

# Standard and Poors Stock Market value over the years\*

Mohammad Sardar Sheikh

27 April 2022

## Abstract

The purpose of this paper is to check the long term trends in the S&P500 stock index value. I want to see whether the numerous recessions and economic booms have had a significant difference on the value of the dividends and related factors. Methods that will be used to fit models will be justified and all assumptions will be checked for them. I have no further information to put in the abstract.

## 1 Introduction

This is still a basic introduction of the paper that I am planning to do. I want to predict long term trends of the S&P 500 and the short term trends. There have been various recessions and booms since it was founded and I want to see what the average rates of return were during these times and whether the stock owners got a different share of dividends during these times. The data that I have is monthly recordings of variables such as the value of the index to the earnings and dividends for the index and long term interest rates. We also have Consumer Price index data that has been used to normalise the values of the variables and get real values in return. I plan to make a number of visualisations that can be used to see the long term trends in the value, and also some visualizations that can be used to focus on smaller time periods, such as recession and or booms. I also plan to include an interactive graph (shiny maybe). After doing some literature review, I plan on fitting a model that can help predict the value of the index and I will choose relevant predictors to include in the model. I will make sure that the predictors have linear or no relationships with each other and that they do not increase the VIF of the model.

I am also interested in how a recession changes the value of the stock and what exactly happens to all the dividends and earnings and the CPI; so what I will do is make line graphs where I will zoom in.

This paper and research was done using R (R Core Team 2020). I used many other packages to help me in the creation of this paper. They include: TidyVerse (Wickham et al. 2019), kableExtra (Zhu 2021) for the informative tables, Patchwork (Pedersen 2020) for the graph layouts, readr (Wickham, Hester, and Bryan 2022) to help load and write data, lubridate (Grolemund and Wickham 2011) to help manage important dates, car (Fox and Weisberg 2019) to help with model assumptions and verification and lme4 (Bates et al. 2015) to help build mixed models.

## 2 key words (i dont know if i need this section)

S&P500. Stock Value. Model. Dividends. Price. Year. CPI. Economic timeline.

#key steps 1. analyze the actual data: what is it trying to tell me is there anything that I need to know to do my research

---

\*Code and data are available at: [LINK](#).

2. do some Heracles on the subject when were the recessions that i am gonna talk about when were the booms how am i gonna manipulate the data to show what i want in to show? damn this sounds pretty amazing doesn't it? lets get started now.
3. Have to gather data on the S&P 500. what is it, what does it actually represent, what stocks are part of it

### 3 Data

have to explain where i got the data from and how I am manipulating the dataset to make everything clean and organized.

For this paper and analysis, the data that I am working with is collected by the Economist Robert Shiller who is an Economist at Yale, and the dataset is open access at his website (Shiller, n.d.). We download the data directly from his website onto a local machine and then we can use that data. We can also access the data using a json file, the details of which are provided in the cleaning script. The next step is cleaning and organizing the data so that it can be used effectively in my analysis. All the variables have to be renamed and significant figures have to be checked so that they match. The dataset contains many variables that we do not have an interest in such as bond returns and values of the annualized stock returns that do not add too much information to the report but make it more complex and harder to replicate. The variables that I choose to keep are value of the index, value of the monthly dividends, value of the monthly earnings (all the variables mentioned above have both real and nominal values included in the final dataset). We also utilize information about the long term real interest rate of the economy (this is very important for investment decisions) and the consumer price index (this can be used to calculate inflation in the economy from year to year, or compared to a base year) for which the base year is 1983. We also include the CAPE ratio as one of the variables. This is important as (ADD INFORMATION ABOUT THE CAPE RATIO AND CITE IT IF POSSIBLE).

Another variable that I had to account for was the date. This needed modification as it was not in proper format. effectively utilizing lubridate allowed me to overcome this and create a clean variable. A variable that I created for the sake of the analysis was a year variable. This was made using the original date variable. The data that we had available was from 1871 onwards. However, I chose to start the analysis from the year 1913 as that is the first year that we have accurate data for regarding the consumer price index (CPI). Some of the variables that I considered important are shown in the graphs.

For displaying the data that we have, instead of choosing monthly values, I summarized them into yearly values and then plotted them so that they were not as cluttered and it would be easier to consider them as one off graphs.

We can see from Figure 1 that the value of the index has increased greatly since it was established in 1923. In 1923, it was launched on a very small scale. it started out with only 90 different stocks, but by 1957, it had increased that to 500 of the largest ones. We can see from the figure (which shows both real and nominal prices of the index) that a great proportion of that increase has taken place 1985 onwards, with very steep increases in market valuation for the stocks.

Figure 2 shows us the real value of the dividends for the index (A dividend is the distribution of some of a company's earnings to a class of its shareholders). We can see that while the value of the dividends rises just like the valuation of the index, it does so at a much slower rate and it is a lot less smoother, with many peaks and troughs.

Using Figure 3 we can see that the earnings increase steadily over a great period of time. We can also see near the end how the earnings take a huge plunge. This can most likely be attributed to the great recession of 2007-2008 and we will look deeper into this.

By the aid of 4, we can see that the average price of a bundle of goods in the economy has increased greatly since the measure was introduced in 1913. The CPI is used to determine the level of inflation between

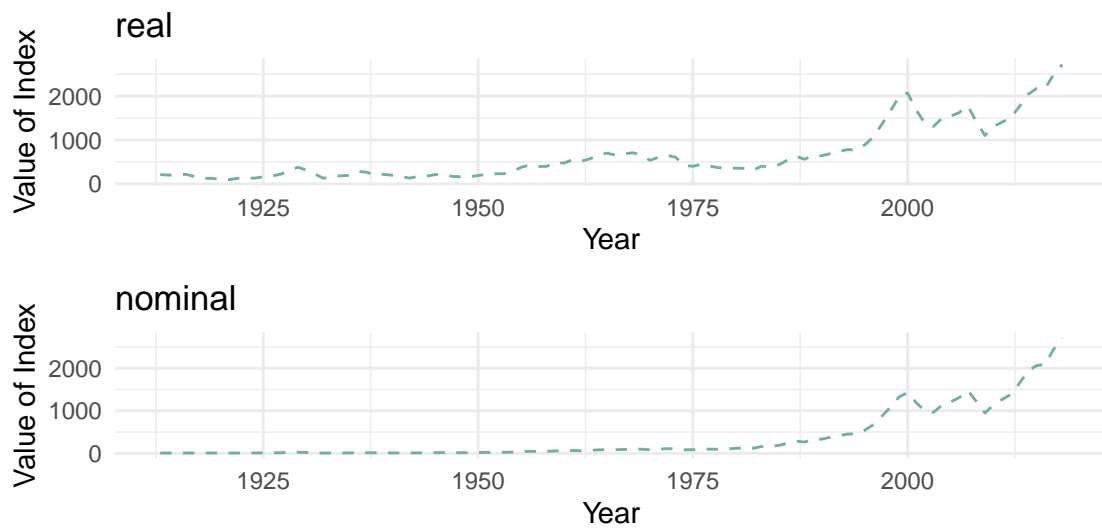


Figure 1: Average yearly S&P500 index value

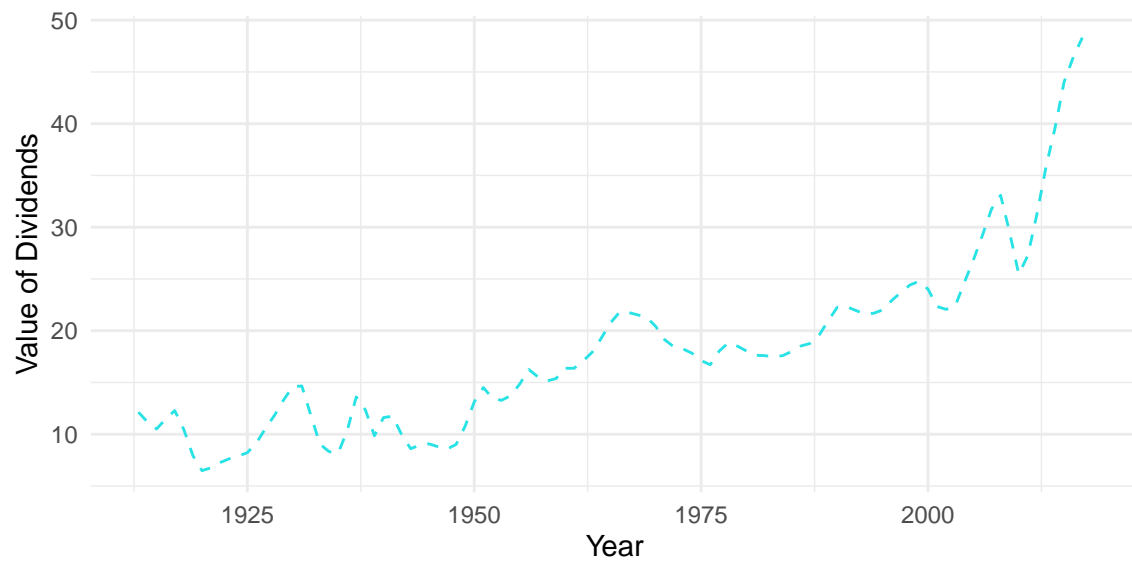


Figure 2: Average yearly S&P500 dividends (real)



Figure 3: Average yearly S&P500 earnings (real)

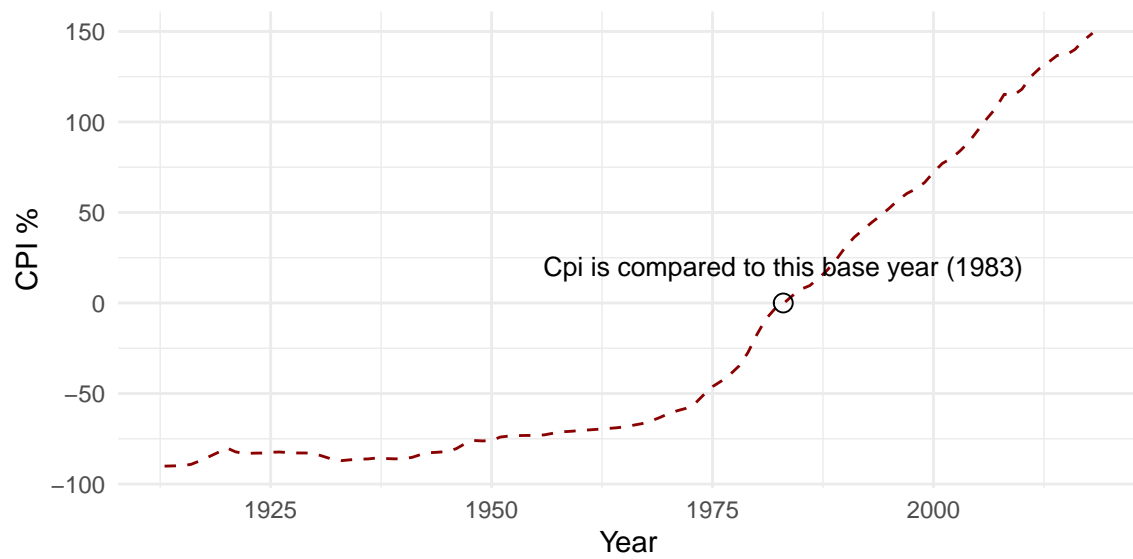


Figure 4: Average yearly Consumer Price Index

years. The US considers the base year for the CPI to be 1983. This means that all valuation of bundles and percentages are relative to 1983 dollars and compared to the price of the same bundle of goods in 1983. For example, we can see from the figure that the value of the CPI in 2005 was around 100%. We can interpret this as: The same bundle of goods in 2004 costs us twice as much compared to if it were purchased in 1983. So the inflation is effectively 100% between the two periods.

## 4 Model

I want to create 2 models as I want to be able to predict 2 things. the first one is the long term interest rate. the second model that i create is the real value of the index Now we will start thinking about what sort of model to use for our predictions. we can use any model that we prefer about it has to be well explained. i think that for my purposes, a linear model will be the best with just a few predictors (real values instead of nominal ones)

We want to be able to predict the value of the S&P 500 index based on the predictors that we have. Or can we do something about inflation? as that has an effect on the economy as a whole. . . . WE WILL USE long term real interest rate as the response variable. what i have to do is justify why i think this is an important thing to create a model for. Do I have to do any calculations for this?

the model that i chose has real interest rate as the response variable. now I have to justify why I think that a linear model is suitable and why interest rate is a suitable choice for the model

There are some things that I would like to predict based on the data that we have. Firstly, I would like to predict what the value of the index will be in the future. The S&P500 is one of the most popular indexes in the world, and its valuation is closely followed by millions. We can predict the state of the economy based on the value of the stock at a current period in time. based on the data exploration that has been done, and fulfilling model assumptions, the most appropriate predictors to use for the model would be the value of the dividends, value of earnings, the long term interest rate and the consumer price index (measure of inflation). We also take the natural log for the response (in this case it is the real value of the stock), to fulfill model assumptions. We will fit a linear regression model which has the following formula:

$$\log(\gamma_i) = \beta_0 + \beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3 + \beta_4\alpha_4$$

Explaining the different model things

- $\beta_0$  is the intercept term.
- $\gamma_i$  is the real price of the stock (we take the log).
- $\alpha_1$  is the long term real interest rate.
- $\alpha_2$  is the value of the dividends that investors get from the index.
- $\alpha_3$  is the value of the earnings (profits) from the index for a given period.
- $\alpha_4$  is the the consumer price index (base year is 1983).

Another model that is important to fit is to predict the value of the long term interest rate. The interest rate has been used by investors to figure out if a specific period is worth investing in. High interest rates dry up the investment in the economy as they increase the risk associated with investing, as there is a greater interest to be paid on the principal amount. During periods of high interest rates, people choose to save instead and this slows economic growth. The model that seems appropriate to fit includes the predictors such as real value of dividends, real value of earnings, consumer price index and the pe10 ratio (CAPE ratio). For the sake of fulfilling model assumptions, we take the logarithm of the response variable (interest rate). We also make sure that there isn't too much collinearity between the predictors and that the VIF (variance inflation factor) of the model is not too high. the final model can be represented as:

$$\log(\gamma_i) = \beta_0 + \beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3 + \beta_4\alpha_4$$

Explaining the different model things

- $\beta_0$  is the intercept term.
- $\gamma_i$  is the real interest rate that we get from the output of the model (we take the log for this to make it normal).
- $\alpha_1$  is the value of the dividends that investors get from the index.
- $\alpha_2$  is the value of the earnings (profits) from the index for a given period.
- $\alpha_3$  is the the consumer price index (base year is 1983).
- $\alpha_4$  is the CAPE ratio.

A linear model seems appropriate to predict both of these values.

#### 4.1 Model for interest rate:

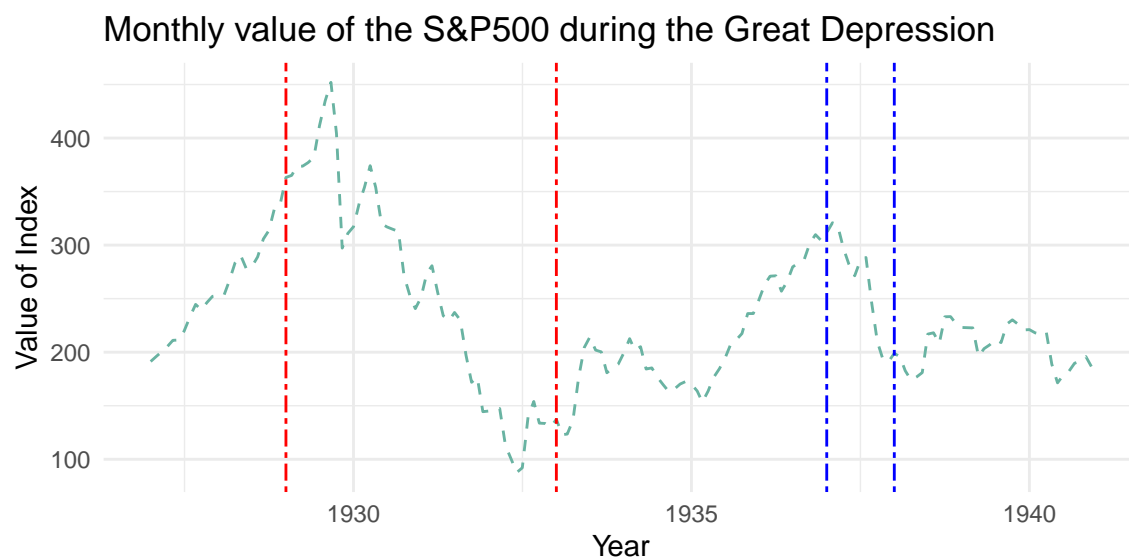
explain the VIF and things as well.

#### 4.2 Model for predicting price of the stock.

### 5 Results

I want to isolate the periods in the history of the United States when there were major recessions and then analyze how the price of the stock varied in those times. As explained, a recession is a period when the overall GDP growth rate of the economy falls as compared to the long run average growth (can go negative). In all of the recessions that I analyze, the GDP decreased by a huge margin and I want to analyze if this had any significant effects on the price of the S&P500 index value. I also want to predict both the value of the long term interest rate based on the components of the S&P500, and i want to be able to predict the value of the index in the future. To do this, I have created two models which will be interpreted.

#### 5.1 Great Depression

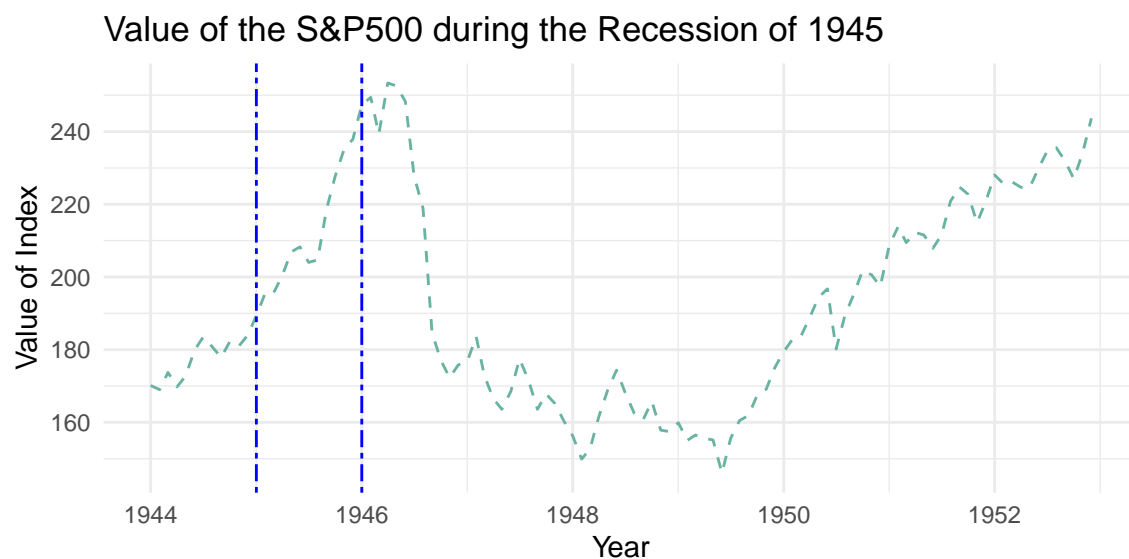


So we can see from the plot above that when the great depression hit (the largest recession in american history, where GDP fell by 27%), the value of the S&P500 fell by a great deal (add proper values here?).

this can be seen from the vertical red lines in the figure. that was the first wave of the recession. the second wave of the recession hit in 1937 and only lasted one year, but we can see that it already started to take effect on the value of the index (this is represented by the blue lines). HAVE TO ADD SOME NUMBERS ON HERE AS WELL.

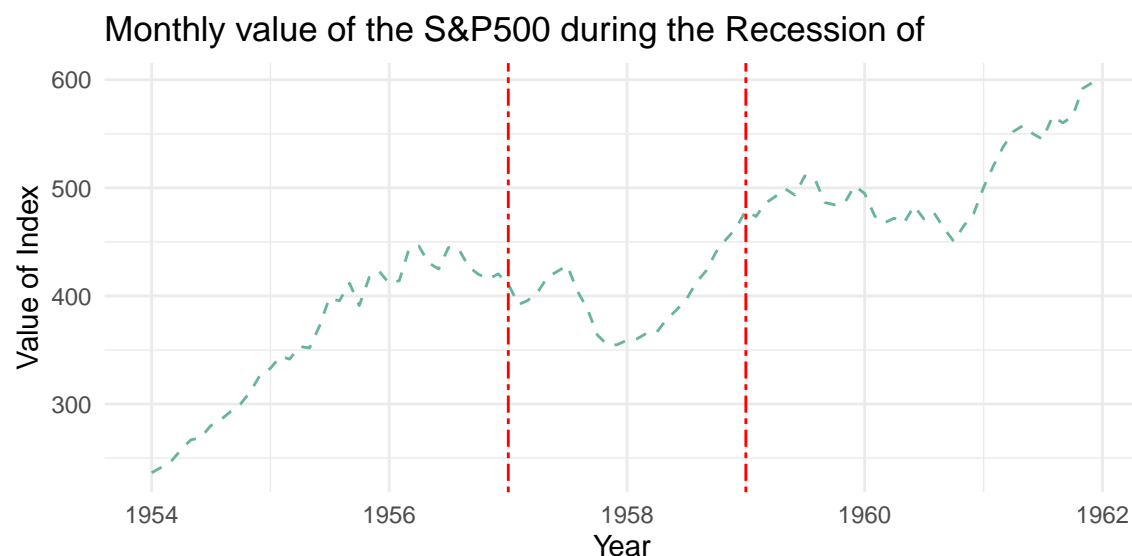
## 5.2 Recession at the end of world war 2

But with the surrender of both Germany and Japan in 1945, military contracts were slashed and soldiers started coming home, competing with civilians for jobs. As government spending dried up, the economy dipped into a serious recession with GDP contracting by a whopping 11 percent. this was actually a fiscal cliff



From the graph above we can see that the value of the index actually increased during the recession of 1945. This seems surprising as this recession in particular hit the United States very hard and there was a decrease of 12.7% in GDP. yet we can see that the value of the index remained strong. We have a counter argument since the effects of a recession are visible for a long period even after the recession has subsided. we can see from the figure above that this seems to be the case. since the recession was officially over within 8 months, the index did not have time to adjust to the fluctuations in the market and we can see this as a lagged effect since we observe a huge and long drop in the value of the index which took a few years just to recover and reach its original level. ADD SOME MORE AFTER SOME RESEARCH

### 5.3 recession on 1957-1958 Eisenhower recession.



We can see during this regression as well that the value of the index also fell somewhat. maybe add some numbers here as well? just include this as a figure to take up some space. what else can be added here. we can add some information about the specific regression and what caused it and how it affected the index (research wise).

This recession was significant in the post world war 2 period and unemployment rose greatly while corporate profits fell by as much as 25%. A weak fiscal policy contributed to this and there was significant stagflation as well as just a reduction in output. Stagflation occurs when the economy deteriorates and there is high inflation at the same time. this leads to a very bad situation for the country and it is what was seen in this particular recession.

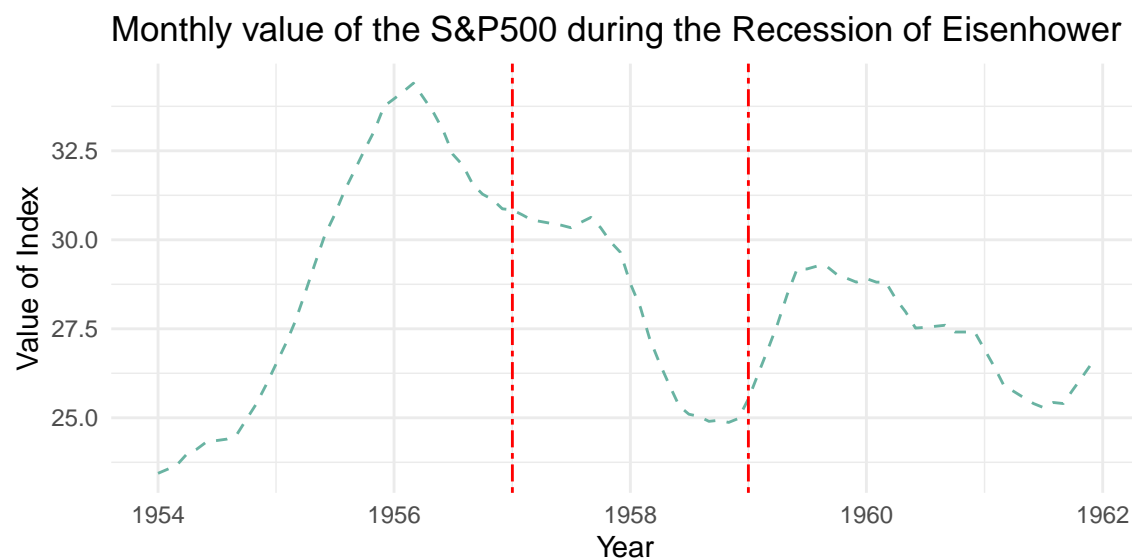




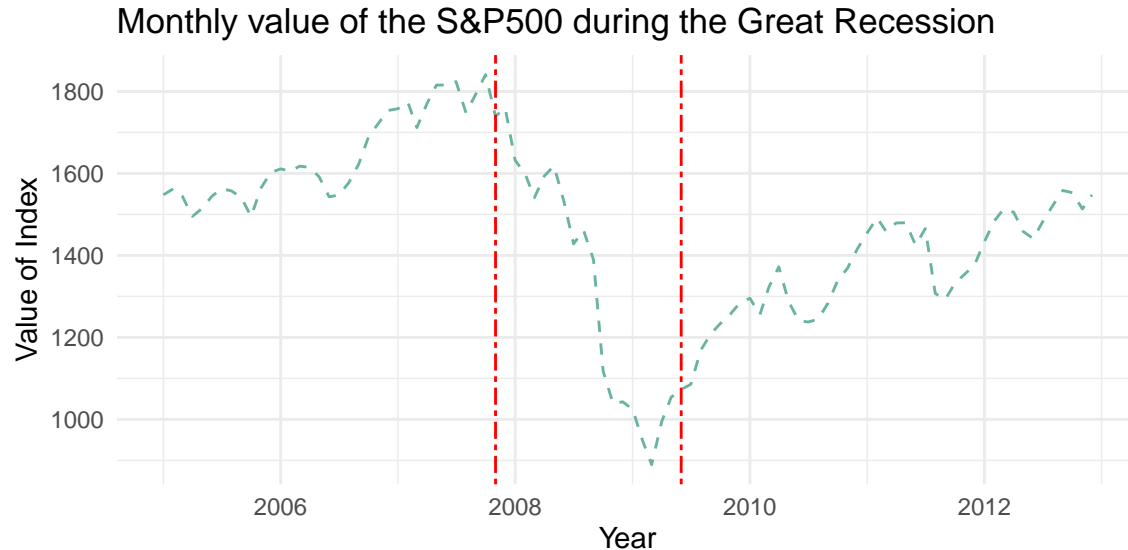
Table 1: Coefficients from Interest Rate Model

term	estimate	standard error	statistic	p-value
(Intercept)	6.275	0.229	27.353	0.000
real_dividend	-0.030	0.022	-1.414	0.158
real_earnings	0.010	0.007	1.395	0.163
consumer_price_index	0.013	0.002	6.412	0.000
pe10	-0.134	0.012	-10.766	0.000

Table 2: Coefficients from Price Model

term	estimate	standard error	statistic	p-value
(Intercept)	-65.394	18.127	-3.607	0
long_interest_rate	-33.915	1.957	-17.331	0
real_dividend	21.168	1.546	13.694	0
real_earnings	6.602	0.495	13.343	0
consumer_price_index	3.186	0.153	20.854	0

#### 5.4 the great recession:



AHH the great recession everyone that is reading this paper has experienced this period in our history. Not only was this specific regression in the United States, its effects spread worldwide, effecting so many other countries and continents. We can see just how devastating this regression was for the value of the stocks that were traded in the index as we can see a huge dip in the value of the standard and poors index in just the span of one year. going from a local maximum of 1800 to a local minimum of 900 dollars in real terms.

THEse are all the recessions that i have to account for.

model section of the results section

interest rate model:

talk about the results and what they mean here

price of stock model:

## **6 Discussion**

### **6.1 interpret the results of how the value of the stock fell during the major recessions**

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

### **6.2 The relationship between Long term interest rates and the economy**

### **6.3 The relationship between price of the stock and how it appears to be important (can include how the other stocks and the strength of the economy can be indicated by the performance of the S&P500 index and so this can be expanded into a well written point.)**

### **6.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

Include something about how I should get more information from the other indexes and how I can use the data to get more accurate results. Next steps can include. idk think about this.

## Appendix

### A Additional details

## References

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Shiller, Robert. n.d. “S&p 500 Data.” *Online Data - Robert Shiller*. <http://www.econ.yale.edu/~shiller/data.htm>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.