How Does the Performance in Different Industries Affecting the US Market*

Finding correlation between S&P 500 and 5 different industries during 3 famous recessions

Zecheng Wu

27 April 2022

Abstract

The S&P 500 index is the best overall measurement for the US market in which it is important for investors to determine whether they should keep more cash or invest more in the market. We have used the S&P 500 PE ratio, a ratio that measures the current share price relative to its earnings per share, to find out the time period for 3 recessions in 2002, 2018 and 2020. Then three tables are created which only contain the closing value for the S&P, AAPL(technology), MAR(real estate), SAFM(agriculture), NEM(gold) and XOM(gas) for each recession. The discovery from the result concludes that agriculture does not seem to get affected by recessions from the market and vice versa. In addition, I have created a multiple linear regression model by using the data from last year till this day to see if we can see any signs that the market might crash. We never know when the market will crash, but what we can do is to search for similar patterns from history and avoid our lost in the future.

#Introduction

Recession has always been the fear for people since 2020 due to the COVID pandemic. Governments have printed more money and tried to save businesses when many of them declared bankruptcy that caused the overall market to go down. The action of printing more money creates inflation, but the price for many fundamentals and food didn't seem to inflate in 2021. Many large investing companies have left their position in the trading market and kept more cash in their hands. This would cause the market to have less cash flow and many people believe that the market will crash eventually. However, nothing happened until 2022 February 24, 2022 when Russia invaded Ukraine and many investors started to trade again. Many sanctions against Russia appeared and the current market is at a very dangerous position right now. Nobody knows when the war will end, and some people believe that this war is a buffer for the recession. Recession is not a bad thing, it gives opportunities for many new companies to startup in their country. Uber, Instagram, WhatsApp, etc that became successful after the recession in 2008(foxbusiness). Thus, we want to know which industries might fall and which industries might rise during a recession.

We begin with the S&P 500 PE ratio (S&P PE ratio Y chart) which simply means a higher PE ratio implies the S&P 500 index is more overvalued. Since the S&P 500 is a U.S index, all the companies which I have selected to compare with the index are also from the U.S. I marked the starting and ending point for the peaks. Surprisingly, the years within these peaks are at the exact same time period as the recessions. Knowing that in 2002 it was the tech bubble and in 2008 it was the housing bubble, I have selected Apple(AAPL) as the representative for the technology industry and Marriott International(MAR) for the real estate industry. The sanctions on Russia have largely affected the gas price worldwide recently, thus I have added the Exxon Mobil Corporation(XOM) for the gas industry. Sanderson Farms(SAFM) and Newmont Corporation (NEM) are the representatives for agriculture and gold. I collected these data from yahoo finance starting from 1990 January 1 to 2022 April 21 by daily and insert them into R studio(R Core Team 2020). The first table combines the closing value for S&P 500 and these industry representatives from 2001 January 01 to 2002 september 30. This is because the first peak for the S&P 500 PE ratio is between 2001 first quarter to 2002

---

*Code and data are available at: https://github.com/nostestwu/market_prediction

third quarter. Next, I have created a multiple linear model setting the S_P as the response variable and five other industries as the explanatory variables. Lastly, I use the summary function to see which variables are significant to our model. I have repeated the same process for the "Great Recession" from 2008/07/01 to 2010/01/01 and the "Covid Recession" from 2020/01/01 to 2021/09/30. In both 2002 and 2008 recession model summary, the agriculture industry does not appear to be significant meaning that this industry is not correlated to the market. This concludes that our best fit model will only contain the values from four industries.

The predicting model I have created starts from 2021/04/01 to 2022/04/02 to test out whether our previous assumption was reliable. The exact same process was done except I have created 8 different models to see which one would be the best one to use for our current date. The result shows model 6 was the best model fit, which somehow it is the exact model that we got from the previous three recessions. These findings make me wonder if we know the performance for these four industries, does it mean that we can see the signs for recession? Before we want to predict these stocks, we want to see if the S&P 500 PE ratio has met our criteria. According to the PE ratio chart, every recession has hit above 40 at least. Currently we are at 24.09 right now which means a recession won't happen too soon.

The reason why this model is important is because most people want to have a brief idea about the market and how it is working but don't know where to start. By understanding the model, we can also think about our society logically. If everyone is losing money in the market, people will still spend money on food even if everything is inflated. The sanctions for the Russian War had a huge impact on gas and energy, but people will still pay more for them since these are the essential things they need to survive just like food. Thus, peace is the best solution for the economy.

The remainder of this paper is : Section **??** explains the data that we collected from yahoo finance.Section **??** shows the model that we are using. Section **??** explains what we found from the models. Section **??** explains interesting discoveries and future improvements.
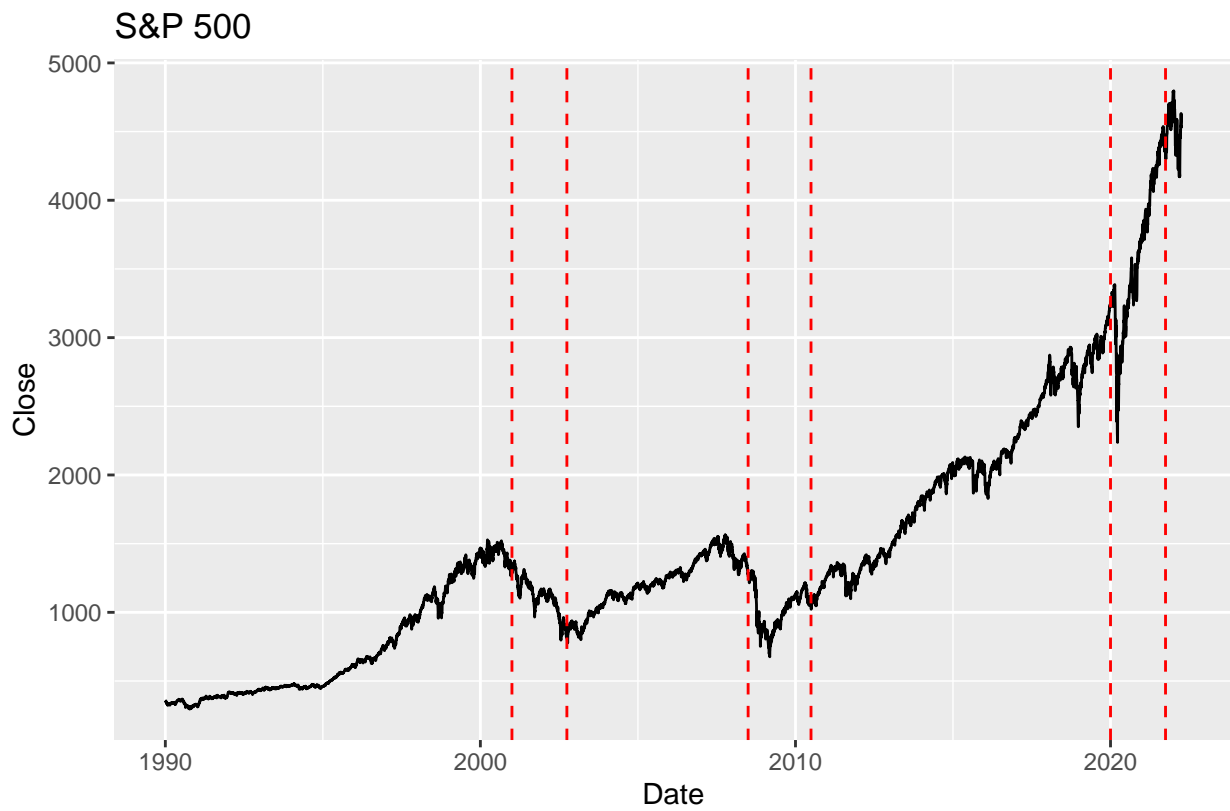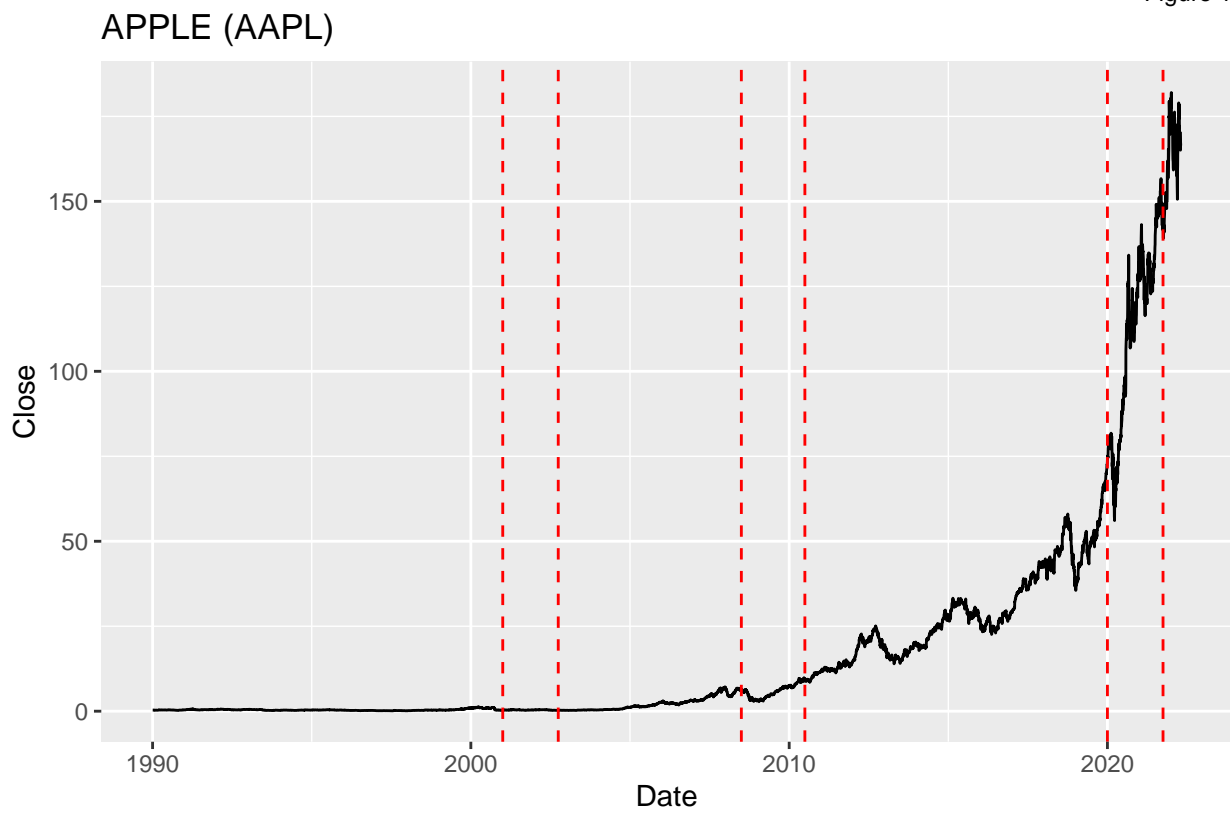
1 Data

S&P 500

Figure 1

APPLE (AAPL)

Figure 2

Marriott International (MAR)

Figure 3

Sanderson Farms, Inc. (SAFM)

Figure 4

5

Newmont Corporation (NEM)

Figure 5

Exxon Mobil Corporation (XOM)
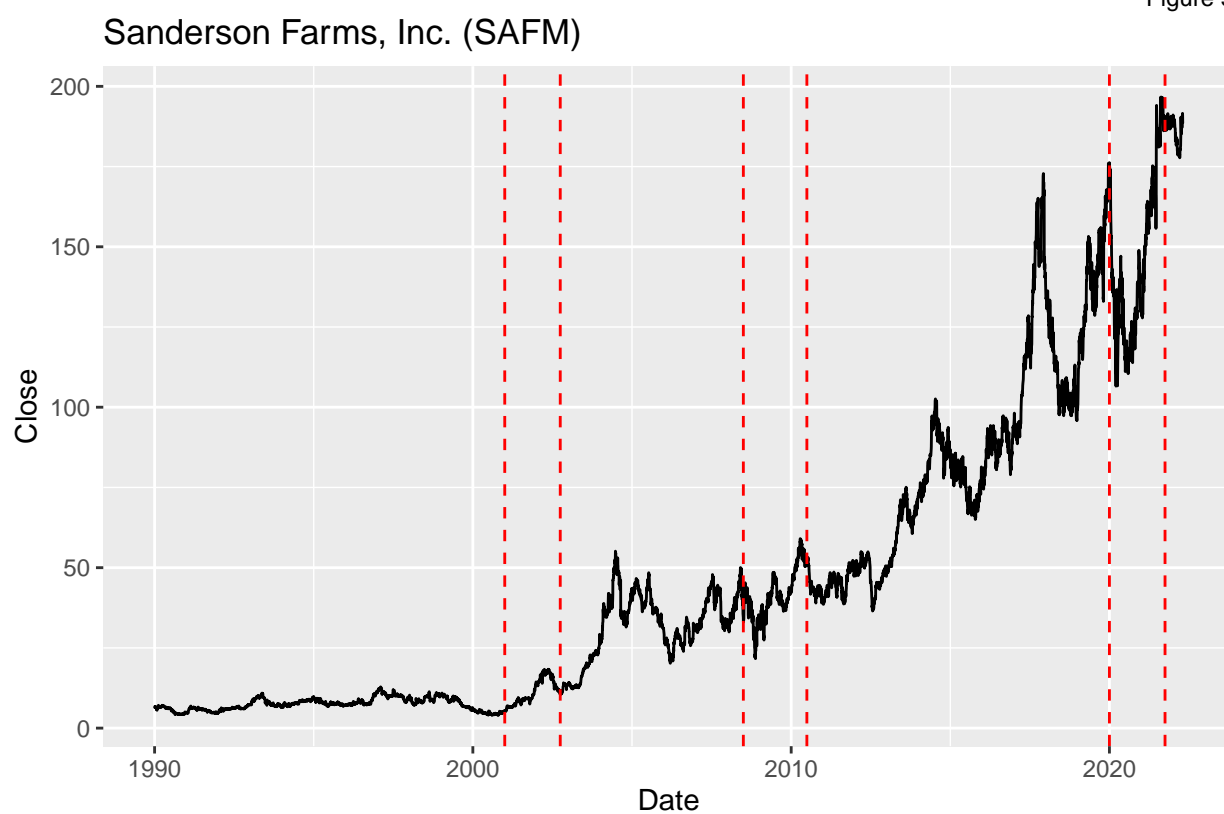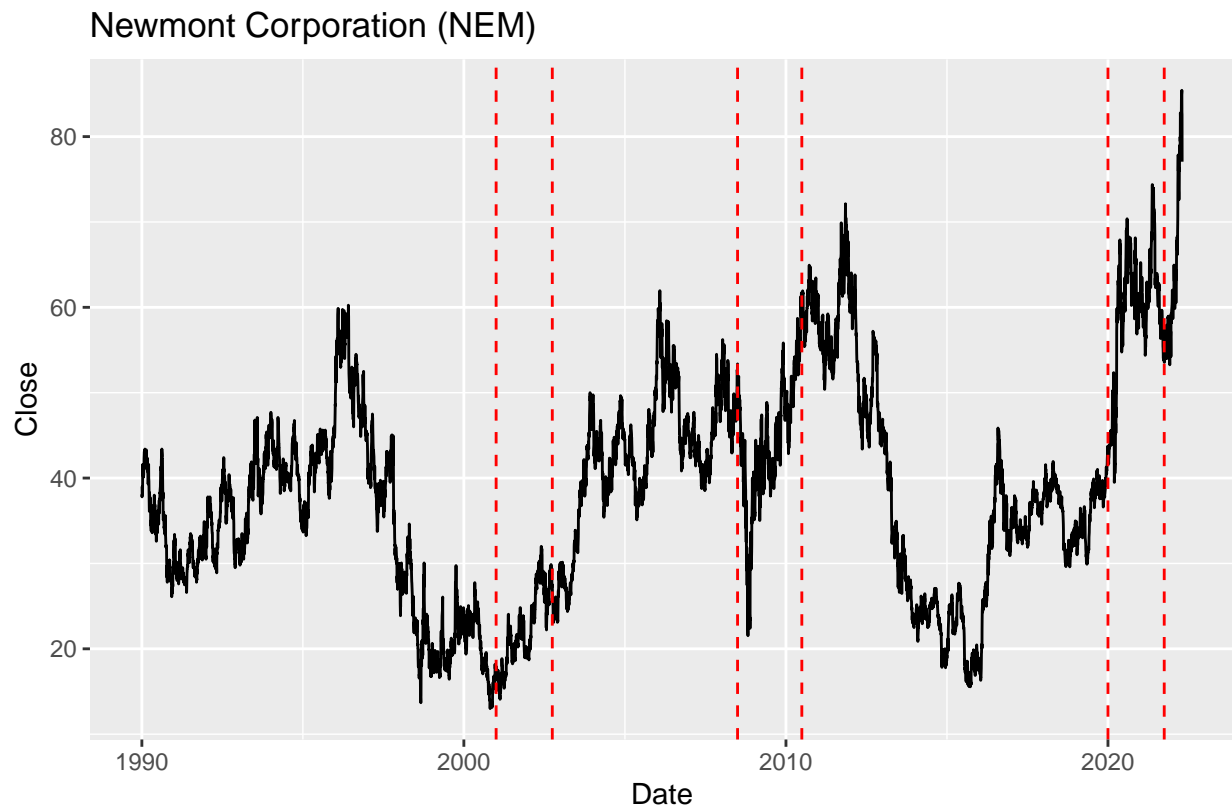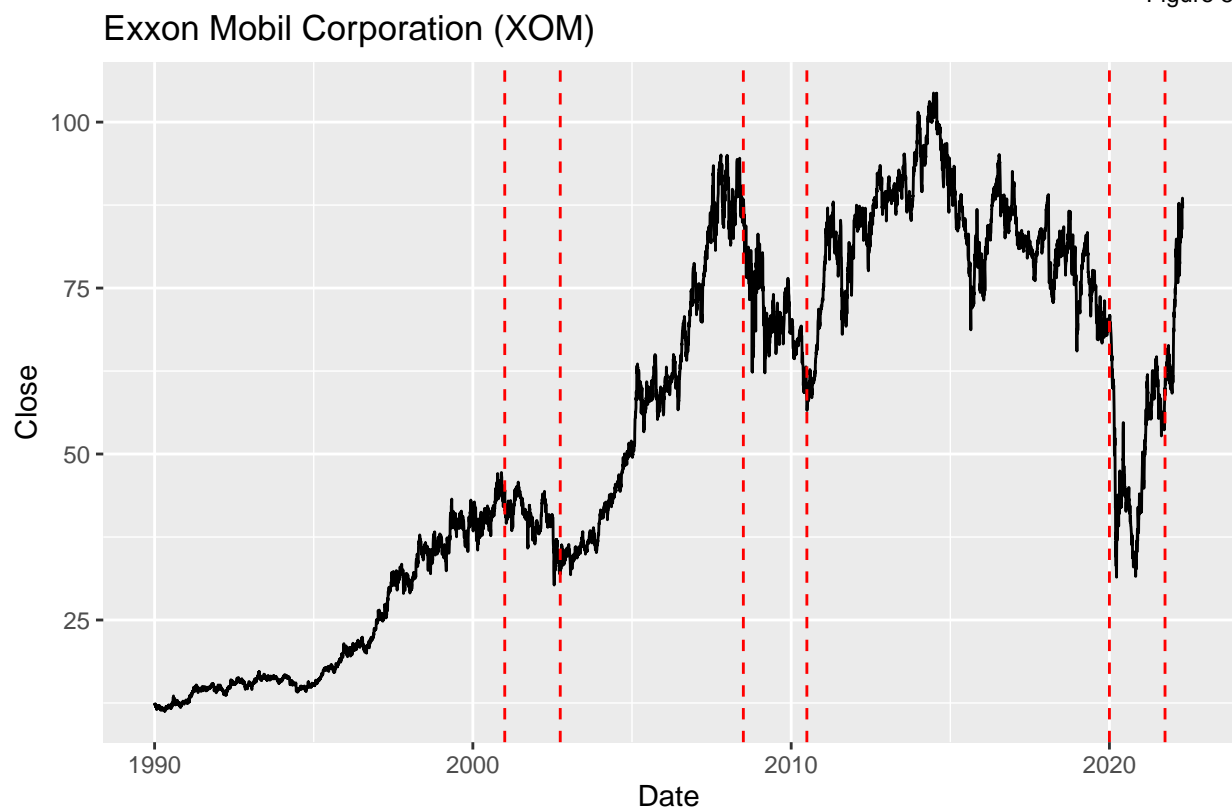

Figure 6

6

S&P 500 index tracks 500 publicly traded domestic companies in the U.S and the most well known index for people to use and see the overall U.S market. PE ratio stands for price per earning ratio, which means how many times the company is trading its earnings per share. For instance, a company has an earning of $100000 and 50000 shares outstanding. This means the earnings per share is $2 meaning each share you hold worth $2 dollars of its earnings. If this company offers each share $100, this means the PE ratio is 50, meaning this company is giving out the shares 50 times higher than its earning. It is obvious to see that we want to invest in companies that sell their shares very cheap and the earnings are super high which have a low PE ratio. Therefore, this is the reason why I consider the high PE ratio in S&P 500 to see when companies have their shares at a super high price while they are not making much money.

The five companies I have chosen to compare with the index are Apple(AAPL), Marriott International(MAR), Sanderson Farms(SAFM), Newmont Corporation (NEM), and Exxon Mobil Corporation(XOM). The reason why I have selected these companies is because all of these companies are one of the largest in their own industry and they have data since 1990. The first step I did is to download the daily data for each of these companies from yahoo finance and import them into rstudio. Then I went to the S&P 500 PE ratio website to record the time interval for each PE ratio peak. Unfortunately I do not have the account else I would have downloaded the actual data from the website. Then I have created a line graph for each of the companies and the index to see a general trend from 1990 to 2022 as we can see from Figure 1 to Figure 6. Instead of considering the entire chart, we want to break them into smaller parts which is the PE ratio time intervals we found earlier. I have combined the closed values from each company, renaming them and creating a model having the S&P as the y variable and different companies as the x variables. The first recession period starts from 2001/01/01 to 2002/09/30. I have repeated the same procedure for intervals from 2008/07/01 to 2010/06/30 and from 2020/01/01 to 2021/09/30, the summary tables are displayed as following:

```
## 
## Call:
## lm(formula = S_P ~ AAPL + MAR + SAFM + NEM + XOM, data = recession1)
## 
## Residuals:
##      Min      1Q   Median      3Q      Max 
## -107.738  -23.766    1.229   24.405  124.524 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 484.1241    35.9103  13.481  < 2e-16 ***
## AAPL        399.0394    57.0618   6.993 1.03e-11 ***
## MAR          14.0409     1.3439  10.448  < 2e-16 ***
## SAFM         -4.7873     1.0948  -4.373 1.54e-05 ***
## NEM         -10.2015     0.8797 -11.596  < 2e-16 ***
## XOM          12.7457     1.1765  10.834  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 42.19 on 430 degrees of freedom
## Multiple R-squared:  0.8892, Adjusted R-squared:  0.8879 
## F-statistic: 689.9 on 5 and 430 DF,  p-value: < 2.2e-16

## 
## Call:
## lm(formula = S_P ~ AAPL + MAR + SAFM + NEM + XOM, data = recession2)
## 
## Residuals:
##      Min      1Q   Median      3Q      Max 
## -129.395  -37.599   -2.671   35.392  141.673 
## 
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -382.4322    44.8769  -8.522  < 2e-16 ***
## AAPL          33.0345     4.4283   7.460 3.88e-13 ***
## MAR           20.8439     1.3667  15.251  < 2e-16 ***
## SAFM          -0.1320     0.6039  -0.219    0.827
## NEM           -4.5139     0.5958  -7.576 1.75e-13 ***
## XOM           13.2689     0.5534  23.978  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.27 on 498 degrees of freedom
## Multiple R-squared:  0.8808, Adjusted R-squared:  0.8796
## F-statistic: 736.1 on 5 and 498 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = S_P ~ AAPL + MAR + SAFM + NEM + XOM, data = recession3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -185.046  -46.356    7.147   49.129  142.232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 327.2688    38.1572   8.577   <2e-16 ***
## AAPL         13.3322     0.2727  48.886   <2e-16 ***
## MAR           2.9488     0.3143   9.382   <2e-16 ***
## SAFM          4.2479     0.2689  15.795   <2e-16 ***
## NEM           5.9555     0.6124   9.725   <2e-16 ***
## XOM           9.4123     0.7862  11.972   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68.16 on 435 degrees of freedom
## Multiple R-squared:  0.9847, Adjusted R-squared:  0.9845
## F-statistic:  5604 on 5 and 435 DF,  p-value: < 2.2e-16
```

2 Model

The model that we will be using is the multiple linear regression model. We will write the model as the following:

$$\bar{Y} = \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \beta_1 X_3 + \beta_1 X_4$$

$\bar{Y}$: the estimated S&P 500 index value

β_0: intercept

$\beta_1...\beta_n$: coefficents for X_1 to X_n

X_1: AAPL

X_2: MAR

X_3: NEM

X_4: XOM

3 Result

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A   Additional details

B   References

foxbusiness, https://www.foxbusiness.com/markets/startups-great-recession