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

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) 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 1 explains the data that we collected from yahoo finance.Section 2 shows the model that we are using. Section 3 explains what we found from the models. Section 4 explains interesting discoveries and future improvements.

1 Data

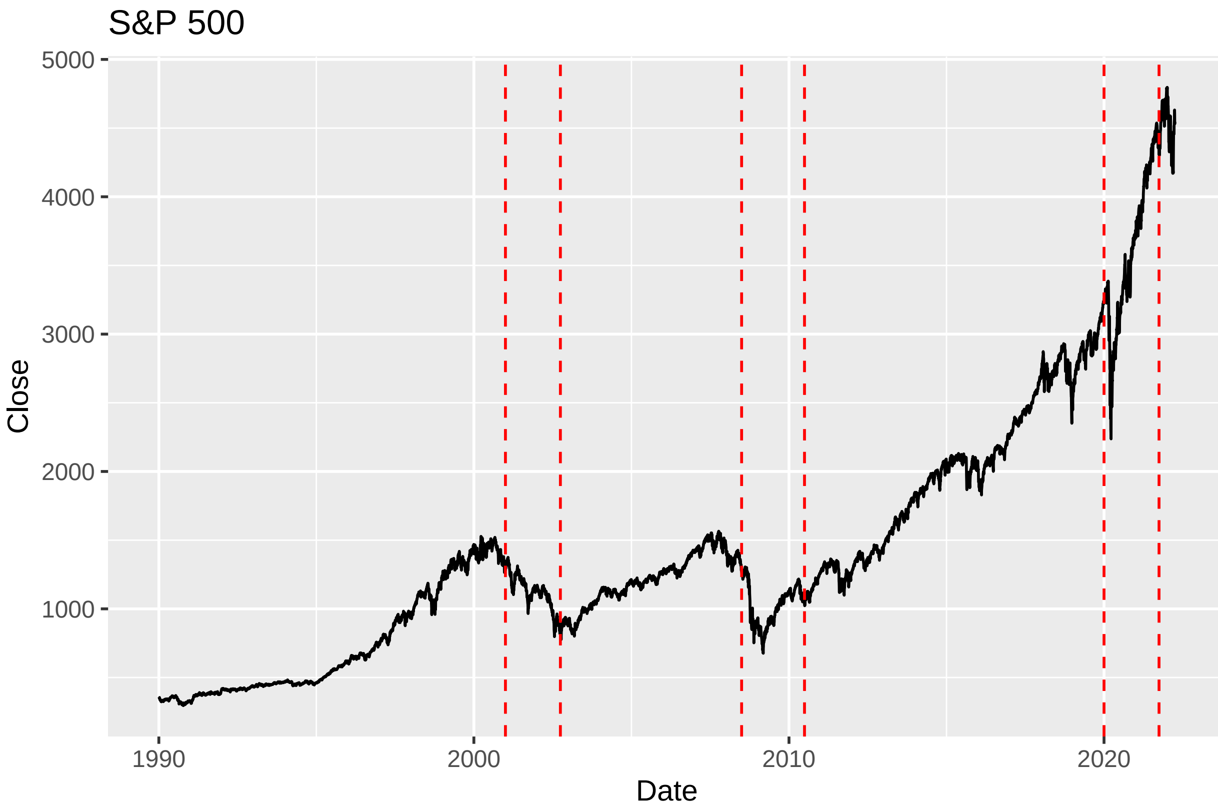

Figure 1

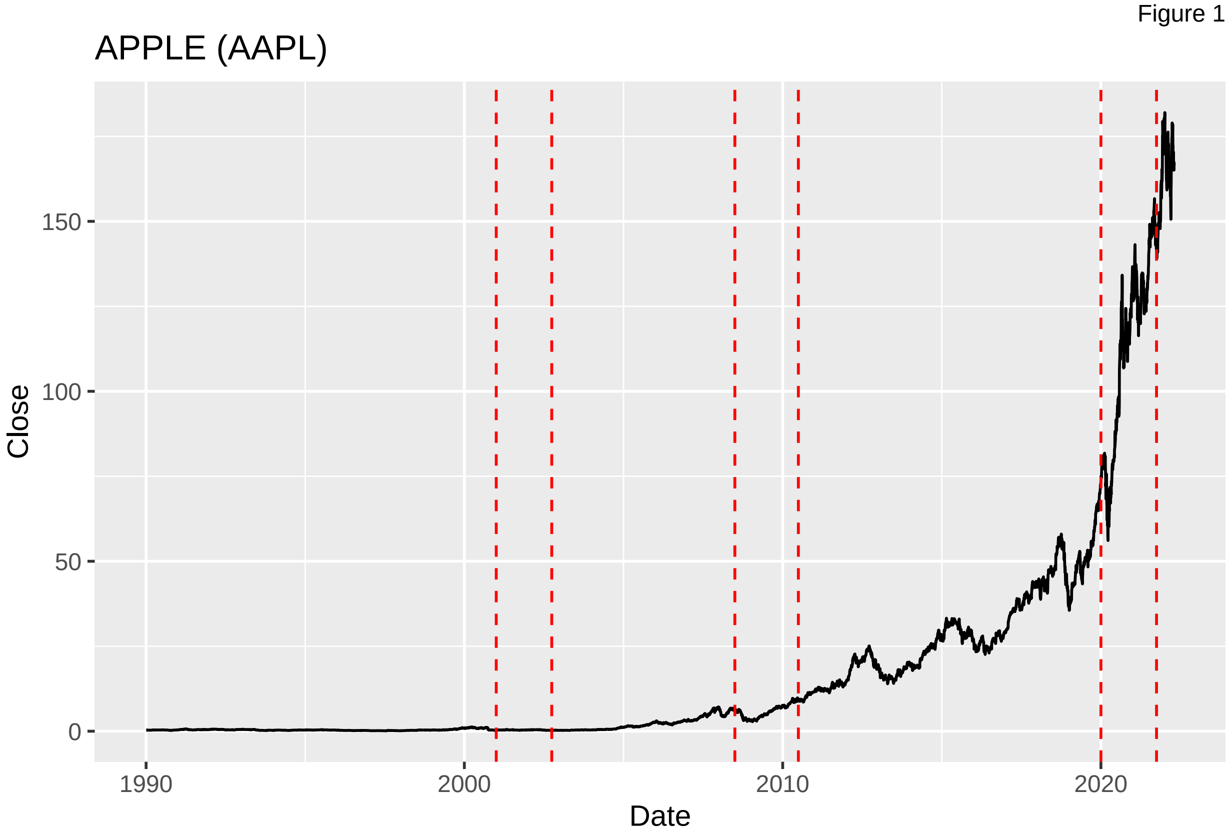

Figure 2

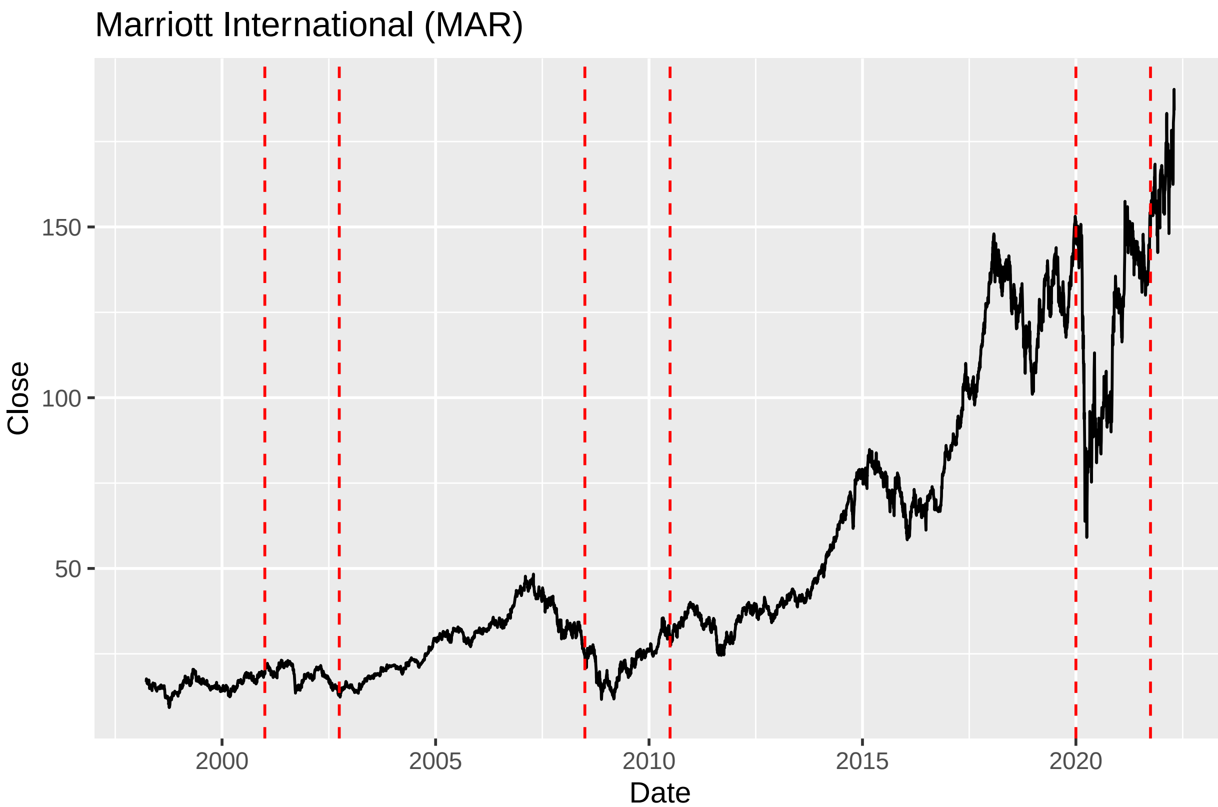

Figure 3

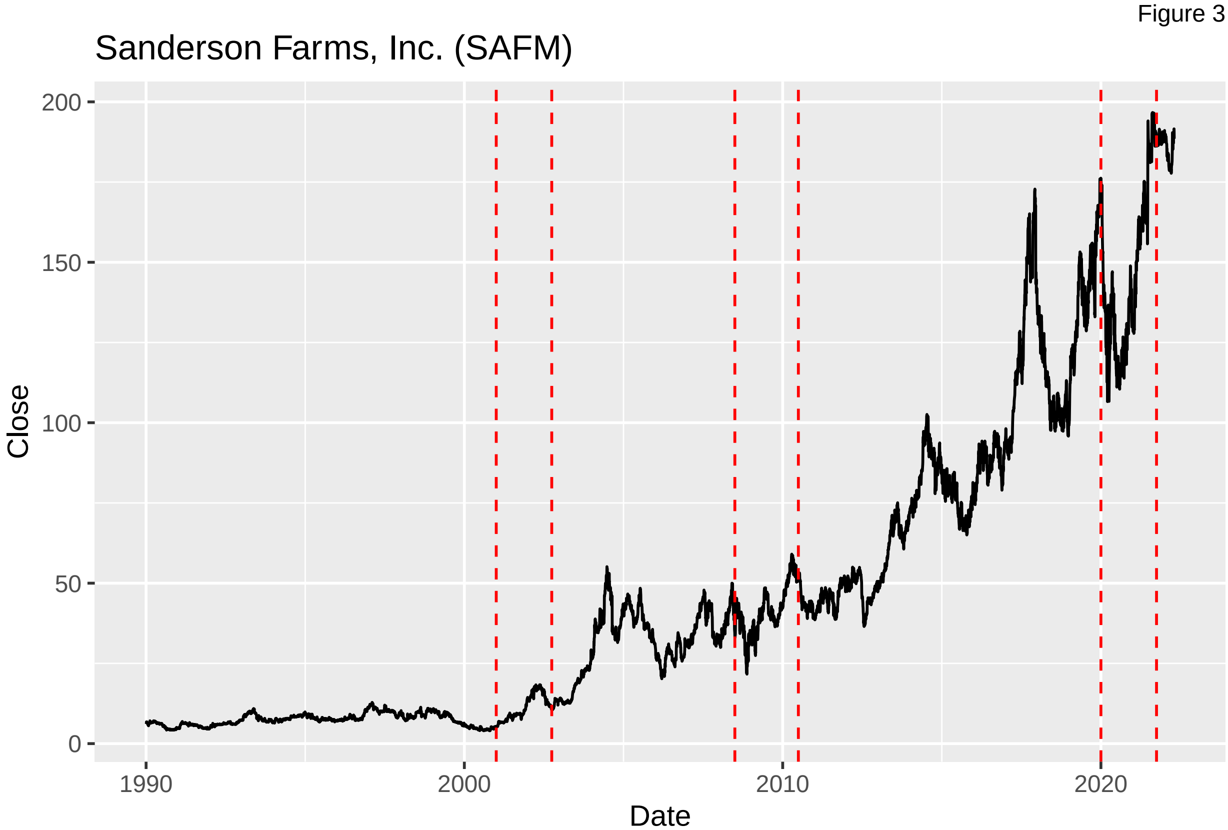

Figure 4

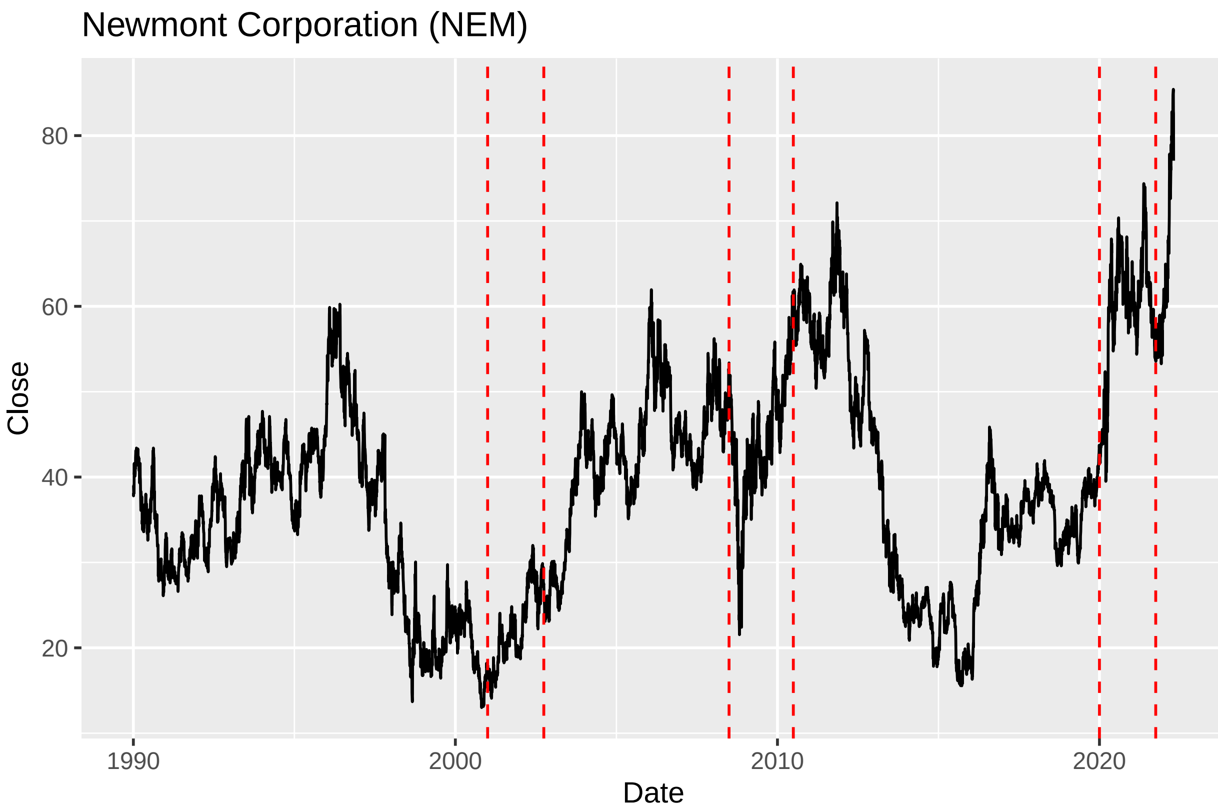

Figure 5

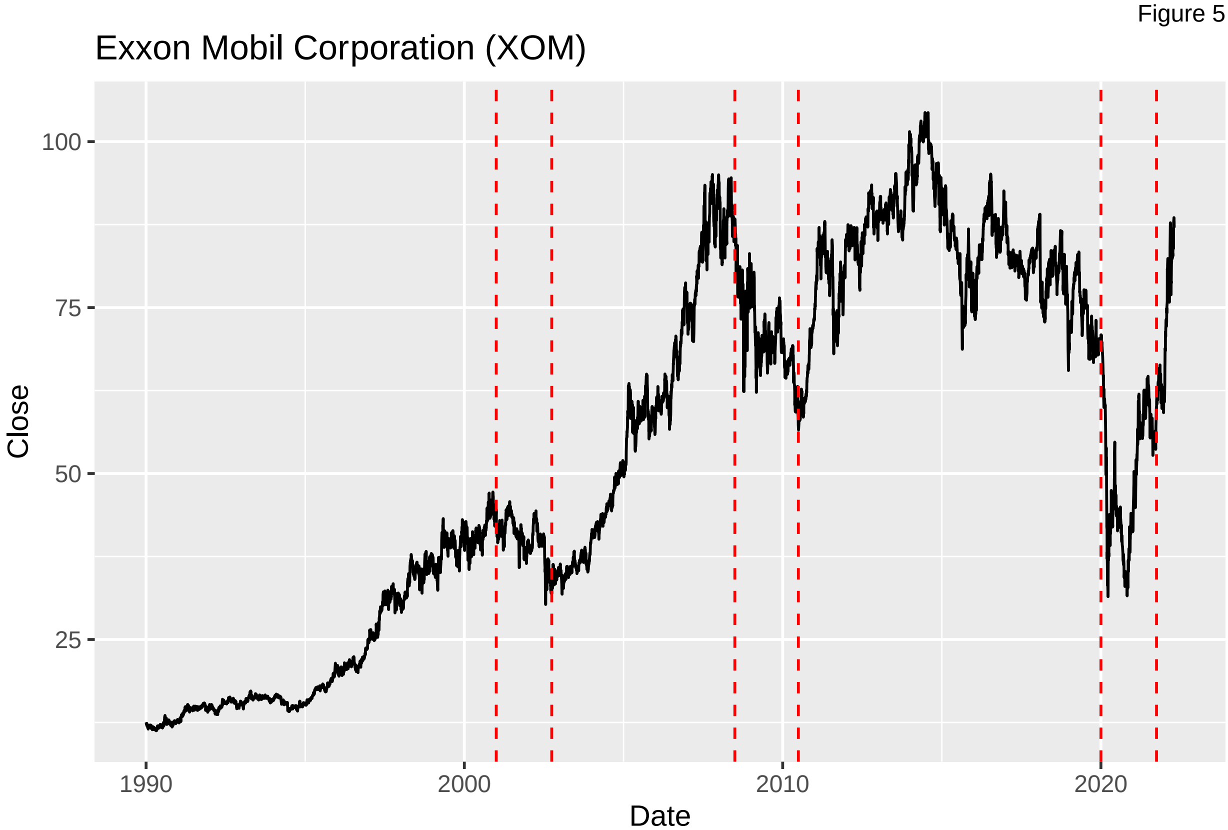

Figure 6

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
```

If we only consider the summary for the first and second recession, we can clearly see that SAFM is always the least significant variable for our model. Even though in recession 1, the P value for SAFM is very significant. However, the P value it has is relatively greater than all other variables. Thus, we don't want to include this variable in our prediction model.

After finding out the variables that are significant to determine the market, wouldn't it be more interesting if we can predict the future market trend? If we look closely at the time period for these recessions, we can see that they are around 1 year and 9 months to 2 years. Thus, I decided to take the data within a year from 2021/04/01 to 2022/04/01 and predict the market in the next 9 months. Simply repeating the same procedures except for the final step which we want to compare different models by keeping different x variables. I have created 8 different models and compared the summary between them. Then I picked 5 of the best summary models to predict the S&P value and compare it with the actual value on 2022/04/07, 2022/04/14 and 2022/04/22 which is a one week interval. I sum up the difference for each week and compare between these 5 models to see which one has the smallest difference. It turns out model6 has the smallest difference which we will be using this model for our prediction.

Table 1: Model Prediction

date	model5	model6	model7	model10	model11	Actual_S_P
2022/04/21	4453.653	4398.248	4399.394	4521.964	4401.448	4393.66
2022/04/14	4301.774	4327.545	4310.554	4491.662	4311.831	4392.59
2022/04/7	4256.542	4399.516	4409.574	4468.331	4408.805	4500.21

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 value from model6 (9.686)

X_2: MAR value from model6 (2.2753)

X_3: NEM value from model6 (-5.9445)

X_4: XOM value from model6 (-6.9975)

3 Result

If we compute the sum of the differences for all the models from Table 1, we would get -274.49 for model 5, -161.15 for model 6, -166.94 for model 7, 195.50 for model 10 and -164.38 for model 11. We can clearly see that model 6, model 7 and model 11 have similar values. Model 6 has 4 variables, model 7 has 3 variables and model 11 has 2 variables. Normally we would choose the model that has the least variable which is model 11. Considering the previous 3 recessions where each one of these variables are significant, the summary table in mode 6 shows that all variables are significant for our model.

Model 6 has positive coefficients for AAPL and MAR and negative coefficients for NEM and XOM. This is not something ideal to see because the gas price is increasing significantly after the Russian war started. If we see the exports from Russia, gold is also one of the top exports according to the OEC website (OEC) Sanctions on Russia could potentially raise the stock prices for both gas and gold industries since there would be less resources in the country without the imports from Russia. Our model shows that there is a negative correlation for both gas and gold with the S&P 500 index, a rapid increase of gas and gold price could lead to a rapid decrease in the overall market. In addition, many technology companies have moved out from Russia. This reduces the number of Russian customers, causing the price for technology companies to go down. Knowing that technology has a positive correlation with the index, decrease in technology will lead to a decrease in the overall market. This seems like the action of sanction for the U.S is speeding up the time for a new recession.

```
##
## Call:
## lm(formula = S_P ~ AAPL + MAR + NEM + XOM, data = clean20)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -198.544  -59.507    2.016   49.338  254.196
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3434.3374   109.4792  31.370  < 2e-16 ***
## AAPL           9.6860     0.5302  18.269  < 2e-16 ***
## MAR            2.2753     0.8061   2.823  0.00515 **
## NEM           -5.9445     1.2427  -4.784 2.95e-06 ***
## XOM           -6.9975     1.3666  -5.120 6.10e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.06 on 249 degrees of freedom
## Multiple R-squared:  0.7733, Adjusted R-squared:  0.7697
## F-statistic: 212.4 on 4 and 249 DF,  p-value: < 2.2e-16
```

4 Discussion

4.1 First discussion point

If we observe the graph for each company and the index, we can see that in 2015 the overall market dropped without any sign. After researching, it appears to be a market selloff during 2015 to 2016. This means even if the PE ratio for S&P doesn't rise, the overall market could still crash if all the investors are panicking about the current market. This could be the reason that explains why governments are printing more money to save the current market because everyone was too scared to invest after the COVID recession in 2020.

4.2 Second discussion point

After researching the 505 companies within the S&P 500, it turns out all the companies except for SAFM are included. This is why all companies have a strong correlation with the index except for SAFM. Even though this is a bad company selection, the purpose of this paper is to find the patterns during recession periods. In addition, we also want to know whether the variable is positively or negatively correlated with the S&P index. One of the methods that can improve this is to choose more companies from the same industry but not within the index.

4.3 Weaknesses and next steps

One major weakness is that we only selected one representative from each industry to conclude everything. If we can use more than 5 companies from each industry, then the data is more reliable to represent the entire industry growth. If we can compare the data with the index from other countries, then there might be some interesting observations that we can make. I had a quick glance at the SSE Index in China and the FTSE Index in London, both of them seem to have a rapid squeeze to the index and a rapid fall. Unlike the S&P where it falls rapidly at start but recovers back at a similar pace.Thus, comparing the data worldwide could find out more correlations and signs before each recession.

Appendix

```
setwd("/Users/nostest/Desktop/Github/market_prediction")
library(tidyverse)
library(ggplot2)

# Read in the raw data.

S_P <- readr::read_csv("inputs/data/HistoricalPrices.csv")
```

```
## Rows: 8127 Columns: 5
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (4): Open, High, Low, Close
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
AAPL <- readr::read_csv("inputs/data/tech_AAPL.csv")
```

```
## Rows: 8140 Columns: 7
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): Open, High, Low, Close, Adj Close, Volume
## date (1): Date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
MAR <- readr::read_csv("inputs/data/house_MAR.csv")
```

```
## Rows: 6062 Columns: 7
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): Open, High, Low, Close, Adj Close, Volume
## date (1): Date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
SAFM <- readr::read_csv("inputs/data/agri_SAFM.csv")
```

```
## Rows: 8140 Columns: 7
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): Open, High, Low, Close, Adj Close, Volume
## date (1): Date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
NEM <- readr::read_csv("inputs/data/gold_NEM.csv")
```

```
## Rows: 8140 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): Open, High, Low, Close, Adj Close, Volume
## date (1): Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
XOM <- readr::read_csv("inputs/data/gas_XOM.csv")
```

```
## Rows: 8140 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): Open, High, Low, Close, Adj Close, Volume
## date (1): Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#Cleaning S&P
S_P$Date <- as.Date(S_P$Date, "%m/%d/%y")

#### What's next? ####

S_P_graph <- ggplot(data = S_P, aes(x = Date, y = Close)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2001/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2002/09/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2008/07/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2010/06/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2020/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2021/09/30")), color = 'red', linetype = 2)


AAPL_graph <- ggplot(data = AAPL, aes(x = Date, y = Close)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2001/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2002/09/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2008/07/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2010/06/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2020/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2021/09/30")), color = 'red', linetype = 2)

MAR_graph <- ggplot(data = MAR, aes(x = Date, y = Close)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2001/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2002/09/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2008/07/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2010/06/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2020/01/01")), color = 'red', linetype = 2) +
```

```r
  geom_vline(xintercept = as.numeric(as.Date("2021/09/30")), color = 'red', linetype = 2)

SAFM_graph <- ggplot(data = SAFM, aes(x = Date, y = Close)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2001/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2002/09/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2008/07/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2010/06/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2020/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2021/09/30")), color = 'red', linetype = 2)

NEM_graph <- ggplot(data = NEM, aes(x = Date, y = Close)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2001/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2002/09/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2008/07/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2010/06/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2020/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2021/09/30")), color = 'red', linetype = 2)

XOM_graph <- ggplot(data = XOM, aes(x = Date, y = Close)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2001/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2002/09/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2008/07/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2010/06/30")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2020/01/01")), color = 'red', linetype = 2) +
  geom_vline(xintercept = as.numeric(as.Date("2021/09/30")), color = 'red', linetype = 2)


# Recession 1: tech bubble

r1_S_P <- S_P %>% filter(Date >= as.Date("2001/01/01") & Date <= as.Date("2002/09/30"))
r1_AAPL <- AAPL %>% filter(Date >= as.Date("2001/01/01") & Date <= as.Date("2002/09/30"))
r1_MAR <- MAR %>% filter(Date >= as.Date("2001/01/01") & Date <= as.Date("2002/09/30"))
r1_SAFM <- SAFM %>% filter(Date >= as.Date("2001/01/01") & Date <= as.Date("2002/09/30"))
r1_NEM <- NEM %>% filter(Date >= as.Date("2001/01/01") & Date <= as.Date("2002/09/30"))
r1_XOM <- XOM %>% filter(Date >= as.Date("2001/01/01") & Date <= as.Date("2002/09/30"))

#combining all tables

r1_S_P <- rename(r1_S_P, S_P = Close) %>%
  select(Date, S_P)
r1_AAPL <- rename(r1_AAPL, AAPL = Close) %>%
  select(Date, AAPL)
r1_MAR <- rename(r1_MAR, MAR = Close) %>%
  select(Date, MAR)
r1_SAFM <- rename(r1_SAFM, SAFM = Close) %>%
  select(Date, SAFM)
r1_NEM <- rename(r1_NEM, NEM = Close) %>%
  select(Date, NEM)
r1_XOM <- rename(r1_XOM, XOM = Close) %>%
  select(Date, XOM)
```

```r
clean1 <- inner_join(r1_S_P, r1_AAPL, by = "Date")
clean2 <- inner_join(clean1, r1_MAR, by = "Date")
clean3 <- inner_join(clean2, r1_SAFM, by = "Date")
clean4 <- inner_join(clean3, r1_NEM, by = "Date")
clean5 <- inner_join(clean4, r1_XOM, by = "Date")

#creating model

model1 <- lm(S_P ~ AAPL + MAR + SAFM + NEM + XOM , data = clean5)

# Recession 2: housing bubble

r2_S_P <- S_P %>% filter(Date >= as.Date("2008/07/01") & Date <= as.Date("2010/06/30"))
r2_AAPL <- AAPL %>% filter(Date >= as.Date("2008/07/01") & Date <= as.Date("2010/06/30"))
r2_MAR <- MAR %>% filter(Date >= as.Date("2008/07/01") & Date <= as.Date("2010/06/30"))
r2_SAFM <- SAFM %>% filter(Date >= as.Date("2008/07/01") & Date <= as.Date("2010/06/30"))
r2_NEM <- NEM %>% filter(Date >= as.Date("2008/07/01") & Date <= as.Date("2010/06/30"))
r2_XOM <- XOM %>% filter(Date >= as.Date("2008/07/01") & Date <= as.Date("2010/06/30"))

#combining all tables

r2_S_P <- rename(r2_S_P, S_P = Close) %>%
  select(Date, S_P)
r2_AAPL <- rename(r2_AAPL, AAPL = Close) %>%
  select(Date, AAPL)
r2_MAR <- rename(r2_MAR, MAR = Close) %>%
  select(Date, MAR)
r2_SAFM <- rename(r2_SAFM, SAFM = Close) %>%
  select(Date, SAFM)
r2_NEM <- rename(r2_NEM, NEM = Close) %>%
  select(Date, NEM)
r2_XOM <- rename(r2_XOM, XOM = Close) %>%
  select(Date, XOM)

clean6 <- inner_join(r2_S_P, r2_AAPL, by = "Date")
clean7 <- inner_join(clean6, r2_MAR, by = "Date")
clean8 <- inner_join(clean7, r2_SAFM, by = "Date")
clean9 <- inner_join(clean8, r2_NEM, by = "Date")
clean10 <- inner_join(clean9, r2_XOM, by = "Date")

#creating model

model2 <- lm(S_P ~ AAPL + MAR + SAFM + NEM + XOM , data = clean10)

# Recession 3: COVID

r3_S_P <- S_P %>% filter(Date >= as.Date("2020/01/01") & Date <= as.Date("2021/09/30"))
r3_AAPL <- AAPL %>% filter(Date >= as.Date("2020/01/01") & Date <= as.Date("2021/09/30"))
r3_MAR <- MAR %>% filter(Date >= as.Date("2020/01/01") & Date <= as.Date("2021/09/30"))
r3_SAFM <- SAFM %>% filter(Date >= as.Date("2020/01/01") & Date <= as.Date("2021/09/30"))
r3_NEM <- NEM %>% filter(Date >= as.Date("2020/01/01") & Date <= as.Date("2021/09/30"))
r3_XOM <- XOM %>% filter(Date >= as.Date("2020/01/01") & Date <= as.Date("2021/09/30"))
```

```r
#combining all tables

r3_S_P <- rename(r3_S_P, S_P = Close) %>%
  select(Date, S_P)
r3_AAPL <- rename(r3_AAPL, AAPL = Close) %>%
  select(Date, AAPL)
r3_MAR <- rename(r3_MAR, MAR = Close) %>%
  select(Date, MAR)
r3_SAFM <- rename(r3_SAFM, SAFM = Close) %>%
  select(Date, SAFM)
r3_NEM <- rename(r3_NEM, NEM = Close) %>%
  select(Date, NEM)
r3_XOM <- rename(r3_XOM, XOM = Close) %>%
  select(Date, XOM)

clean11 <- inner_join(r3_S_P, r3_AAPL, by = "Date")
clean12 <- inner_join(clean11, r3_MAR, by = "Date")
clean13 <- inner_join(clean12, r3_SAFM, by = "Date")
clean14 <- inner_join(clean13, r3_NEM, by = "Date")
clean15 <- inner_join(clean14, r3_XOM, by = "Date")

#creating model

model3 <- lm(S_P ~ AAPL + MAR + SAFM + NEM + XOM , data = clean15)


# Predicting model

p_S_P <- S_P %>% filter(Date >= as.Date("2021/04/01") & Date <= as.Date("2022/04/01"))
p_AAPL <- AAPL %>% filter(Date >= as.Date("2021/04/01") & Date <= as.Date("2022/04/01"))
p_MAR <- MAR %>% filter(Date >= as.Date("2021/04/01") & Date <= as.Date("2022/04/01"))
p_SAFM <- SAFM %>% filter(Date >= as.Date("2021/04/01") & Date <= as.Date("2022/04/01"))
p_NEM <- NEM %>% filter(Date >= as.Date("2021/04/01") & Date <= as.Date("2022/04/01"))
p_XOM <- XOM %>% filter(Date >= as.Date("2021/04/01") & Date <= as.Date("2022/04/01"))

#combining all tables

p_S_P <- rename(p_S_P, S_P = Close) %>%
  select(Date, S_P)
p_AAPL <- rename(p_AAPL, AAPL = Close) %>%
  select(Date, AAPL)
p_MAR <- rename(p_MAR, MAR = Close) %>%
  select(Date, MAR)
p_SAFM <- rename(p_SAFM, SAFM = Close) %>%
  select(Date, SAFM)
p_NEM <- rename(p_NEM, NEM = Close) %>%
  select(Date, NEM)
p_XOM <- rename(p_XOM, XOM = Close) %>%
  select(Date, XOM)

clean16 <- inner_join(p_S_P, p_AAPL, by = "Date")
clean17 <- inner_join(clean16, p_MAR, by = "Date")
clean18 <- inner_join(clean17, p_SAFM, by = "Date")
```

```
clean19 <- inner_join(clean18, p_NEM, by = "Date")
clean20 <- inner_join(clean19, p_XOM, by = "Date")

#creating model

model4 <- lm(S_P ~ AAPL + MAR + SAFM + NEM + XOM , data = clean20)
model5 <- lm(S_P ~ MAR + NEM + XOM , data = clean20)
model6 <- lm(S_P ~ AAPL + MAR + NEM + XOM , data = clean20)
model7 <- lm(S_P ~ AAPL + MAR + NEM , data = clean20)
model8 <- lm(S_P ~ AAPL + MAR + SAFM, data = clean20)
model9 <- lm(S_P ~ AAPL + MAR + SAFM + NEM , data = clean20)
model10 <- lm(S_P ~ AAPL + MAR + SAFM + XOM , data = clean20)
model11 <- lm(S_P ~ AAPL + NEM , data = clean20)
```

A References

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

Recessions history, https://en.wikipedia.org/wiki/List_of_recessions_in_the_United_States

Russia resources, https://oec.world/en/profile/country/rus#:~:text=Exports%20The%20top%20exports%20of,and%20Germany

Russian sanctions, https://graphics.reuters.com/UKRAINE-CRISIS/SANCTIONS/byvrjenzmve/

S&P 500 Historical Data, https://www.wsj.com/market-data/quotes/index/SPX/historical-prices

S&P 500 PE ratio, https://ycharts.com/indicators/sp_500_pe_ratio

Apple (AAPL), https://finance.yahoo.com/quote/AAPL/history?p=AAPL

marriott international(MAR), https://finance.yahoo.com/quote/MAR/

Sanderson Farms, Inc. (SAFM), https://finance.yahoo.com/quote/SAFM/history?p=SAFM

Newmont Corporation (NEM), https://finance.yahoo.com/quote/NEM/

Exxon Mobil Corporation (XOM), https://finance.yahoo.com/quote/XOM/

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.