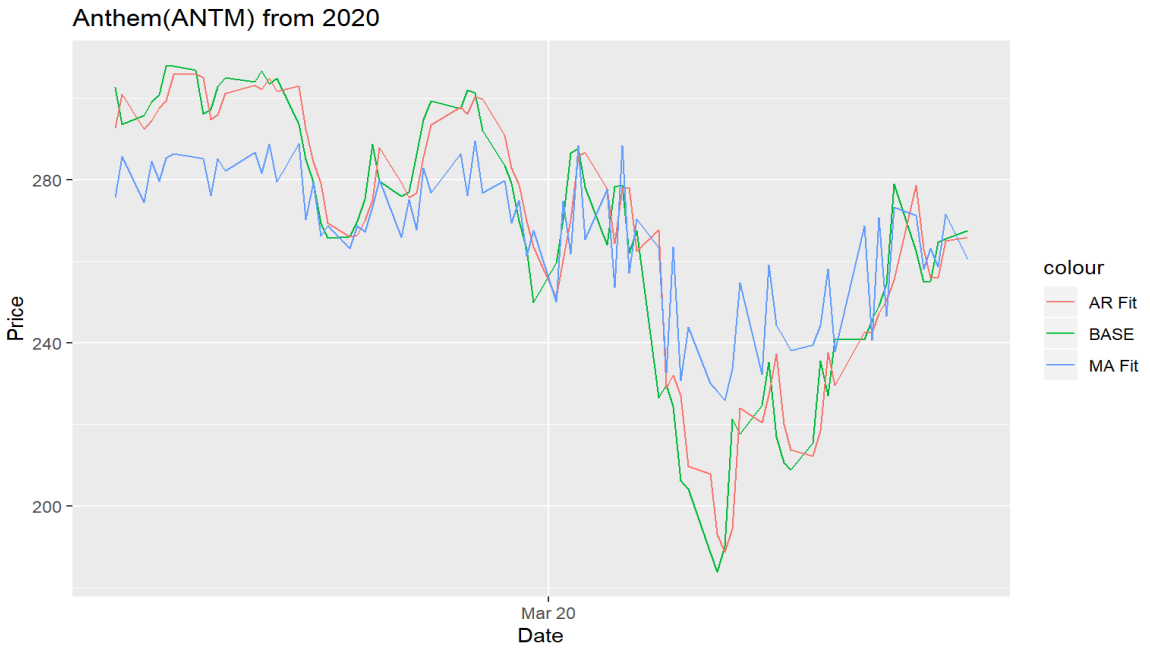
I. DISCUSSION AND CONCLUSION

If we compare the data and results from AR and MA Model , we noted that:

MA model seems to be doing better with predication when we used data till Dec 2019.

For instance from the figure in right we can see very clearly a sharp drop in price , due to COVID-19 Pandemic. Letss see how these two Model predicted if we go by full data:

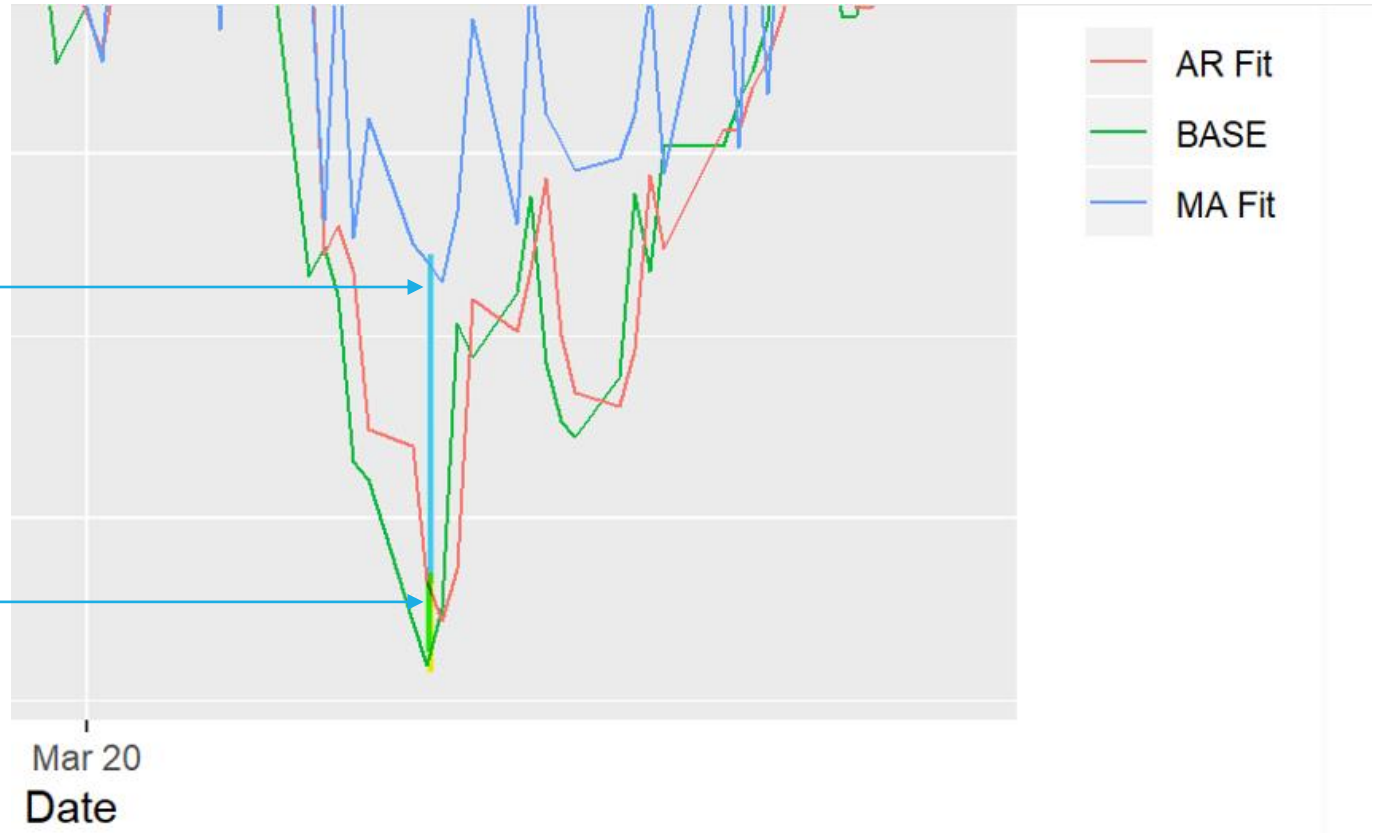
Cont. Next slide….



II. DISCUSSION AND CONCLUSION

Diff of MA model and AR Model

Diff of AR Model Actual Data



When we use full data set to predict the value for year 2020.

AR Model does good with data set with random variation and model is able to predict better in comparison with MA model.

III. DISCUSSION AND CONCLUSION

Based upon our underacting of this time series analysis we noted that :

Different model can be used to better predict same set of time series

AR model is would always perform better for few prediction if market is not stable

MA model may give better predication when market is very unstable

Training and testing in Time series data depends on portioning data by date , Random selection of such data may not be accurate choice to better check the efficiency of the model.

REFERENCE

Data Camp R cheat-sheet

Introduction to Stock Analysis

R for Data Science cheat-sheet

A little book of R for Time Series

Applied Time Series Analysis for Fisheries and Environmental Sciences

Autoregressive Models

Moving-average model

APPENDICES

Project GitHub link

https://github.com/Rajwantmishra/DATA621_CR4/tree/master/Final

THANK YOU