# CS510_final

Hesam Sadri

Introduction

This is the final project for the **CS510**.

In this program I am trying to use timeseries analysis to predict future direction of the different stocks. I used interactive format where user could insert it's own favorite stock symbol and see the results.
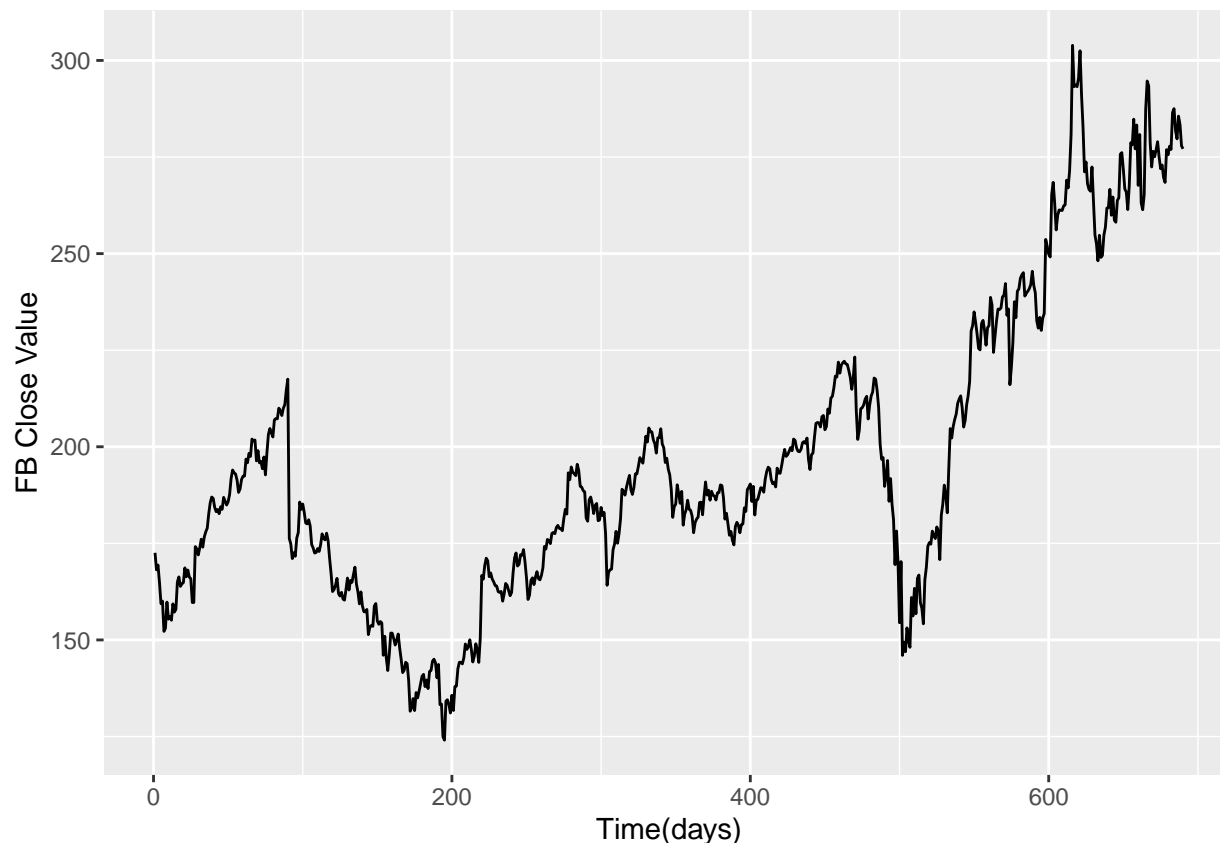
The data for stocks are getting from yahoo finance and there are up-to-dated.For instance when we put WMT we will get the stock of Walmart data for the last 1000 days.

Here we can see different charts related to the Walmart stock that we have downloaded.

Walmart
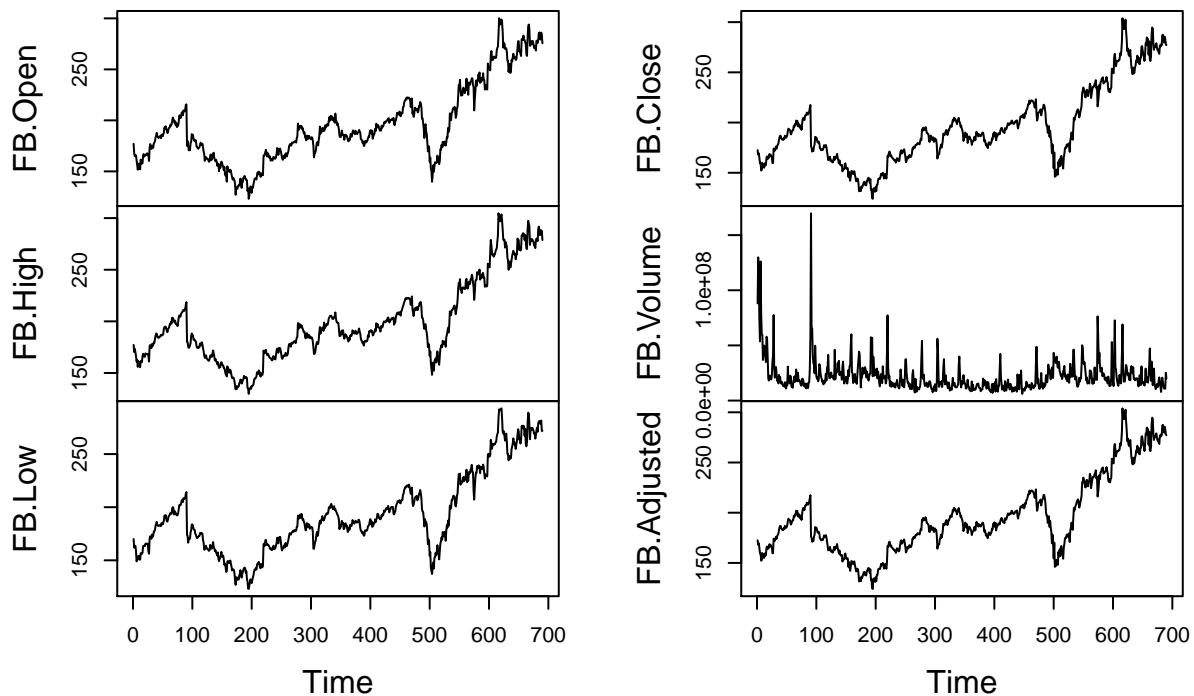
In this program I am going to use the close data which is related to the daily close value of the stock.

After downloading data, I will make a sliding window with a predetermined length(I used 50) and analyze this data to predict the value for the next day and then I compare the result of the prediction with the real value, if both of them are in the same direction I count this as a successful prediction. Then I shift the window one day ahead and do the calculation again. I will make prediction for 100 days and then will calculate the success rate.

I also make two different models, one of them use auto.arima which try to find and fit the data to the best arima model available and the other model use exponential smoothing model(Holt-winter).

Models:

Predicting future value or even direction of a stock market would bring us fortune. There are lots of big financial institutions and expert individuals who spent huge amount of time and money to achieve a reasonable success rate of forcasting, nevetheles there are rare case of consistent profitable investment managers.This is enough to show how complicated is to do prediction of stock market. In this project I am going to use two famous models in the timeseries:

1. **auto.arima**: This model which is in the **tsa package** try to find and fit the available data on the available data and later using **forcast package** I make the prediction.

2. **Holt-Winters**: **Holt-Winter** is one of the methods which use exponential smoothing methods. This methods try to find and seperate the seasonality and trend terms from the

chart and then make a prediction.

Additionally, there is another fact that I want to take under consideration, this fact declare that if we could fit a timeseries data properly to a model, the remainders of real values and values from the moodel should be individually independent from each other. Thus, I also want to see how many of my models have pass the **Box_Pierce** test and how much of them are successful predictors.

This program has two functions:

1. Get_Sym: Which use yahoo.financial website to get the data of the selected symbol. To run this part of the code we require internet connection and the symbol of selected stock. For example "FB","GOOG" are symobls of FaceBook and Google.

2. predict_fn : This function do all the calculation and make mentioned models and predictions and finally print a table of the result.

Results and Analysis:

Here I want to show the results for the Walmart and Google. For the Walmart results looks like:

```
## [1] "Here is the success rate: "

##   symbol ARMA.good HW_good ARMA.poor HW_poor GOOD_PERCENT
## 1    WMT      0.48    0.45       NaN     NaN            1
```

And for the Google we get:

```
## [1] "Here is the success rate: "

##   symbol ARMA.good HW_good ARMA.poor HW_poor GOOD_PERCENT
## 1   GOOG      0.48    0.49       NaN     NaN            1
```

The results shows the succcess rates are about 50% which is not good to rely on the predictions, if we can achieve a success rate above 60 percent that would be nice. Finally I want to try something else, let's increase the length of our window, for instance let's use a window size of 200 and compare the results on the both of the above symbols:(Be aware that this is a computationaly expensive step and it will take some time to see the results)

```
## [1] "Here is the success rate: "

##   symbol ARMA.good HW_good ARMA.poor HW_poor GOOD_PERCENT
## 1    WMT      0.53    0.52       NaN     NaN            1
```

```
## [1] "Here is the success rate: "

##   symbol ARMA.good   HW_good ARMA.poor HW_poor GOOD_PERCENT
## 1   GOOG 0.6060606 0.5858586         1       0         0.99
```

The results looks intresting we get better results, I wonder if we can get even better results with a bigger window size(size=400)?

```
## [1] "Here is the success rate: "
```

```
##   symbol ARMA.good HW_good ARMA.poor HW_poor GOOD_PERCENT
## 1    WMT      0.55    0.56       NaN     NaN            1

## [1] "Here is the success rate: "

##   symbol ARMA.good HW_good ARMA.poor HW_poor GOOD_PERCENT
## 1   GOOG      0.63    0.54       NaN     NaN            1
```

There is an improvement on both of the results, and this could be a good starting point for the future experiments. Now we should try different window length and find the optimal one. Does all stocks perform better when they have larger window length?