

# A Linear Regression Analysis on Predicting Economic Wellness Using the Stock Market Index

Chen Hao Yu

December 20, 2020

## Abstract

The study's purpose is to test if there is a correlation between the stock market index price to the wellness of the US economy, and whether or not we can use the index price as an indicator of the US economy. Using simple linear regression, the paper concludes that we can show a negative relationship between the S&P 500 index and the US unemployment rates, and a positive relationship between the index and the US weekly median income. We cannot use the index to accurately predict the US real GDP and real GDP per capita.

## Introduction

On March 9, 2020; the S&P 500 Index, a stock market index that measures the stock prices of the top 500 companies in the United States, dipped by a staggering 7.6% (BBC, 2020). The drop had been caused by the cut in oil prices and the coronavirus outbreak, and marked the end of the decade long bull market since the end of the Great Recession in 2008 (Mashayekhi, 2020). According to the Federal Reserve Bank of St. Louis, unemployment rates skyrocketed from a low 3.88% in the fourth quarter of 2019 to a high 13.07% at the end of the first quarter of 2020, while GDP per Capita had dropped by \$5243 US Dollars. However when the S&P index recovered to an record high of 3389.78 on August 18 (Schneider, 2020), the unemployment rate did not recover fully and was still at 8.4%, which was 4.7% higher than August 2019 (U.S. Bureau of Labor Statistics, 2020). Yet, President Trump has frequently attributed his ability to manage the US economy whenever the stock market seems to rise (Bloomberg, 2020). This leads us to an interesting question, can we use the S&P 500 index alone to measure the US economic performance?

To test if the S&P 500 index is a good indicator of the economy, we will be using the index as an explanatory variable to actual economic indicators using linear regression. We will be testing four economic indicators; unemployment rate, real GDP per capita, median weekly income (measured every quarter), and real GDP (Pettinger, 2019). The purpose of using the linear regression model is to see if we can find any correlation between the explanatory variable and the four response variables.

## Methodology

Our null hypothesis is that the changes in the S&P 500 index should have no effect on any of the four economic indicators, and we will be utilizing a significance level of 0.05.

## Data

Data from the S&P 500 Index will be extracted from Quandl, and will be average price of S&P 500 index every quarter. These prices will be adjusted to inflation in October 2020 dollars. In total there are 600 observations of the index prices ranging as far back as the first quarter of 1871; however, in our study we are only able to utilize data starting from the last quarter of 1946 due to missing observations from our response variables.

The four response variables are all from the Federal Reserve Bank of St. Louis. GDP per capita is measured every quarter.

Real GDP per capita is frequently quoted as a measure of standard of living for a country's population (Hall, 2020), and is calculated by the GDP divided by the population number of a country and then adjusted for inflation. The benefits of this measure is that it ignores inflation, and is used to assess the consumer spending of citizens (Amadeo, 2020). There are disadvantages with this measure; real GDP per capita does not incorporate the effects of pollution, health, and unpaid work such as housework (Amadeo, 2020). Additionally, GDP per capita is an average of GDP per person and does not include income inequality into the picture (Amadeo, 2020). Our data set for the US real GDP per capita is calculated every quarter from the last quarter of 1946 and all the way until the end of the second quarter of 2020.

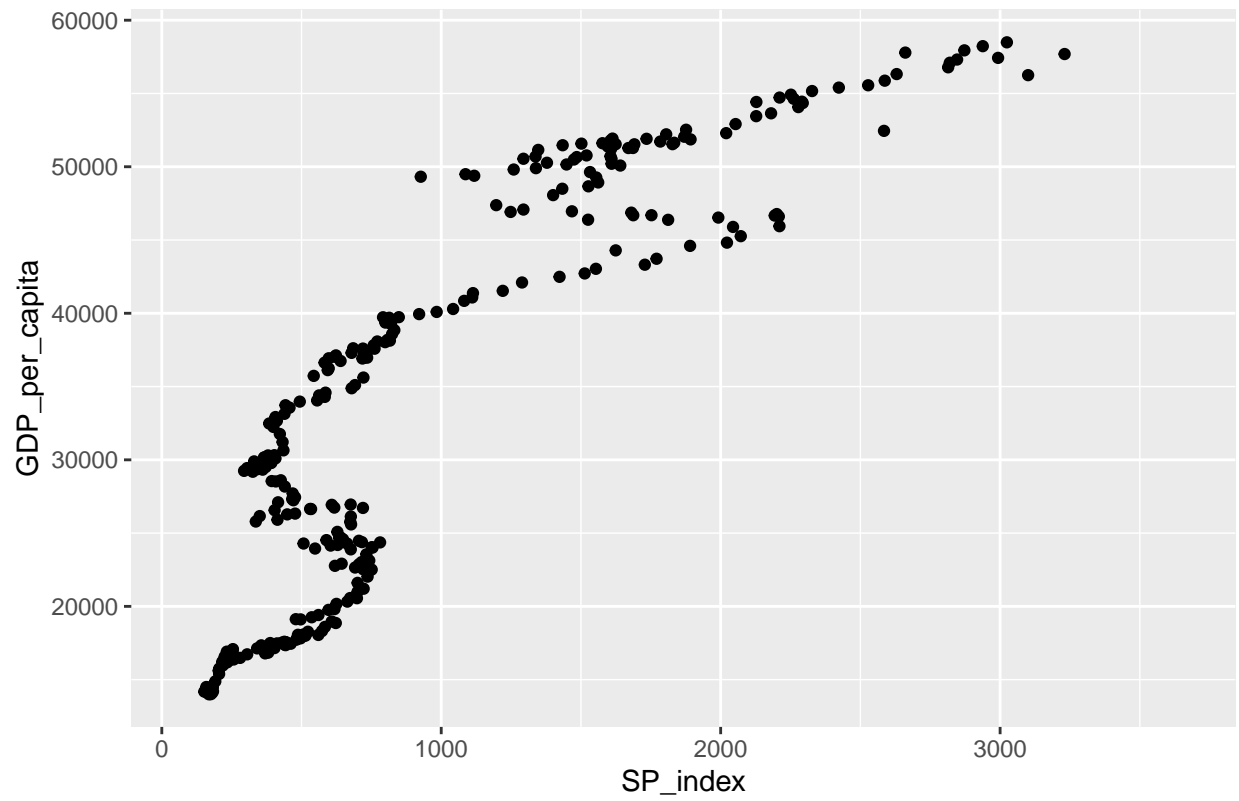
The unemployment rate represents the proportion of people in the labour force that do not have a jobless. The labor force includes people who are employed and jobless people who are actively searching for a job. Unemployment rate is a strong indicator of the economy as rising unemployment equates to a reduction of spending by consumer, which can cause a domino effect of companies cutting more workers as demand for goods decreases (Picardo, 2020). The disadvantages of unemployment rate is that it frequently underestimates the true unemployment rate, as workers can often be discouraged from finding a job due to economic conditions (Amadeo, 2020). Data for the US unemployment rates dates back to the last quarter of 1969 and all the way to second quarter of 2020 and is measured every quarter.

The US weekly median income is measured every quarter, and is calculated by looking at the weekly income of all workers and taking the median of it. Income is used as an economic indicator because it represents consumer spending, and therefore the strength of economic activity (Friedberg Direct, 2020). A similar category to the weekly median income is the gross national income, which calculates the income of a nation, and is used as a factor to calculate the Human Development Index of a nation (Hall 2020). However, the measure does not take into account of actual spending, because consumers can always decided to set aside money for saving. The data is collected using a survey of both public and private sector employees (but not the self-employed) from the last quarter of 1978 to the first quarter of 2020, and is adjusted to inflation.

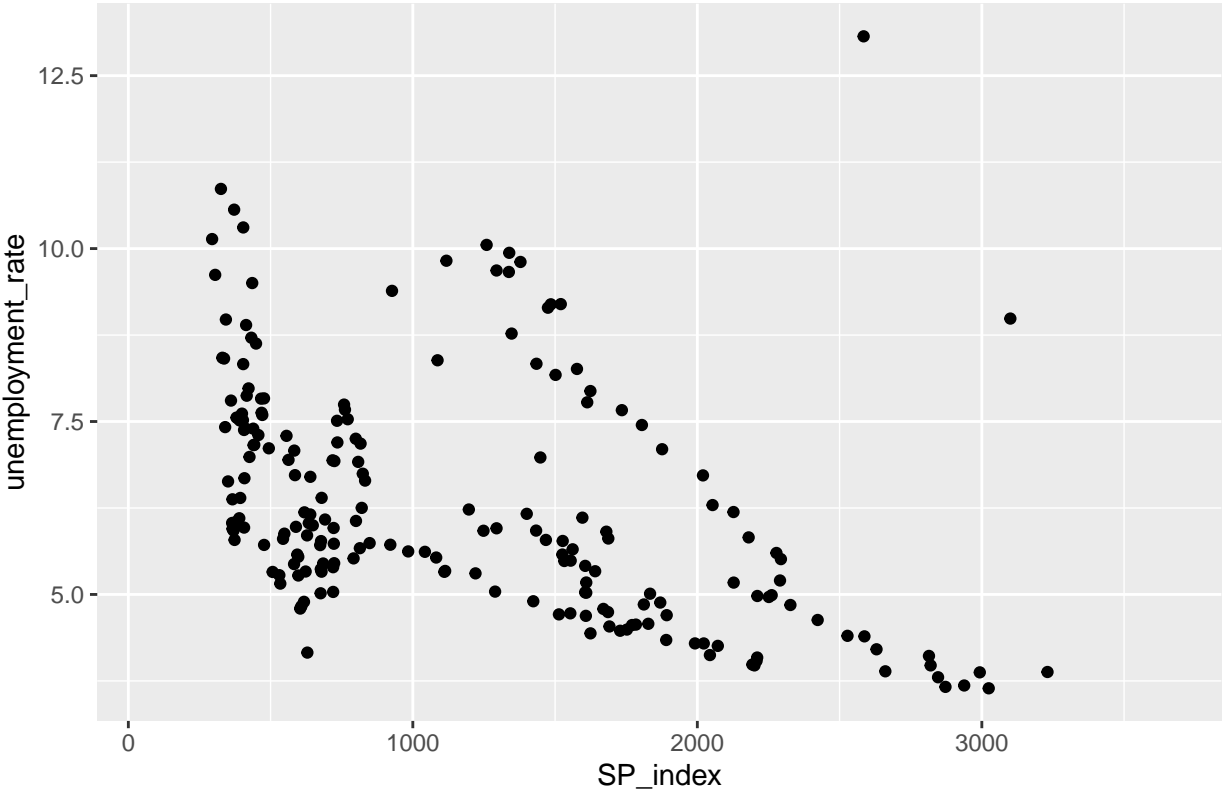
Real GDP represents the total output of goods and services by a country adjusted to inflation and is calculated by looking at four categories; consumption, investment, government expenditures, and net exports (Hall, 2020). A country's total GDP is expected to increase when the economy performs well; for example, increases in consumer spending helps businesses grow and hire more employees, and infrastructure projects by government spending results in a higher standard of living for people (Amadeo, 2020). The problems with GDP as an economic indicator is that it does not include citizens living overseas that send part of their wages back to their home country, resulting in a understatement of the income of the nation, and similar to GDP per capita, it does not address the problems with income inequality (Amadeo, 2020). US real GDP is calculated at the end of every quarter, and data points go all the way back to the last quarter of 1946 up to the second quarter of 2020.

Here are each of the response variables plotted against the S&P 500 index:

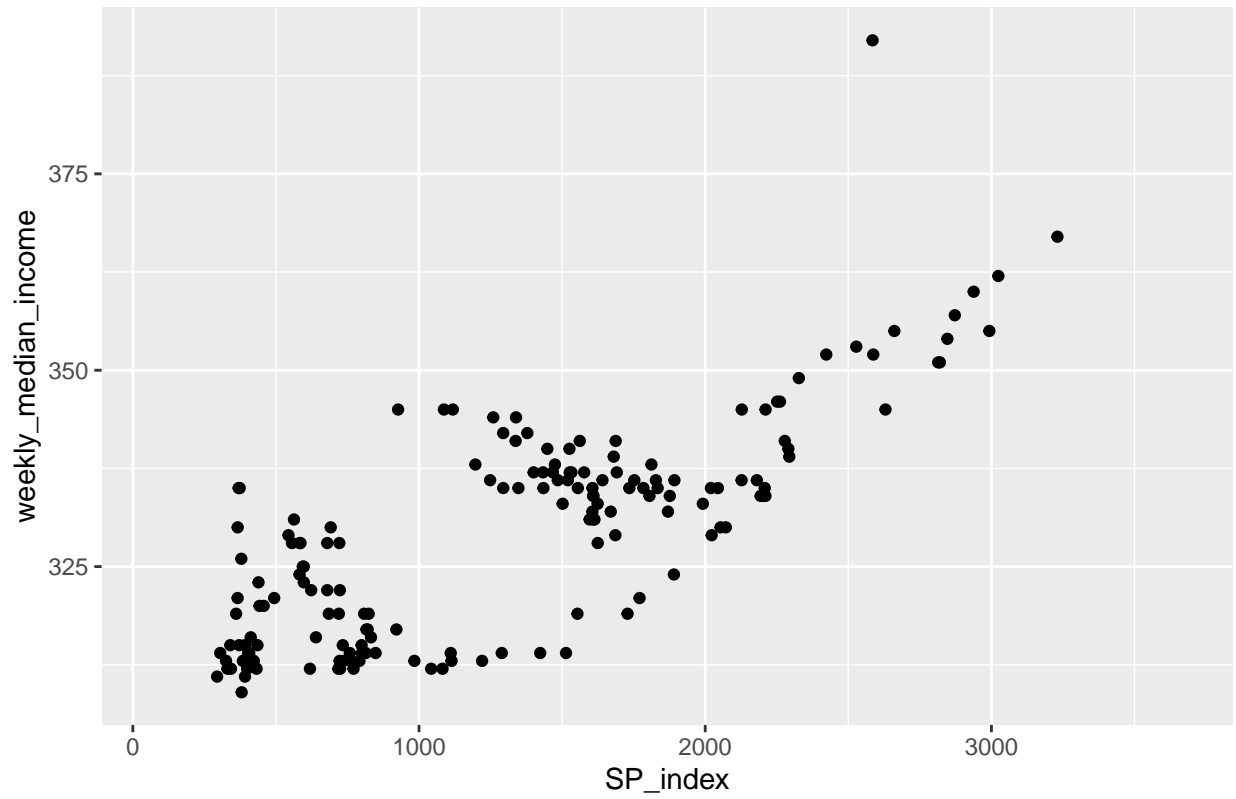
Scatterplot of Real GDP per Capita vs The S&P 500 Index

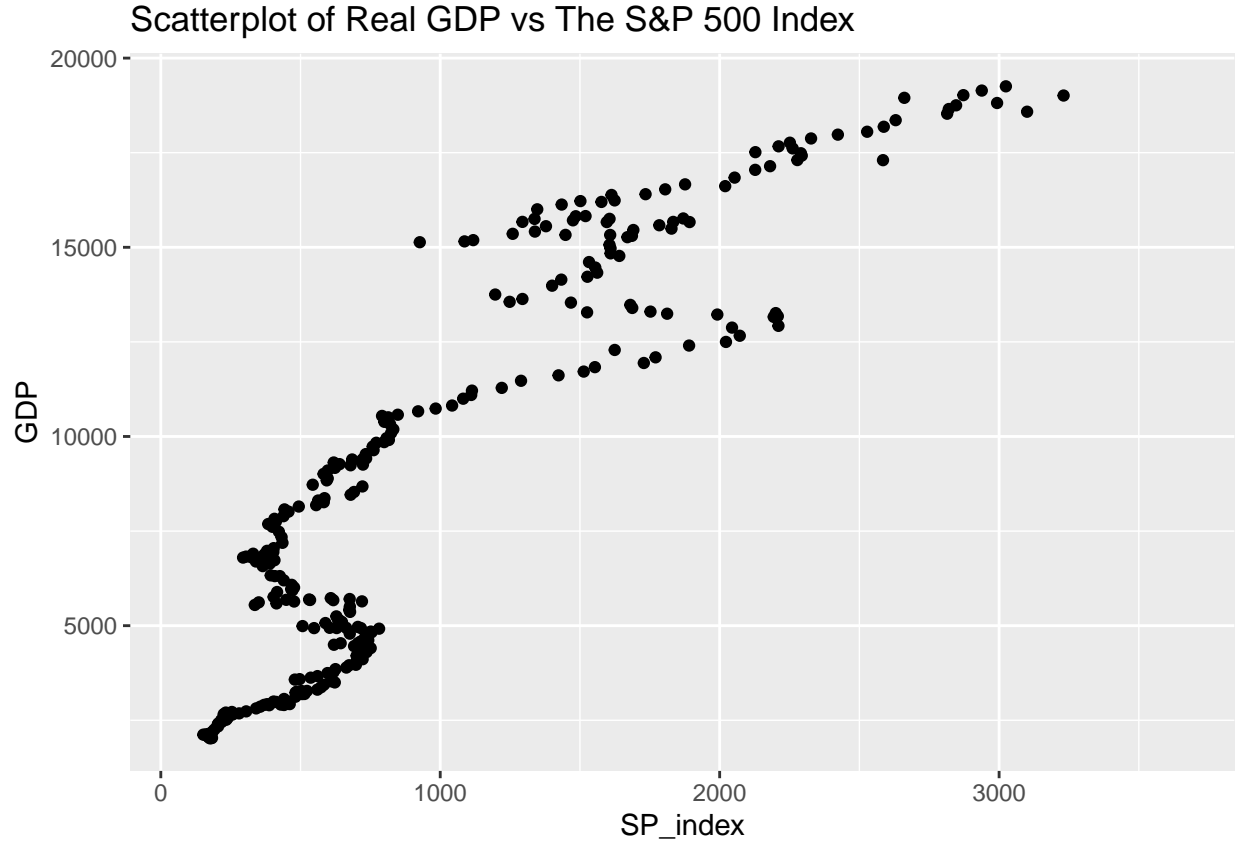


Scatterplot of Unemployment Rate vs The S&P 500 Index



Scatterplot of Weekly Median Income vs The S&P 500 Index





## Model

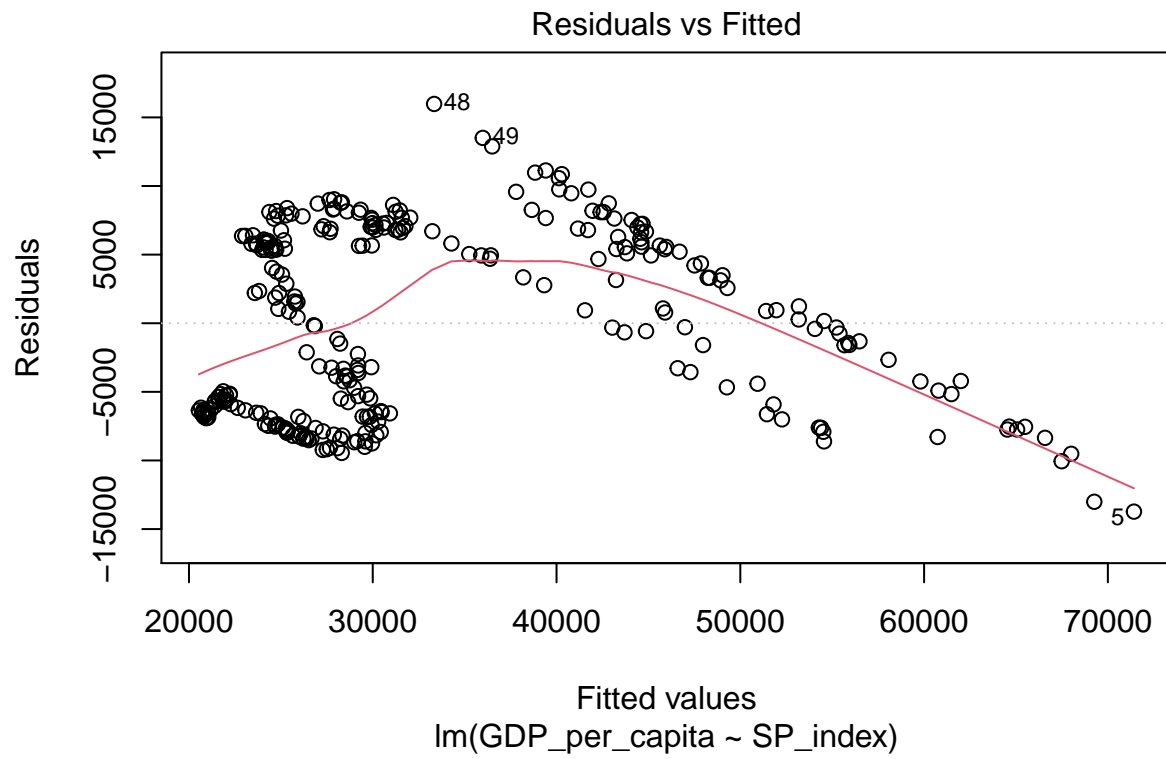
In total 4 linear regression models will be used to tested for the S&P index, one model for each response variable set as our economic indicators. Since there is only one explanatory variable and one response variable for each model, we will be using simple linear regression models. So for all our models, this equation will be used to predict the economic indicator:

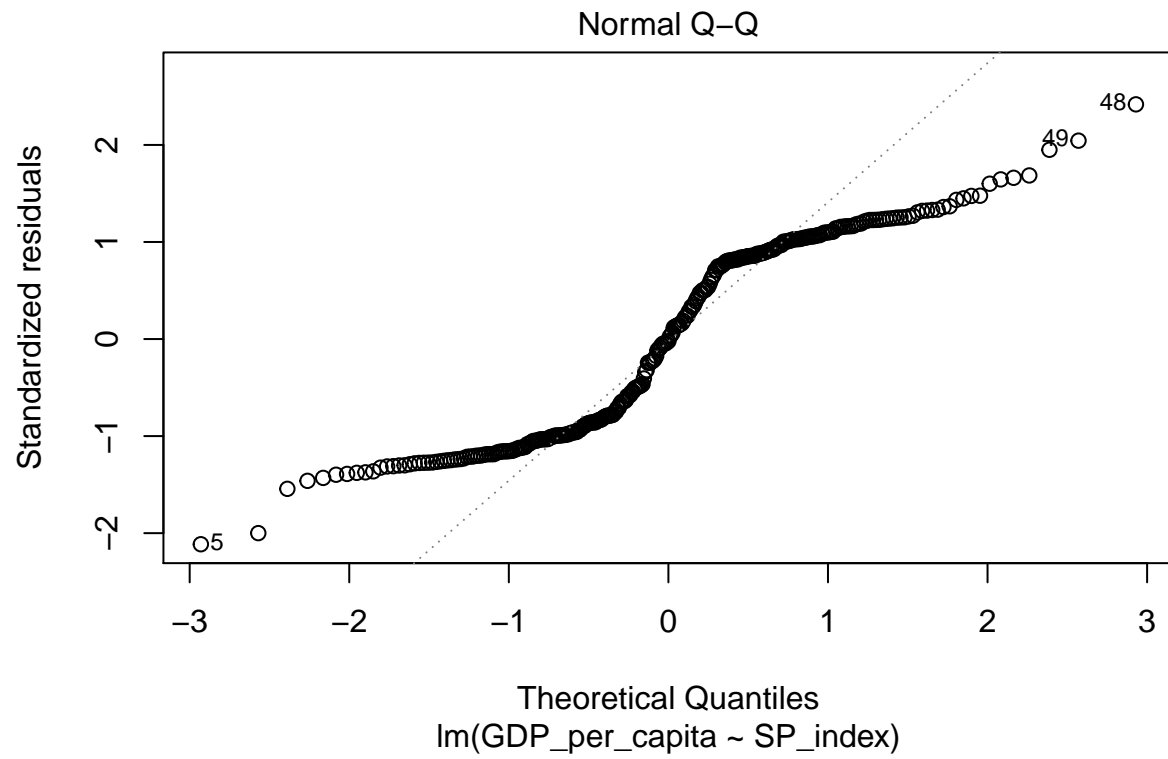
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where  $\hat{y}$  represents the predicted economic indicator amount (real GDP per capita, unemployment rate, weekly median income, or real GDP),  $\hat{\beta}_0$  is the intercept of our model,  $\hat{\beta}_1$  is the slope of regression model and the average change in  $\hat{y}$  if  $x$  increases by 1, and  $x$  represents the S&P 500 index.

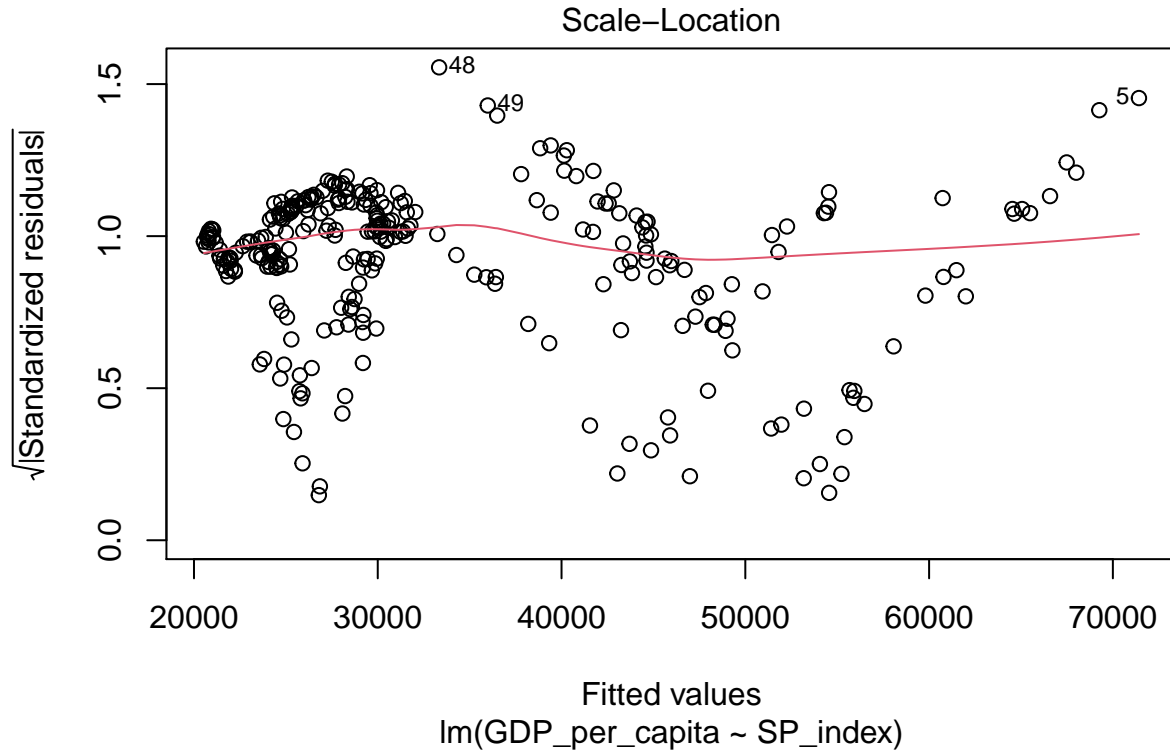
## Results

Our model shows that for predicting real GDP per capita using the S&P 500 index as a explanatory variable, we get an intercept of 1.802e+04 and a slope of 1.653e+01. This means that if the S&P index rises by \$1, then the real GDP per capita is expected to increase by 1.653e+01 US dollars. If the S&P 500 index is \$0 (although this is not within the domain of possibility), then real GDP per capita is expected to be 1.802e+04 US dollars. Both the intercept and the slope of the model have a p-value of 0, which makes this model statistically significant. However, when analyzing the diagnostic plots, we can see several anomalies. The residuals vs fitted graph has a parabolic trend, suggesting that we may be missing another predictor variable. Additionally, it appears that the response variable is not normally distributed, which suggests that this model breaks the assumptions of the SLR model. The constant variance assumption is not broken, as the scale-location graph appears to show no trend.

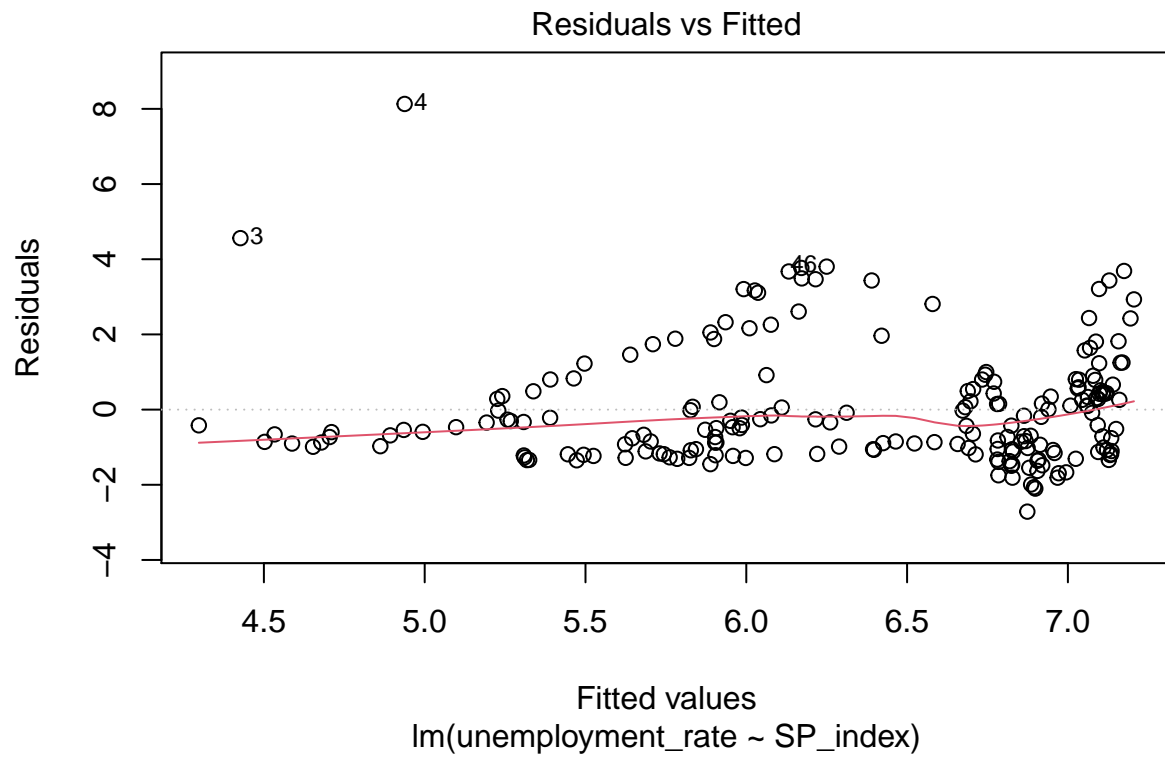


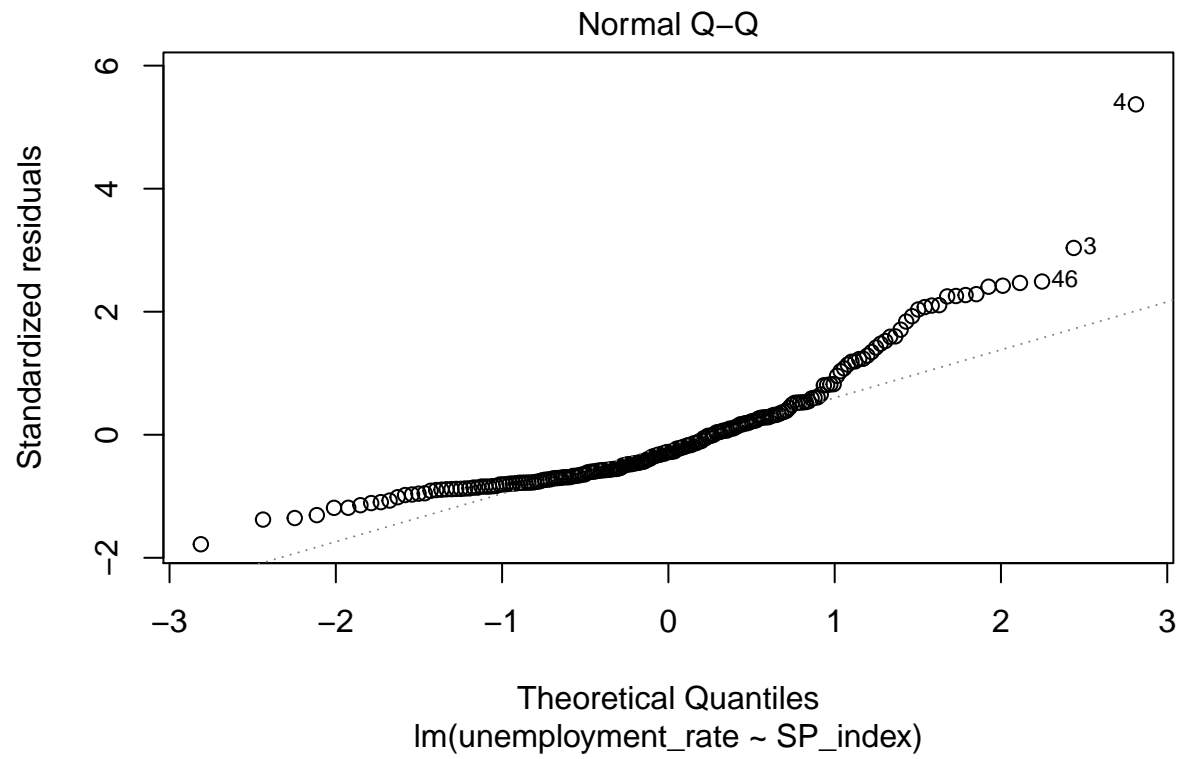


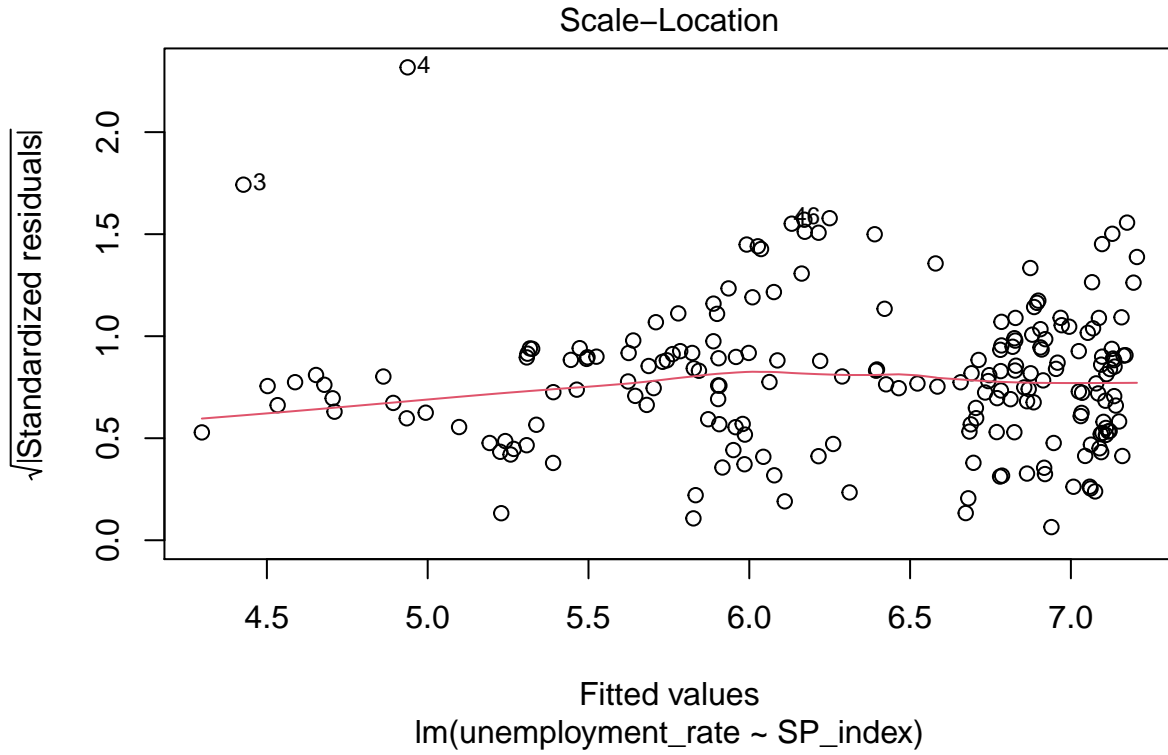




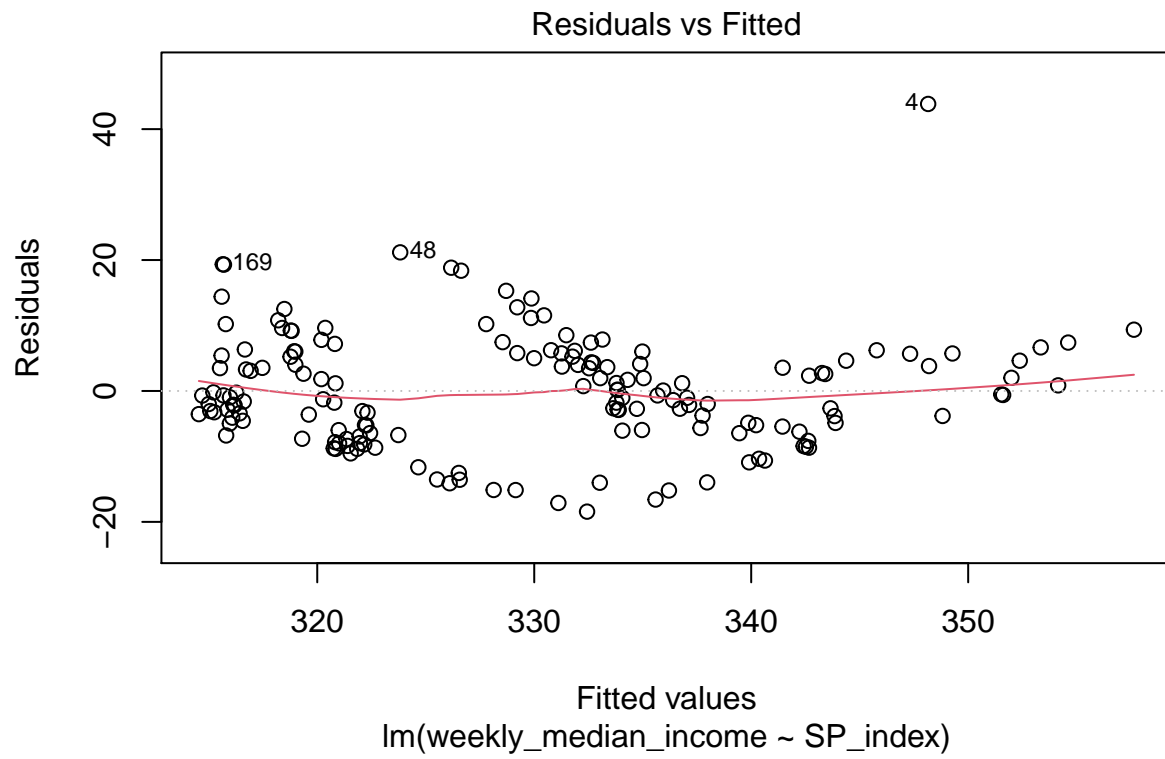
Our second model using the S&P 500 index to predict the unemployment rate in the US has an intercept of 7.4969851 and a slope of -0.0009902. This means if S&P 500 index increases by \$1, then the unemployment rate is expected to decrease by 0.0009902%. If S&P 500 index is \$0, then the unemployment rate is expected to be 7.50%. The p-value for both the intercept and the slope are 0 and 6.64e-11. The three diagnostic plots appear to align with the assumptions of the SLR model, and therefore, we can show that there is a negative correlation between the S&P index and unemployment rates using this simple linear regression model. However, the effect of the change in the S&P 500 index is small, meaning that if the S&P 500 index were to climb \$100 (which is quite an increase in the context of the index price), then unemployment rates would be expected to only decrease by 0.099%.

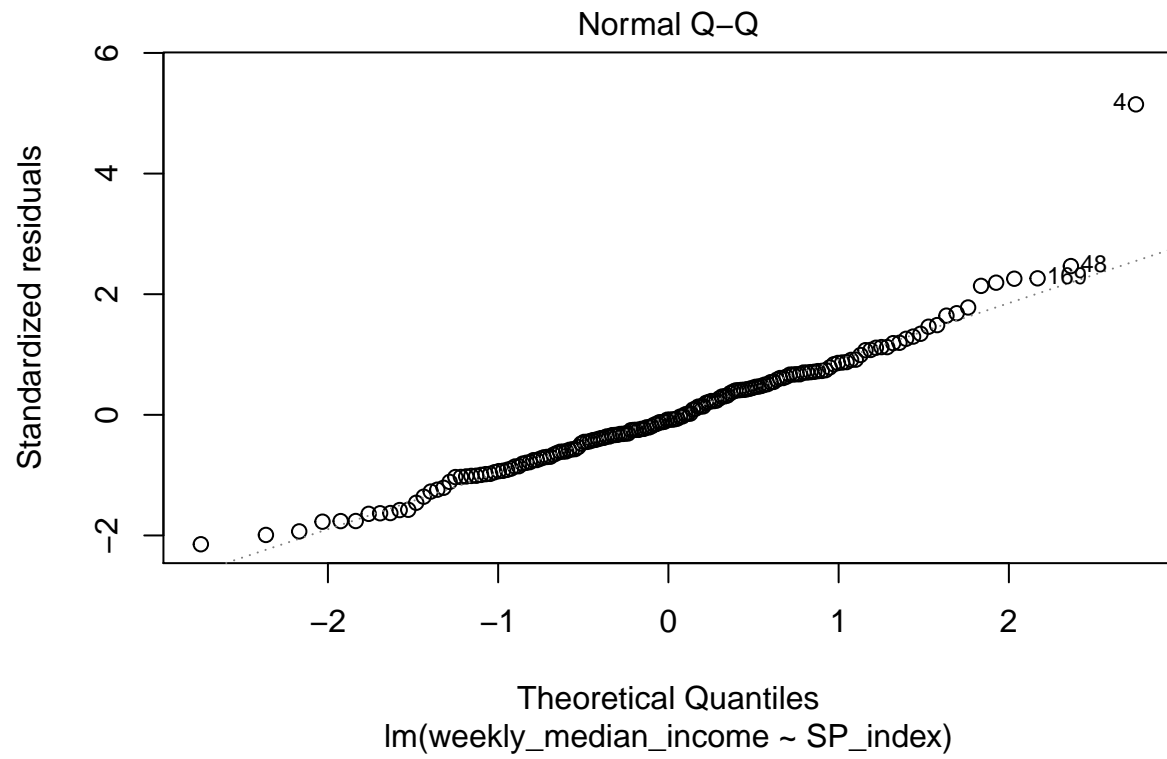


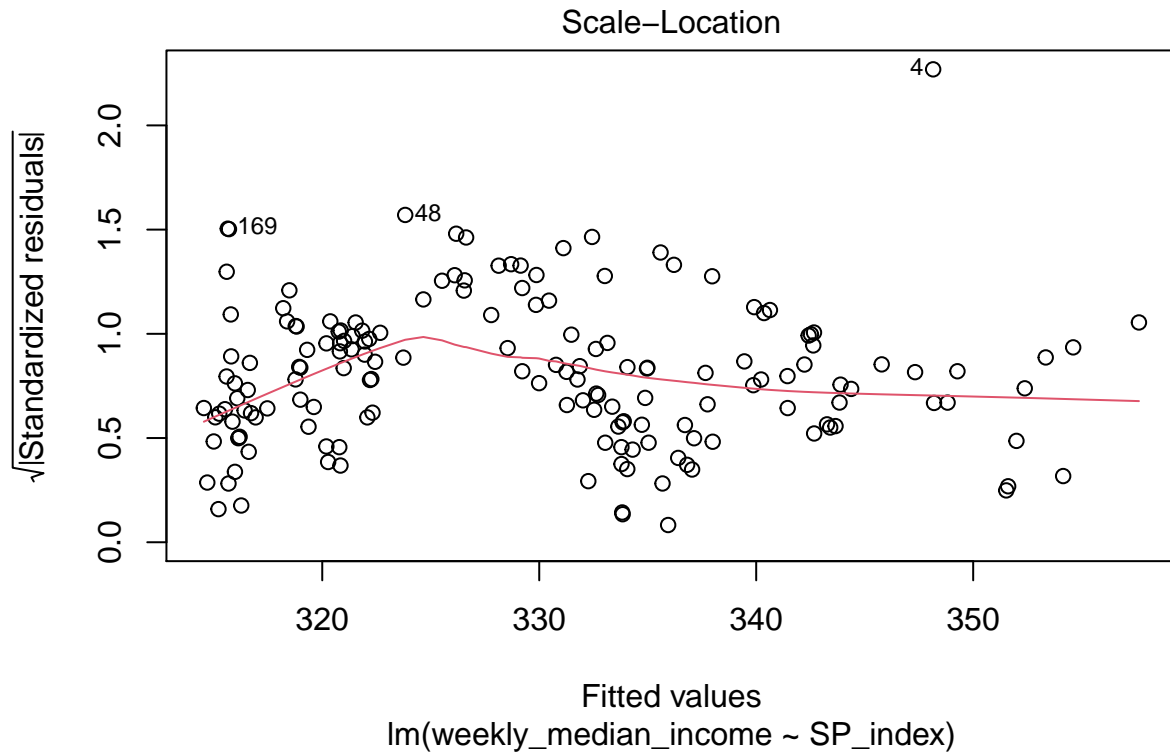




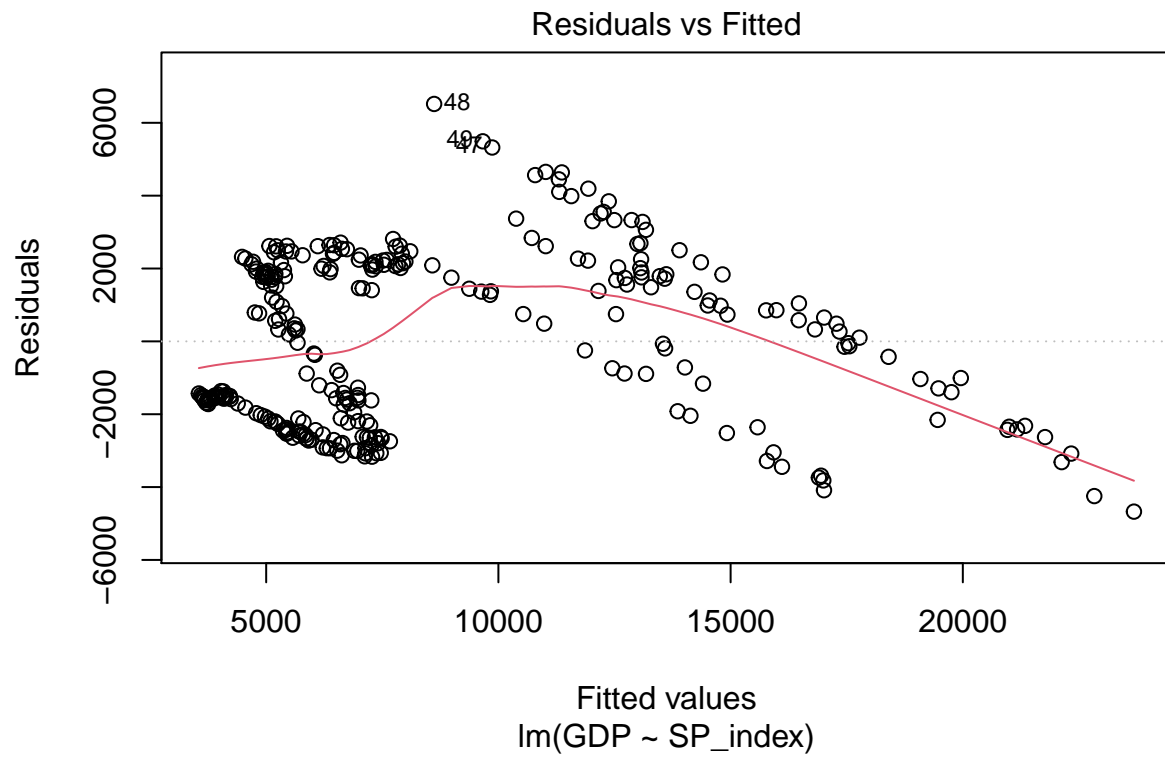
Our third model uses the s&p 500 index to predict the weekly median income of a US worker; the model has an intercept of  $3.102e+02$  and a slope of  $1.468e-02$ , both with p-values of 0 which makes them statistically significant. This means if the S&P 500 index were to increase by \$1, then the weekly median income would be expected to increase \$0.015 in the US. Again, similar to the unemployment rate, wages only appear to increase very little. Using diagnostic plots we can see that there is no trend in the residuals vs fitted graph, the residuals are normally distributed by the Normal QQ-plot, and the model has a constant variance as evident by the lack of a trend in the scale location graph. Therefore we determine that this simple linear model has all of its assumptions fulfilled and is a valid model.



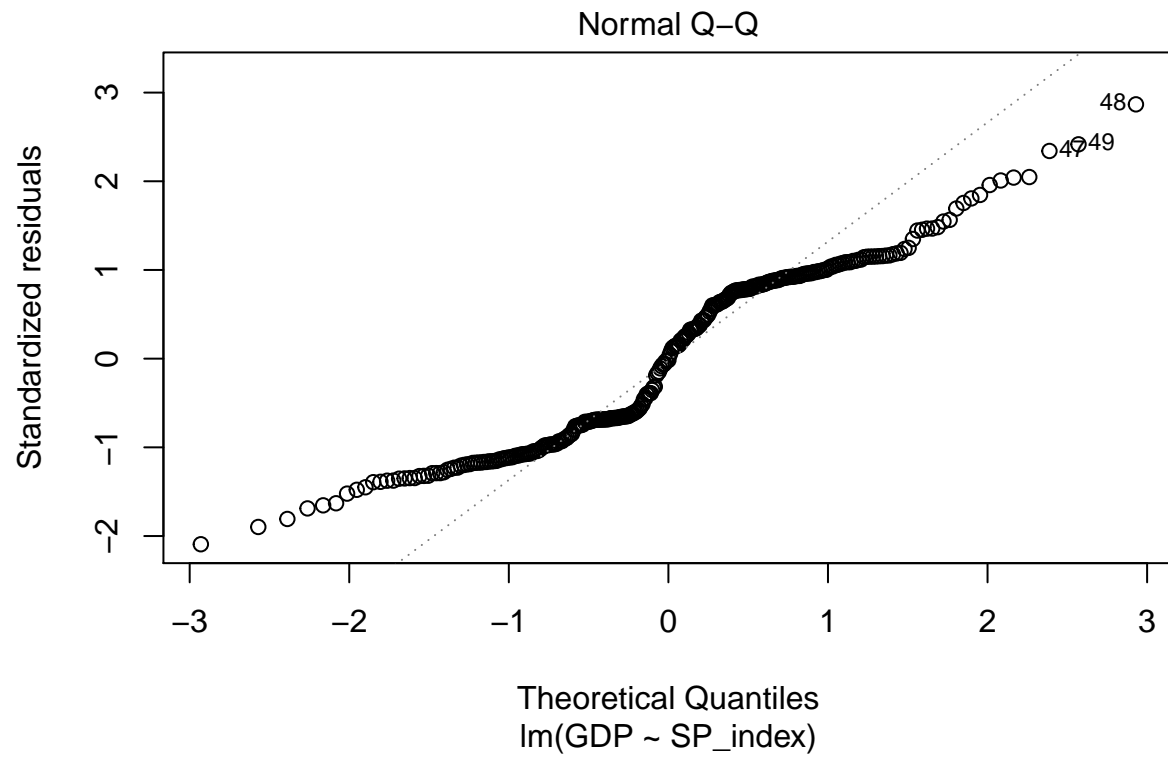


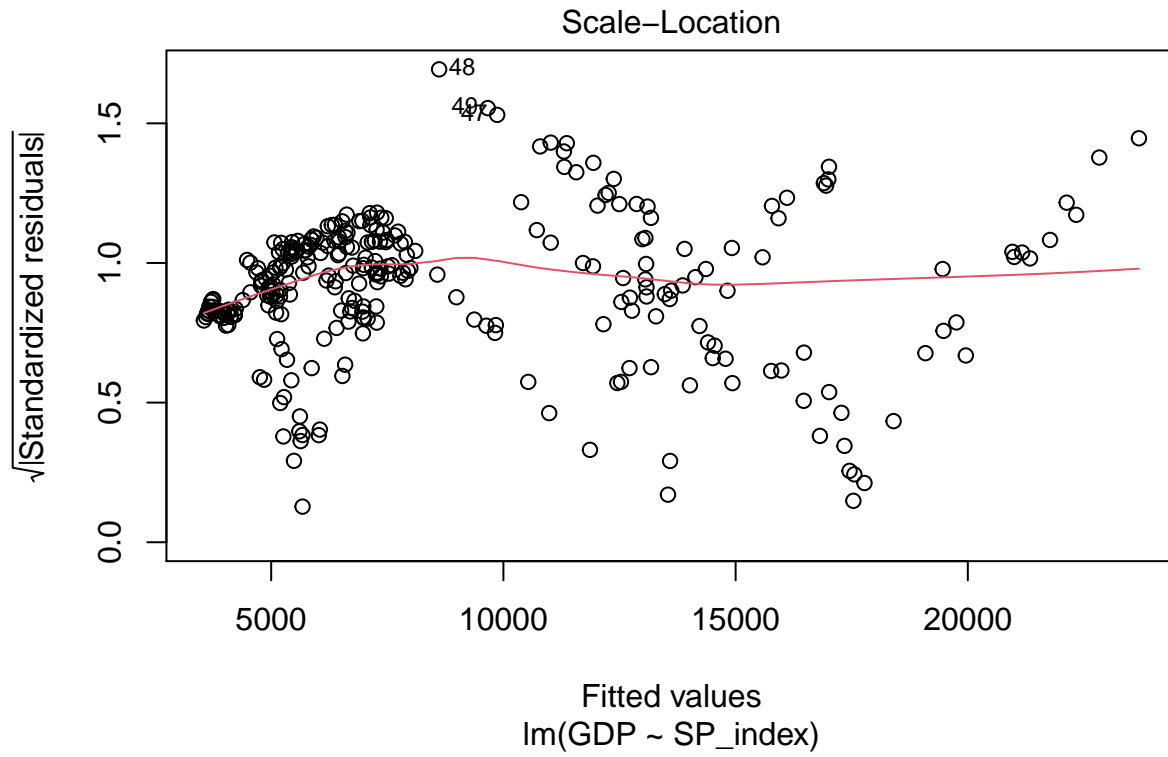


Our last model uses the S&P 500 index to predict US real GDP, and has an intercept of 2554.7449 and a slope of 6.5402. This means that if the S&P 500 index is \$0, then the real GDP would be \$2554.7449 (both of these values are outside the domains of the explanatory variable and response variable respectively). Additionally if the S&P 500 index were to increase by \$1, then real GDP is expected to increase by 6.5402 US dollars. Like our previous 3 models, the intercept and slope are statistically significant. Using the diagnostic plots, we can see the similar problem this model shares with the GDP per capita model. Both the residuals vs fitted model and the normal qq-plot model show that the SLR model assumptions are not met, and suggests that we cannot use the simple linear regression model and that we are likely missing another explanatory variable.









## Discussion

### Summary

Data for the historic quarterly S&P 500 index, real GDP per capita, unemployment rate, weekly median income, and real GDP were used to build our simple linear regression model. Using simple linear regression to predict the four economic indicators with the S&P 500 index as our predictor variable, we built four models showing their relationship.

### Conclusion

In conclusion, we've shown that S&P 500 index alone cannot be used to predict real GDP and real GDP per index. We can however, use the index to predict unemployment rate and the weekly median income. If the S&P increases by \$1, the expected unemployment rate decreases by 0.0009902% and the weekly median income is expected to increase 1.468e-02 US dollars. In other words, we showed that there is a negative relationship between the S&P 500 index and unemployment rate, and a positive relationship between the index and the weekly median income.

We can say that President Trump is not wrong in showing stock market index price increases as a way to show an increase in median weekly income and a decrease in unemployment rate. However, the extent to which that index price increases have on the two economy indicators are quite minimal. More importantly, our model shows that there is a correlation between index prices, and unemployment rates and weekly median income, but it does not necessarily show that there is a causation effect. It would therefore be wrong to assume that enacting policies that increase index prices will help improve the US economy.

## Weaknesses & Next Steps

There are many weaknesses in the models we have established, many of them were addressed in the data section regarding the flaws of using these response variables as an absolute indicator of the state of the economy. We have also addressed the fact that the index price may not be the only factor in determining the US economic performance; we could also be potentially missing out on another confounding variable that affects both our predictor and response variable, such as the interest rate set by the Federal Reserve. This idea that the stock price is not the sole factor in determining the economy is also echoed by Nobel prize winner economist Paul Krugman, who cites aggressive quantitative easing policies and low interest rates set by the Federal Reserve that made the stock market attractive for investors in contrast to other investment alternatives (Krugman, 2020). This caused the stock market to perform well despite the record unemployment that the US is currently experiencing (Krugman, 2020).

The next steps in our model could be extending the data to include other major countries to see if our findings could be replicated in other economies, or if the results we are seeing are merely applicable to the United States.

## References

- Amadeo, Kimberly. *Where's the Best Standard of Living? Depends Who You Ask*. 30 Aug. 2020, [www.thebalance.com/standard-of-living-3305758#:~:text=Real GDP per capita removes,more of its plentiful production.](http://www.thebalance.com/standard-of-living-3305758#:~:text=Real GDP per capita removes,more of its plentiful production.)
- Amadeo, Kimberly. *Why Every Jobless Person Is Not Counted as Unemployed*. 21 Nov. 2020, [www.thebalance.com/unemployment-rate-3305744#:~:text=The unemployment rate is an,sectors are losing jobs faster.](http://www.thebalance.com/unemployment-rate-3305744#:~:text=The unemployment rate is an,sectors are losing jobs faster.)
- Employed Full Time: Median Usual Weekly Real Earnings: Wage and Salary Workers: 16 Years and Over*. 17 July 2020, [fred.stlouisfed.org/series/LES1252881600Q](http://fred.stlouisfed.org/series/LES1252881600Q).
- Global Shares Plunge in Worst Day Since Financial Crisis*. 9 Mar. 2020, [www.bbc.com/news/business-51796806](http://www.bbc.com/news/business-51796806).
- Hall, Mary. *How Does GDP Affect the Standard of Living?* 23 Oct. 2020, [www.investopedia.com/ask/answers/060115/how-does-gross-domestic-product-gdp-affect-standard-living.asp](http://www.investopedia.com/ask/answers/060115/how-does-gross-domestic-product-gdp-affect-standard-living.asp).
- Income and Wages as an Economic Indicator: Friedberg Direct*. 1 Dec. 2020, [www.avatrade.ca/education/trading-for-beginners/income-and-wages#:~:text=Income and Wages Reports \(IWRs,entities in each time period.&text=Wage is a form of,exchange for labour or service.](http://www.avatrade.ca/education/trading-for-beginners/income-and-wages#:~:text=Income and Wages Reports (IWRs,entities in each time period.&text=Wage is a form of,exchange for labour or service.)
- It's Been 4 Days Since Trump Last Tweeted About the Markets and the Economy*. 17 Dec. 2020, [www.bloomberg.com/features/trump-tweets-market/](http://www.bloomberg.com/features/trump-tweets-market/).
- Krugman, Paul. *Crashing Economy, Rising Stocks: What's Going On?* 30 Apr. 2020, [www.nytimes.com/2020/04/30/opinion/economy-stock-market-coronavirus.html](http://www.nytimes.com/2020/04/30/opinion/economy-stock-market-coronavirus.html).

Mashayekhi, Rey. *Worst Day In a Decade: Nasdaq, S&P, Dow Down Nearly 8% In Massive Market Rout*. 10 Mar. 2020, [fortune.com/2020/03/09/dow-jones-sp-500-today-nasdaq-stock-market-crash-2020/](https://fortune.com/2020/03/09/dow-jones-sp-500-today-nasdaq-stock-market-crash-2020/).

Picardo, Elvis. *How the Unemployment Rate Affects Everybody*. 4 Dec. 2020, [www.investopedia.com/articles/economics/10/unemployment-rate-get-real.asp](http://www.investopedia.com/articles/economics/10/unemployment-rate-get-real.asp).

*Real Gross Domestic Product*. 22 Dec. 2020, [fred.stlouisfed.org/series/gdpc1](https://fred.stlouisfed.org/series/gdpc1).

*Real Gross Domestic Product Per Capita*. 22 Dec. 2020, [fred.stlouisfed.org/series/A939RX0Q048SBEA](https://fred.stlouisfed.org/series/A939RX0Q048SBEA).

“S&P 500 Inflation Adjusted by Month.” *Quandl.com*, 6 Dec. 2020, [www.quandl.com/data/MULTPL/SP500\\_INFLADJ\\_MONTH-S-P-500-Inflation-Adjusted-by-Month](https://www.quandl.com/data/MULTPL/SP500_INFLADJ_MONTH-S-P-500-Inflation-Adjusted-by-Month).

Schneider, Avie. *The Big Comeback: S&P 500 Closes At Record High 6 Months After Coronavirus Plunge*. 18 Aug. 2020, [www.npr.org/sections/coronavirus-live-updates/2020/08/18/903252878/the-big-comeback-s-p-500-closes-at-record-high-6-months-after-coronavirus-plunge](https://www.npr.org/sections/coronavirus-live-updates/2020/08/18/903252878/the-big-comeback-s-p-500-closes-at-record-high-6-months-after-coronavirus-plunge).

*Unemployment Rate: Aged 15-64: All Persons for the United States*. 10 Dec. 2020, [fred.stlouisfed.org/series/LRUN64TTUSQ156S](https://fred.stlouisfed.org/series/LRUN64TTUSQ156S).

*Unemployment Rates Down In 41 States, July 2020 to August 2020*. 24 Sept. 2020, [www.bls.gov/opub/ted/2020/unemployment-rates-down-in-41-states-july-2020-to-august-2020.htm#:~:text=The national unemployment rate fell,higher than in August 2019](https://www.bls.gov/opub/ted/2020/unemployment-rates-down-in-41-states-july-2020-to-august-2020.htm#:~:text=The national unemployment rate fell,higher than in August 2019).