Time Series Analysis and Forecasting Methods of

Tesla and Volkswagen Stock Price

Eswarya Pillai Rahul Sudan

Harrisburg University of Science and Technology

Introduction

The main aim of this project is to forecast the future stock price of Tesla and Volkswagen motor companies.

Tesla Inc is an American company specializes in electric vehicle manufacturing, Solar panel and solar roof tile manufacturing.

Volkswagen is a German based automotive manufacturing company specialize in manufacturing passenger, commercial vehicles and motorcycles. I chose this motor companies to forecast electric based company stock value and automotive based company stock value. The stock market is a place of buying and selling of company stocks. Every stock exchange has its own stock index value. The Index is the average value that is calculated by combining several stocks. The stock market can have a huge impact on people and the country’s economy as a whole.Therefore predicting the stock value is an efficient manner can help to minimize the risk of loss and maximize profit.

Prediction is difficult, by using time series analysis we can assume the value but we can’t predict the accurate value.

Time Series forecasting falls under the category of quantitative forecasting wherein statistical principals and concepts are applied to a given historical data of a variable to forecast the future values of the same variable.

As we know the stock prices are not randomly generated values instead they can be treated as a discrete-time series model which is based on a set of well defined numerical data items collected at successive points at regular intervals of time. Instead of forecasting directly, it is better to identify a model to analyze trends and use various different models like ARIMA, SARIMA,VAR and different other models which gives more authentic and reliable results.

For this project I have used the historical monthly data of the stock price of Tesla and VW of the past 7 years to forecast the next 1 year stock trend by using different Time Series Model.

I have downloaded the data from yahoo finance website.

**MOTIVATION**

In this Research paper, we discuss the different time series models which have been applied for stock prices of Tesla and VW to predict the rise and fall of stock prices before the actual event of an increase or decrease in the stock price occurs. In particular the paper discusses the application of Linear Regression, ARIMA and VAR. The paper introduces the parameters and variables that can be used in order to recognize the patterns in stock prices which can be helpful in the future prediction of stocks.

The main goal of the project was to study and apply as many different Time series model as possible on a dataset in predicting the price of a stock.

HYPOTHESIS

Predicting the stock returns is a difficult task, also predicting stock returns give crucial implications about market efficiency. The hypothesis test of my research paper is to predict the stock price for the Tesla and Volkswagen motor company and compare their predicted values and analyze which company will perform better, which is likely to increase.

DATA AND METHODOLOGY

Monthly stock prices of Tesla and Volkswagen from July 2013 to July 2020 are extracted from the Yahoo finance website. This data set contains the open, high, low, close and adjusted close prices of Tesla and VW stock . It also contains trading volume values on these days. To achieve consistency, the close prices are used as a general measure of stock price over the past seven year.

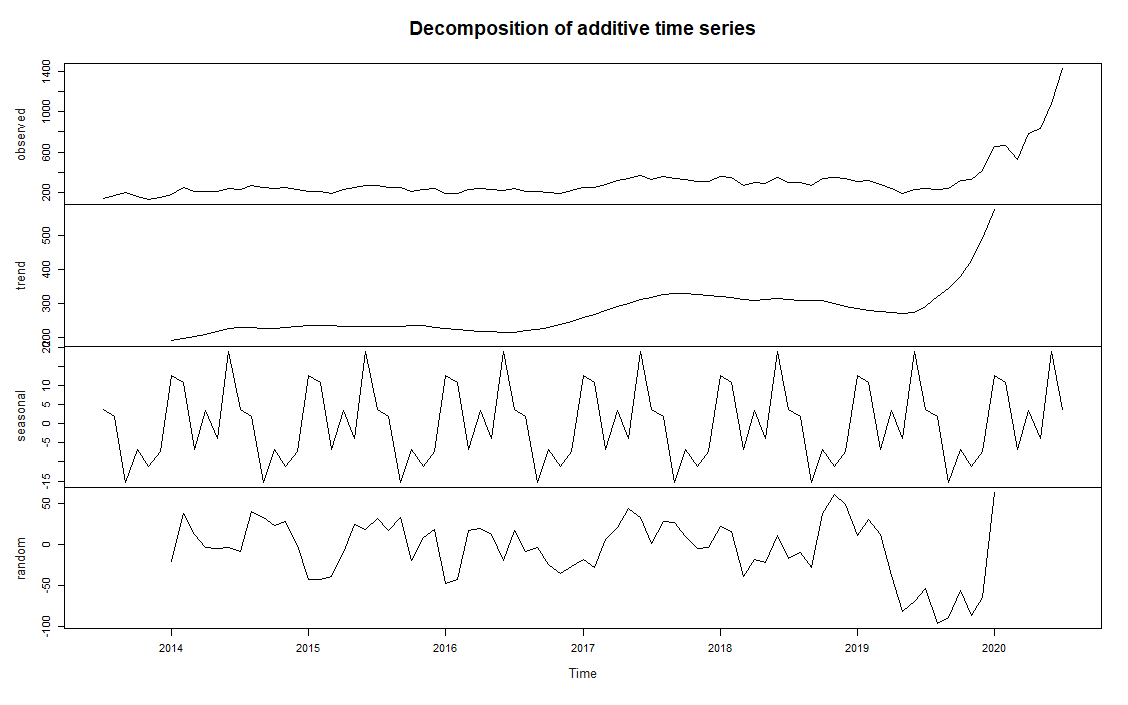
The methodology we opted for in this study is Linear Regression Model, VAR(Vector auto regression model) and ARIMA ( Auto Regressive Integrated Moving Average) models. **The goal of Time Series analysis is** to identify the underlying forces that lead to a particular trend in time series pattern and predict future values of the time series variable. Time series analysis assumes that time-series data consists of some systematic pattern and some random noise

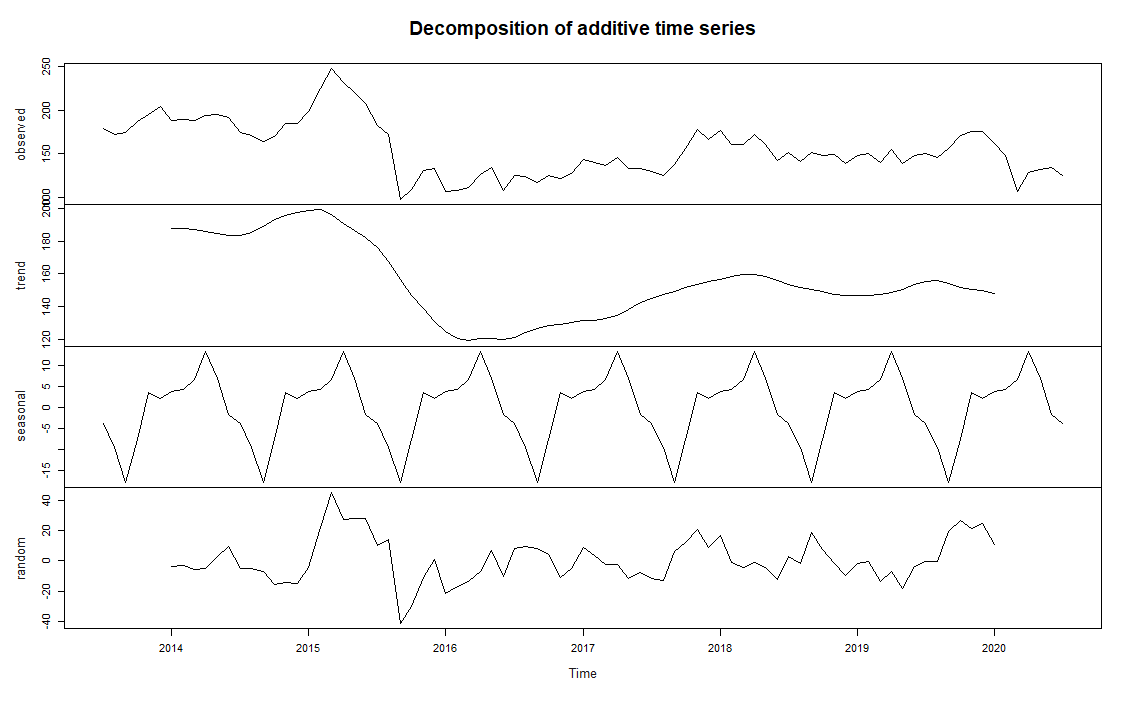
This can be done using Time Series Decomposition. The **decomposition of time series** is a statistical task that deconstructs a time series into several components. Each component represents one of the underlying categories of patterns.

**Types of time series patterns:**

**Trend(T)**- reflects the long-term progression of the series. A trend exists when there is a persistent increasing or decreasing direction in the data. The trend component does not have to be linear.**Cyclic ( C)**— reflects repeated but non-periodic fluctuations.**Seasonal(S)**-reflects seasonality present in the Time Series data, like demand for flip flops, will be highest during the summer season. Seasonality occurs at a fixed period of time could be weekly, monthly, quarterly, etc.**Random(R)**-reflects random or irregular influences. This is residual after we have removed all other components from time-series data.

Below is the Decomposition time series of Tesla and VW stock price. Our data doesn’t seem to have magnitude so I have used addiditive model.





The Tesla stock seems to have a positive trend over a period of time and VW stock has a negative trend, our data doesn’t seem to have a seasonal trend.

**For a good fit of model, our data should be stationary.**

A stationary series is one whose statistical properties like mean, variance and auto-correlation is constant and does not depend on time. A data set that does not exhibit trend or seasonality and fluctuations in data is entirely external.

**Augmented Dickey-Fuller(ADF)Test**

Augmented Dickey-Fuller(ADF) is the most popular statistical method to find if the series is stationary or not. It is also called as Unit Root Test.Unit Root Test determines how strongly a time series is defined by a trend.

H0 - Null hypothesis for ADF test is that time series can be represented by a Unit root, that is not stationary.

Ha- Alternate hypothesis is that time series is stationary.

The ADF test results shows that our data is non-stationary. To convert our non-stationary data to stationary, We can apply the different technique to make the data stationary. This technique helps generate series with constant location and scale

Differencing: calculates the difference between consecutive terms or points in the data. Differencing is performed to get rid of the varying mean.

Log transformation helps to stabilize the non-constant variance of a series.

**Linear Regression Model.**

To perform a linear regression model we have to check whether or data is cointegration or non-cointegration.

Cointegration tests identify scenarios where two or more non-stationary time series are integrated together in a way that they cannot deviate from equilibrium in the long term. The tests are used to identify the degree of sensitivity of two variables to the same average price over a specified period of time.

##### Phillips--Ouliaris Cointegration Test is perfomed on the stationary data , the value of p is less than 0.05, which states that our data is cointegration and we can perform the linear regression model.

I performed the linear regression model by assuming the Tesla stock as dependent variable and VW stock as independent variable, which is not a good fit model.

**ARIMA Model**

Before performing the model the time series data is split into training data contains the Tesla and VW stock price from July 2013 to July 2019 and test data contains the Tesla and VW stock price from July 2019 to July 2020. The training set which is considered as a known output and the model learns on this data in order to be generalized to other data later on. We have the test dataset (or subset) in order to test our model’s prediction on this subset.

The components of ARIMA models—autoregressive, integrated, and moving average—are aimed at explaining the autocorrelation in a series.

The order of the ARIMA(p,d,q) can be chose by examining ACF and PACF plots. ACF plots display correlation between a series and its lags. In addition to suggesting the order of differencing, ACF plots can help in determining the order of the M A (q) model. Partial autocorrelation plots (PACF), as the name suggests, display correlation between a variable and its lags that is not explained by previous lags. PACF plots are useful when determining the order of the AR(p) model. I have used the auto Arima function to get the best order and there is a correlation in my predicted residuals which is not a good fit model.

**VAR Model**

Vector auto regression model is one type of time series model.

The autocorrelation and partial autocorrelation of the stationary data doesn’t have any correlation which is predicted to be good fit and VAR model is used for forecasting the next 1 year stock value, the residuals of the predicted model also doesn’t have any correlation which is a good fit model.

**RESULT AND DISCUSSION**

We performed a test to check whether our data is stationary or no-stationary, by using Augmented Dicky Fuller test, which suggest our data in non-stationary, as we know non-stationary data is not a good model to forecast or predict the future value.

Differencing and Log transformation was implemented to convert our non-stationary data to stationary.

Our stationary data seems to have a cointegration where we can perform a linear regression method, by assuming the Tesla stock as dependent variable and VW stock as independent variable Linear regression method was implemented.

The below is the result suggested that there is no statistical significant relationship between the two stock prices.

|  |
| --- |
| Call:  lm(formula = tesla.ts ~ vw.ts)  Residuals:  Min 1Q Median 3Q Max  -0.46079 -0.14162 -0.01998 0.13288 0.57418  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 0.00203 0.02177 0.093 0.9259  vw.ts 0.36608 0.20210 1.811 0.0738 .  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.1982 on 81 degrees of freedom  Multiple R-squared: 0.03893, Adjusted R-squared: 0.02706  F-statistic: 3.281 on 1 and 81 DF, p-value: 0.07379 |

The p value is 0.07379 which is greater than the alpha value 0.05, no evidence to reject null hypothesis,and conclude that this model is not a best fit model.

Further moving on with the next model ARIMA.

I have split the data into train and test data, to implement the ARIMA model into train data and compare the predicted value with the test data to check whether our model is a good prediction

The next step is to observe the graph of ACF and PACF to suggest the order for the ARIMA model.

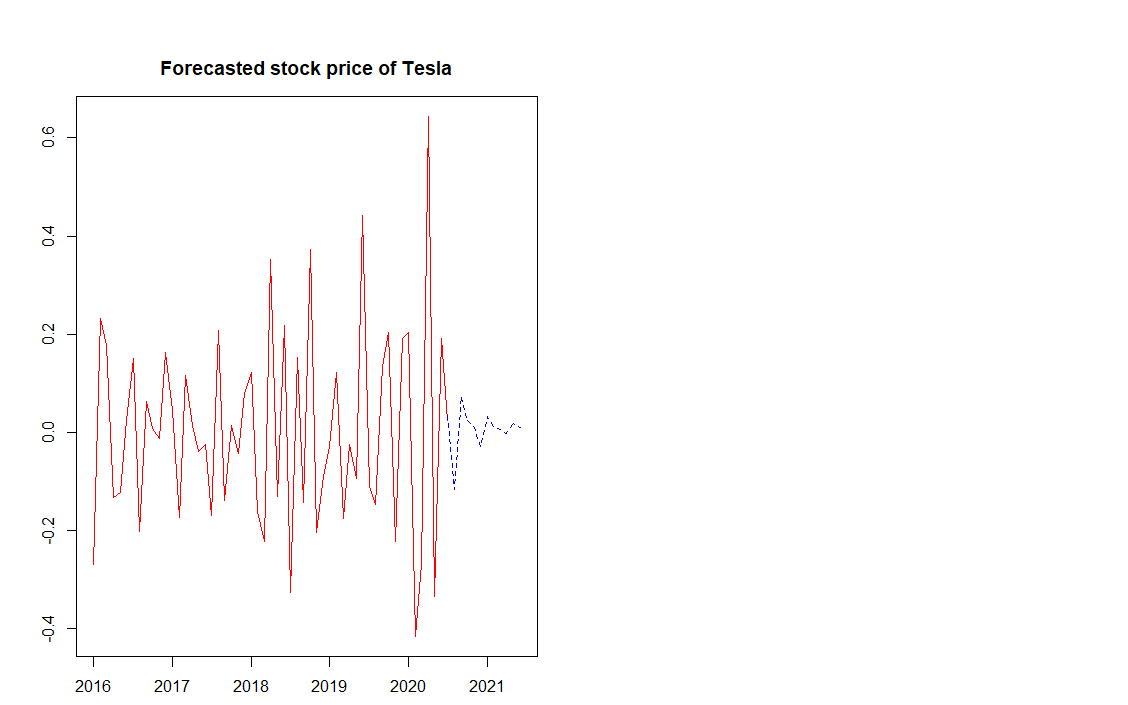
The order suggested (p,d,q) as (1,1,0).

I have used the auto function to get the best order for the model, and predicted the value.

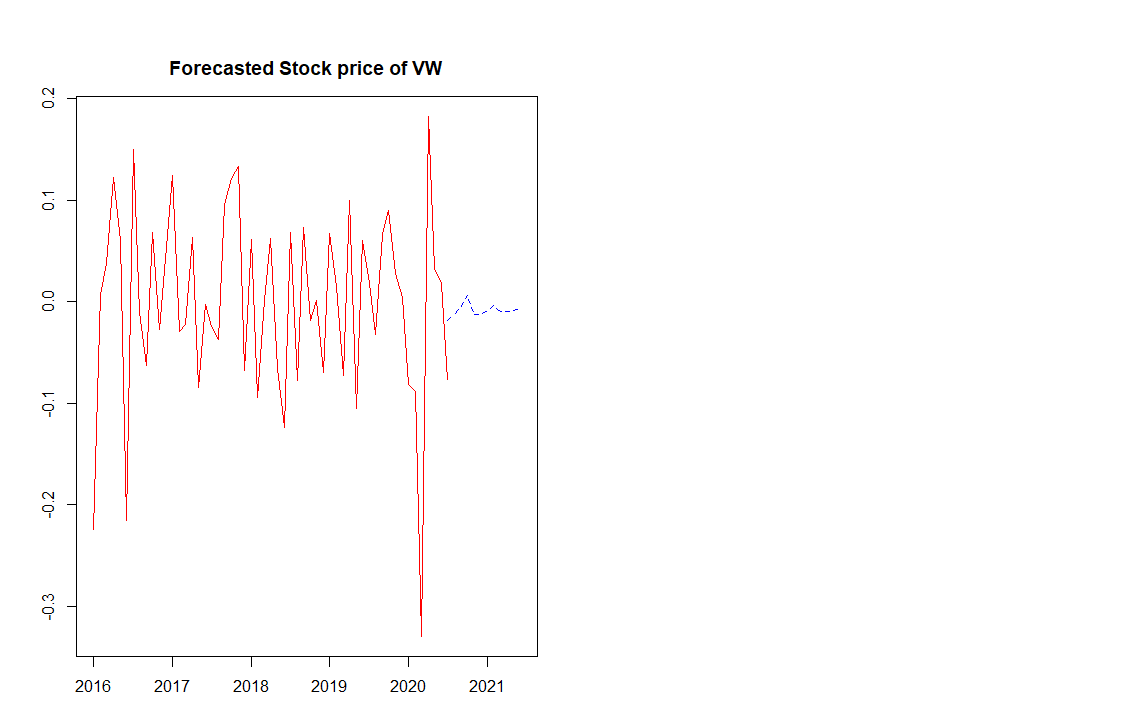
The ACF and PACF of the residuals of the predicted model seems to have lags and has autocorrelation , the mean absolute percentage error is greater than 50% which has a evidence to conclude this is also not a good fit model.

Finally I have performed the VAR(vector auto regression) model for our data.

The residuals of the predicted model doesn’t’ have any autocorrelation and this model is used to predict the value of stock price for the Tesla and Volkswagen for the upcoming 1 year.



From the above graph we can see that for the year 2021 the value of stock price of Tesla is predicted to have up and down trend, the first few month is predicted the price will fall down and then the next few months the price is likely to go upward trend, I suggest that end of 2020 is predicted good time to invest in Tesla stock.



From the above graph we can see that for the year 2021 the value of stock price of Volkswagen is predicted to have slight up and down trend the first few month is predicted the price will increase and then the next few months the price is likely to go downward trend and remain the same, I suggest that the next year is not a good time to invest in Volkswagen stock.

From this analysis we can predict that Tesla stock value seems to go higher than Volkswagen Stock price, and for my data VAR seems to best fit compare to Linear Regression and ARIMA model.

For the future work, we will try to model it using more different approach like SARIMA, TBATS model etc.

REFERENCES:

* Class Assignments and Slides- Pavlo Buryi
* <https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6>
* <https://finance.yahoo.com/quote/NIO/history?period1=1565124664&period2=1596747064&interval=1mo&filter=history&frequency=1mo>
* <https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00727/full#h6>
* <https://blogs.oracle.com/datascience/introduction-to-forecasting-with-arima-in-r>
* <https://github.com/Pyligent/Telsa-Stock-Analysis-R-/blob/master/README.md>

APPENDIX

