Abstract

In this project, the main research question is to find the major factors contributing to personal saving amount based on United States Macroeconomic data collected from Federal Reserve Economic Data (FRED). Initially, I did some exploratory analysis by plotting the distribution of each variable and created scatter plots to find the underlying correlation between the response variable (personal saving) and the independent variables. Then, tried several multiple regression models with different bundles of predictors yet there is an obvious multicollinearity issue based on VIF test results. After some data transformation, I rebuild the models and select the best model with an adjusted R-squared of 0.8259. Based on the residual analysis, I can tell that the issues came across were well resolved and the residual distribution looks good.

Background/Introduction

The aim is to build a parsimonious model that will help interpret different factors that contribute to personal savings. I wanted to analyze how the factors like Consumer price index which measures the average change in prices paid by customers for goods and services, Mortgage rates, Unemployment rate, NASDAQ which plays an important role in local and global economy (measuring the performance of more than 3000 securities that are all listed on the Nasdaq stock market), Disposable Income representing the after-tax income of persons and nonprofit corporations, and another measure called Personal consumption expenditure that tracks the changes in the prices of consumer goods over time affect the variable Personal Savings. In addition to the above measures, I have included recession as well to check how it impacts the personal savings of individuals.

Data Set:

The macroeconomics dataset (Mirashi, 2022) which was downloaded from Kaggle on 20 October 2022, is from the United States Macroeconomic data that was collected from Federal Reserve Economic Data (FRED). The dataset has CPI, Mortgage rate, unemployment rate, NASDAQ, disposable income, personal consumption expenditure, and personal savings as attributes. Additionally, I have created a categorical column named Recession using the Investopedia.com (Team, 2022) reference. Personal savings is the dependent variable. The objective is to build a parsimonious model that will help to interpret factors contributing to personal savings.

Analysis Plan:

I used histograms to identify the skewness of data. Using the scatter plots, I could understand the linear relationship between the variables. Then proceeded with linear regression and observed the significance of independent variables towards the dependent variable. Also, I used the correlation matrix to check the correlation of variables. The VIF tests provided me with information on collinearity between the independent variables of regression models that were ran. To assess the quality of the regression, I utilized residual plots.

Since there is skewness in the original data, I transformed the variables by applying log transformations and carried out a similar process with linear regression analysis. Applying the transformation reduced the skewness to some extent but not fully, resulting in a model with better adjusted R2 value. Then, I have eliminated fields one by one until I came up with our parsimonious model.

Results

The final model is created based on the transformed data fields to resolve the issue of multicollinearity, and it has an adjusted R-squared of 0.8259. The final model is a linear regression with the variables log(personal saving)~ log(Mortgage rate) + log(unemployment rate) + log(NASDAQ) + log(disposable income) + log(CPI) + log(personal consumption expenditure) + recession, and all terms in this model are statistically significant. According to the residual plots, the residuals are randomly distributed centered around 0 and the QQ norm plot is roughly a straight line which indicate the validness of this model.

Limitations

The United States is the third largest country (in population) in the world (Most populous countries in the world, 2022, n.d.), the fourth largest in land area (largest countries in the world [by area], n.d.), and one of the most diverse in the world. This means that there are more macro-economic factors than are listed in the dataset. One should not take the attributes in the dataset as everything that can influence personal savings. In fact, more studies should be completed to get a wider variety of macroeconomic factors that can influence savings.

In addition to macroeconomic factors, there are a wide range of microeconomic factors. Americans save/spend their allotted monies for assorted reasons. Therefore, the personal savings category is very nuanced. The attributes from this experiment cannot be used as the only items attributing to personal savings. One could argue that microeconomic factors are more important than macroeconomic factors. Therefore, one should take this experiment and incorporate other macroeconomic factors and microeconomic factors to get a fuller and richer understanding of personal savings.

One last limitation to this experiment is the size of the dataset. The number of attributes and instances does not allow us to make a complete analysis. There are only 42 years' worth of data. It is possible that there is a time element that can stretch into decades that would not be captured.

Conclusion

Based on all the exploratory analysis and modeling, I can tell that personal saving amount is related to CPI, Mortgage rate, unemployment rate, NASDAQ, disposable income, personal consumption expenditure, and recession. Though all these independent variables are statistically significant in the final model, CPI, personal consumption expenditure, and disposable income are the ones with relatively large coefficients, thus affecting the personal saving more with one unit change.

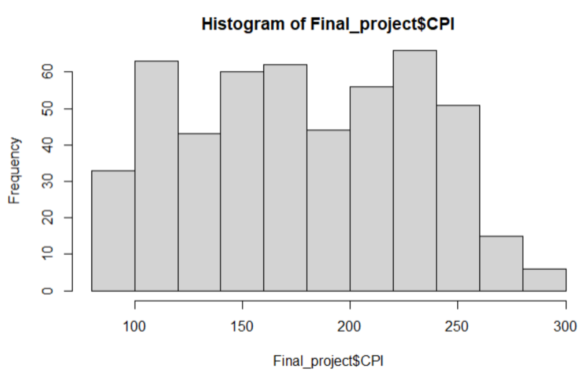
Descriptive Statistics

Summary of the dataset:

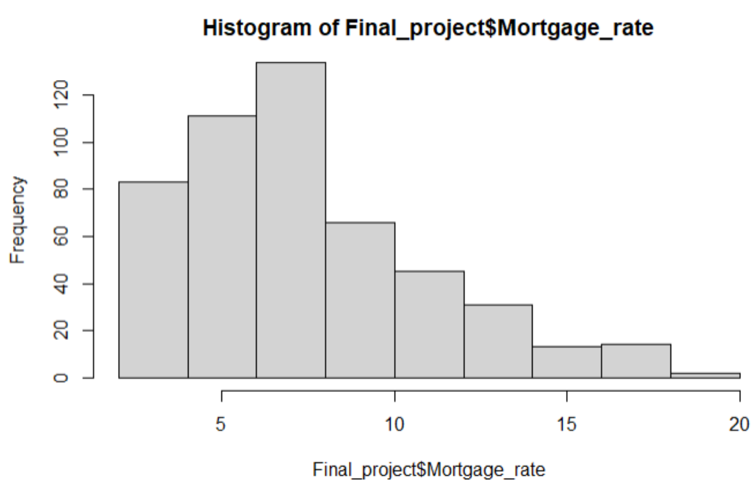
```
     date                CPI          Mortgage_rate      Unemp_rate         NASDAQ          disposable_income Personal_consumption_expenditure
 Length:499          Min.   : 85.6   Min.   : 2.684   Min.   : 3.500   Min.   :  167.4   Min.   : 4965     Min.   : 1827
 Class :character    1st Qu.:134.9   1st Qu.: 4.519   1st Qu.: 4.900   1st Qu.:  479.5   1st Qu.: 6861     1st Qu.: 3910
 Mode  :character    Median :177.4   Median : 6.834   Median : 5.700   Median : 1884.7   Median : 9734     Median : 7082
                     Mean   :178.8   Mean   : 7.421   Mean   : 6.181   Mean   : 2726.4   Mean   : 9802     Mean   : 7626
                     3rd Qu.:227.5   3rd Qu.: 9.338   3rd Qu.: 7.250   3rd Qu.: 3131.4   3rd Qu.:12226     3rd Qu.:10857
                     Max.   :291.5   Max.   :18.454   Max.   :14.700   Max.   :15814.9   Max.   :19120     Max.   :16955
 Recession_categ     personal_savings
 Length:499          Min.   : 2.100
 Class :character    1st Qu.: 5.850
 Mode  :character    Median : 7.200
                     Mean   : 7.487
                     3rd Qu.: 8.600
                     Max.   :33.800
```

Histograms

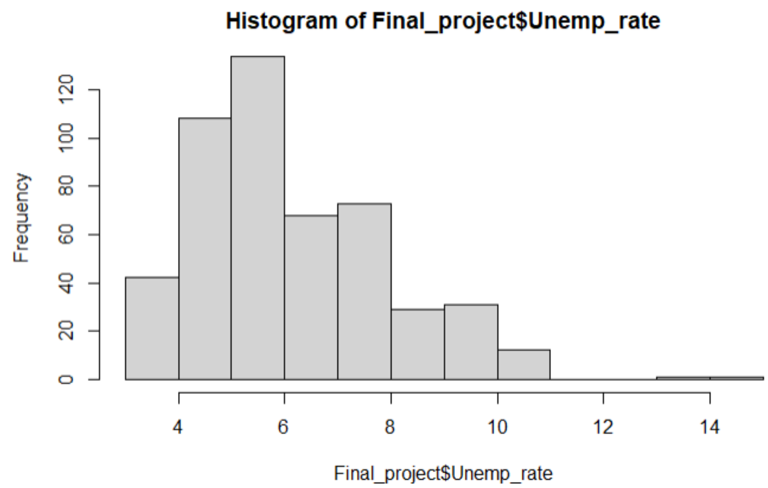
CPI



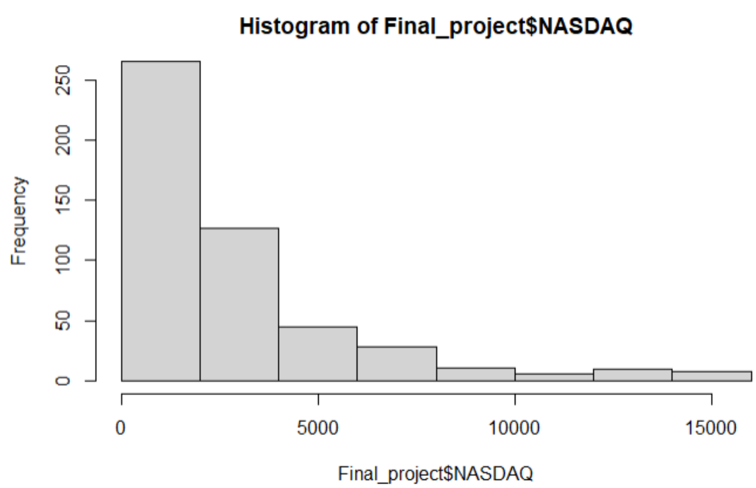
Mortgage Rate:



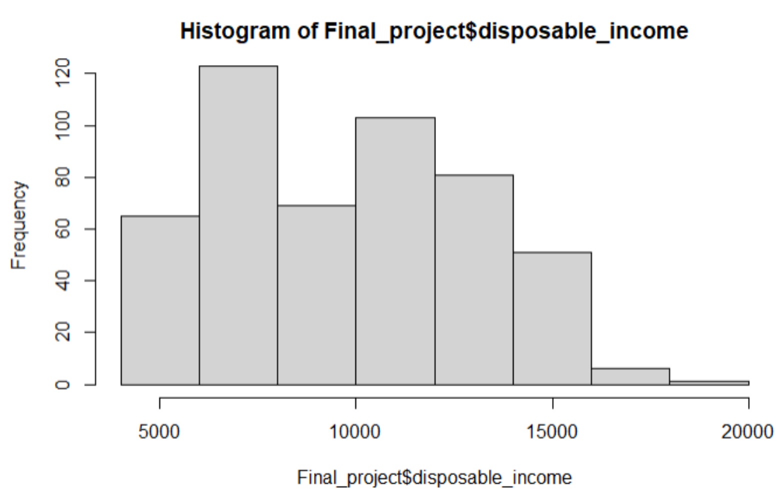
Unemployment Rate:



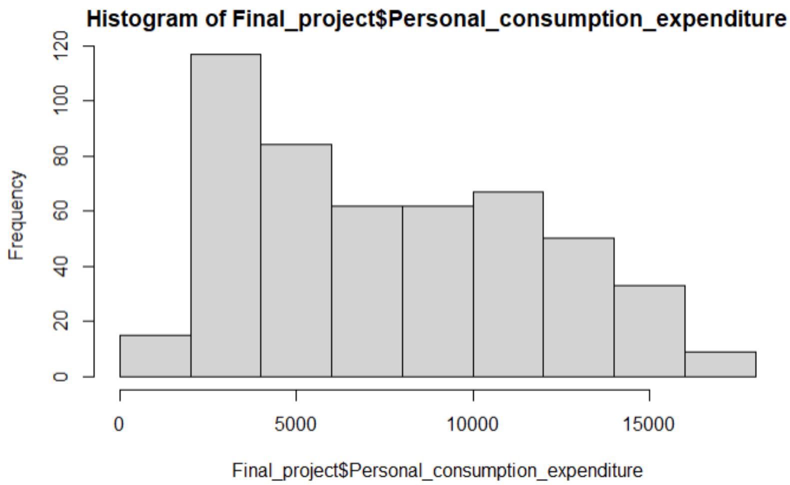
NASDAQ :



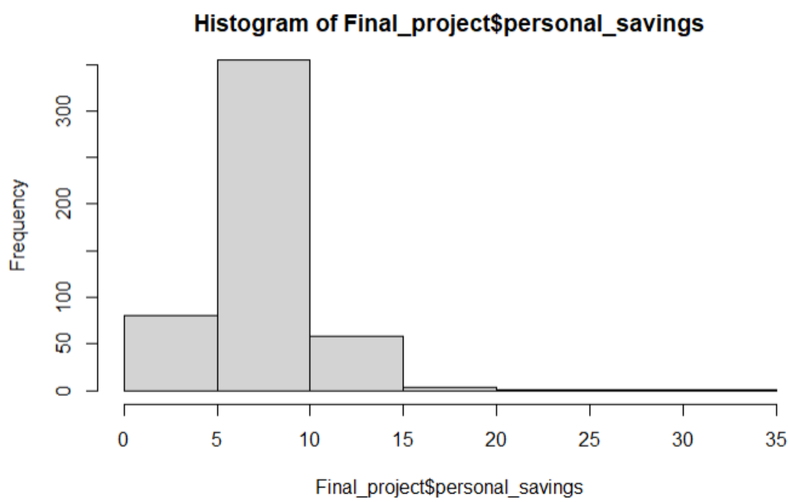
Disposable Income:



Personal Consumption Expenditure:



Personal Savings:

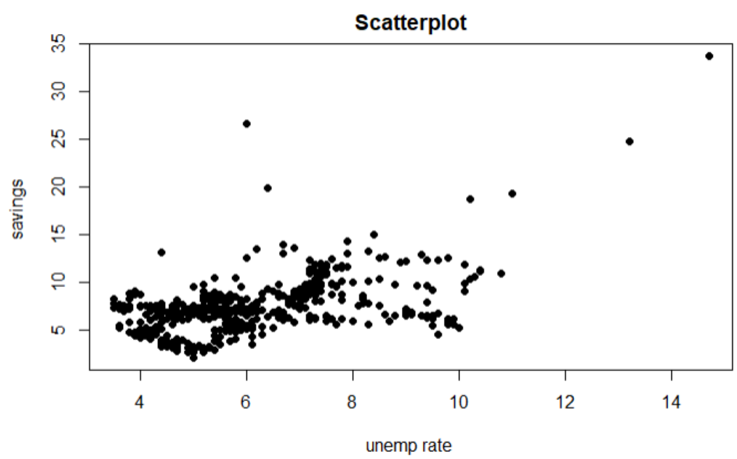


The histograms are generally right skewed, which has led me believe that conducting a linear transformation would lower the skewness of my data.

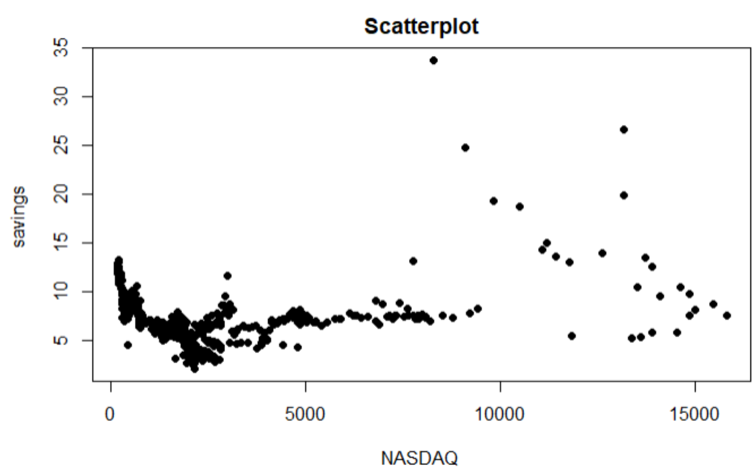
- **CPI**: The CPI histogram is somewhat evenly distributed.
- **Mortgage Rate**: The Mortgage Rate histogram is heavily right-skewed and would benefit from a linear transformation.
- **Unemployment Rate**: The histogram of Unemployment Rate is right-skewed and is a suitable candidate for a linear transformation.
- **NASDAQ**: The NASDAQ histogram is heavily right-skewed and is a suitable candidate for a linear transformation.
- **Disposable Income**: Disposable Income has a slight right-skew but is more or less evenly distributed. Disposable Income will likely benefit from a linear transformation.

- **Personal Consumption Expenditure**: The histogram of Personal Consumption Expenditure is somewhat evenly distributed as there is not one dominant side representing left or right skewness.
- **Personal Savings**: The histogram of Personal Savings is right skewed as the distribution of data is heavily skewed to the Y-axis. Personal Savings represents a good candidate for linear transformations.

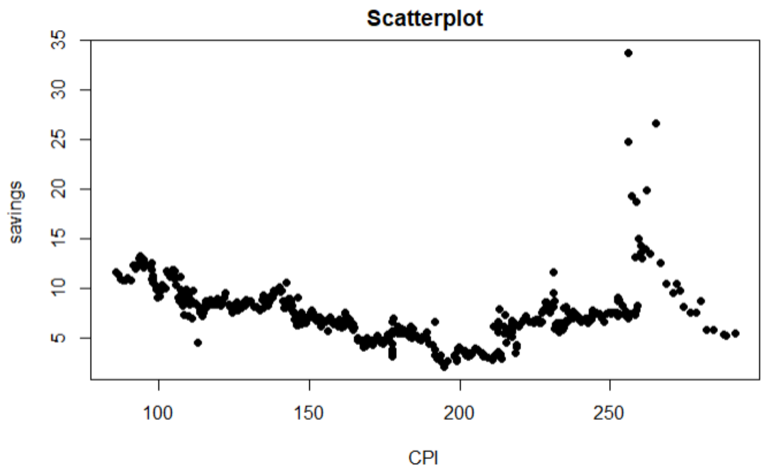
Scatter Plots:



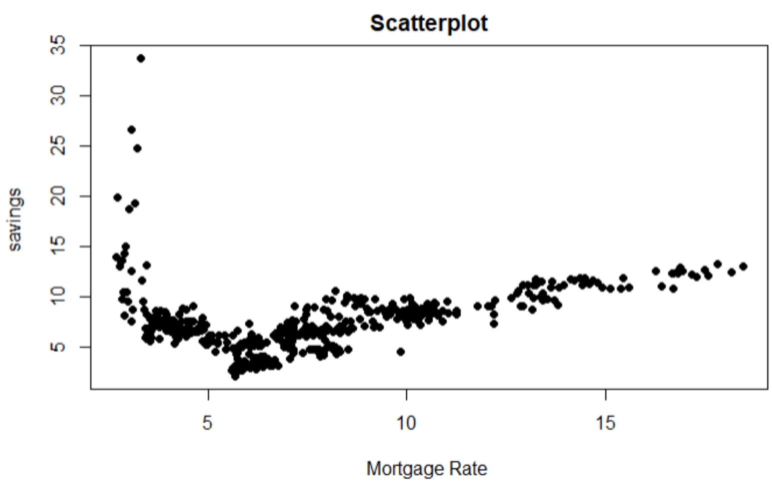
The Scatterplot for Unemployment Rate seems like there could be a small linear relationship between about 4.5 and 7.5 then levels out until 10.5. Before 4.5, then goes in a different direction and after 10.5 there are very few points to make a good confidence level.



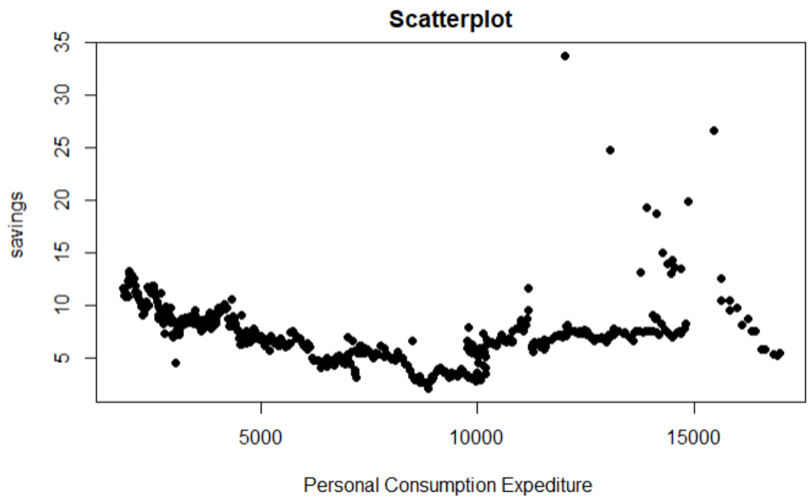
The Scatterplot for NASDAQ shows a sharp negative relationship from 0 to about 2300, then has a positive linear relationship from 2300 to about 7500. After 7500, the plots are more scattered, and the confidence levels go down.



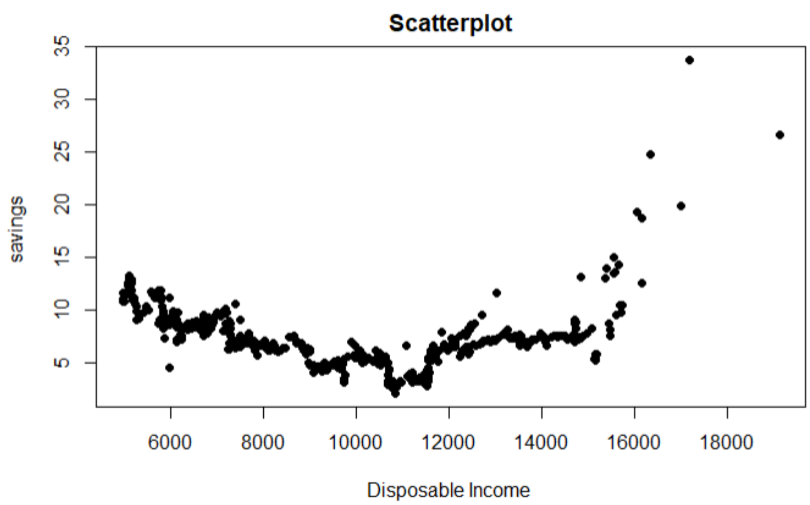
The Scatterplot for CPI has a negative linear relationship from about 50 to about 185. Next, we see a positive relationship from 185 to about 250. After 250, the plots become more erratically scattered and confidence goes down.



The Scatterplot for mortgage rate shows a sharp decline from 0 to 5. After 5, there seems to be a linear relationship for the rest of the rates. However, after 15 the plots become fewer, and the confidence levels drop.



The Scatterplot for personal consumption expenditure looks like a flattened U with the point of inflection at around 8000. After about 13000, the confidence level drops.



The Scatterplot for disposable income looks like a U but has a negative linear relationship from 0 to about 10,000 then a positive linear relation until about 14,000. After 14,000 the plots become more scattered and less reliable.

Regression Output

````
```{r linear model}
fit1<-lm(personal_savings ~ CPI+Mortgage_rate+Unemp_rate+NASDAQ+disposable_income+Personal_consumption_expenditure+Recession_categ,
summary(fit1)
```


Call:
lm(formula = personal_savings ~ CPI + Mortgage_rate + Unemp_rate +
    NASDAQ + disposable_income + Personal_consumption_expenditure +
    Recession_categ, data = Final_project)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0436 -1.3570  0.4379  1.1923  6.5531

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -3.197e+00  2.372e+00  -1.348  0.17832
CPI                              -9.424e-02  1.809e-02  -5.208 2.81e-07 ***
Mortgage_rate                     2.204e-01  7.812e-02   2.822  0.00497 **
Unemp_rate                        8.702e-01  4.196e-02  20.737  < 2e-16 ***
NASDAQ                            8.426e-04  5.154e-05  16.348  < 2e-16 ***
disposable_income                 3.032e-03  1.901e-04  15.952  < 2e-16 ***
Personal_consumption_expenditure -1.511e-03  2.338e-04  -6.462 2.50e-10 ***
Recession_categ                   1.661e-01  2.510e-01   0.662  0.50841
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.54 on 491 degrees of freedom
Multiple R-squared:  0.7428,    Adjusted R-squared:  0.7391
F-statistic: 202.6 on 7 and 491 DF,  p-value: < 2.2e-16
````

Model #1 is the second-best performing model and includes all variables and has an Adjusted R^2 of 0.7391. The Recession variable is not statistically significant at the 0.05 level and will be removed for further analysis. All other variables not including Recission are statistically significant at the 0.05 level. The model in its entirety is statistically significant at the 0.05 level.

````
```{r linear model2}
fit2<-lm(personal_savings ~ CPI+Mortgage_rate+Unemp_rate+NASDAQ+disposable_income+Personal_consumption_expenditure,
summary(fit2)
```


Call:
lm(formula = personal_savings ~ CPI + Mortgage_rate + Unemp_rate +
    NASDAQ + disposable_income + Personal_consumption_expenditure,
    data = Final_project)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0762 -1.3738  0.4398  1.1833  6.5925

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -3.850e+00  2.157e+00  -1.785 0.074844 .
CPI                              -9.067e-02  1.726e-02  -5.253 2.23e-07 ***
Mortgage_rate                     2.448e-01  6.883e-02   3.557 0.000411 ***
Unemp_rate                        8.709e-01  4.193e-02  20.772  < 2e-16 ***
NASDAQ                            8.391e-04  5.125e-05  16.374  < 2e-16 ***
disposable_income                 3.048e-03  1.884e-04  16.179  < 2e-16 ***
Personal_consumption_expenditure -1.550e-03  2.259e-04  -6.861 2.06e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

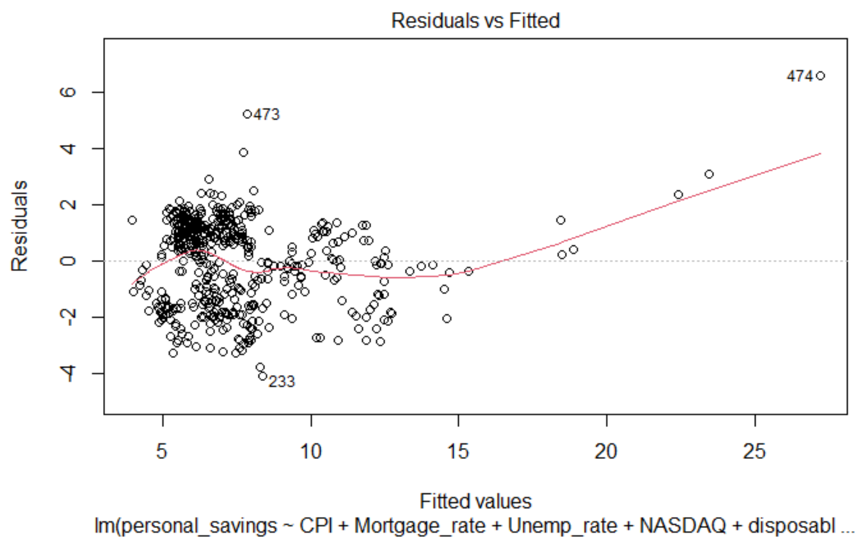
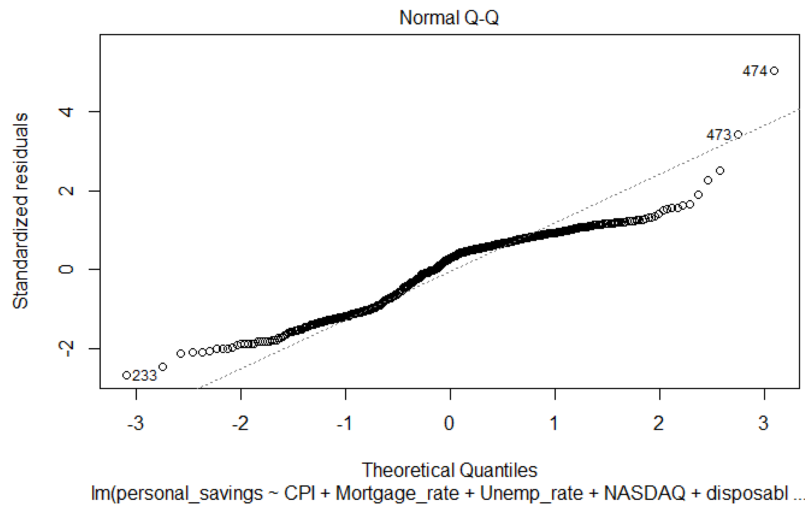
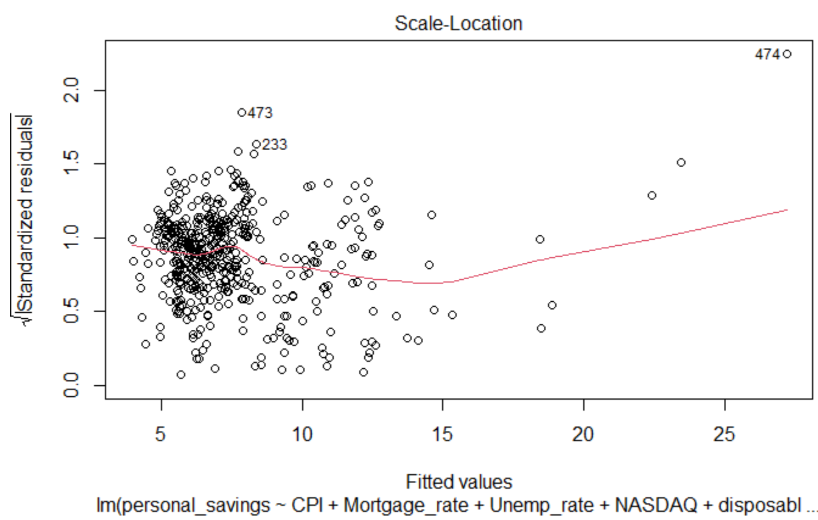
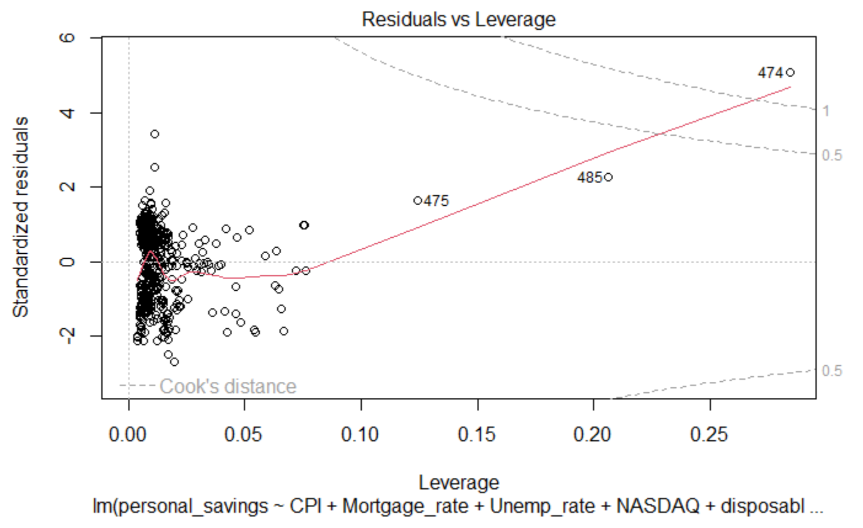
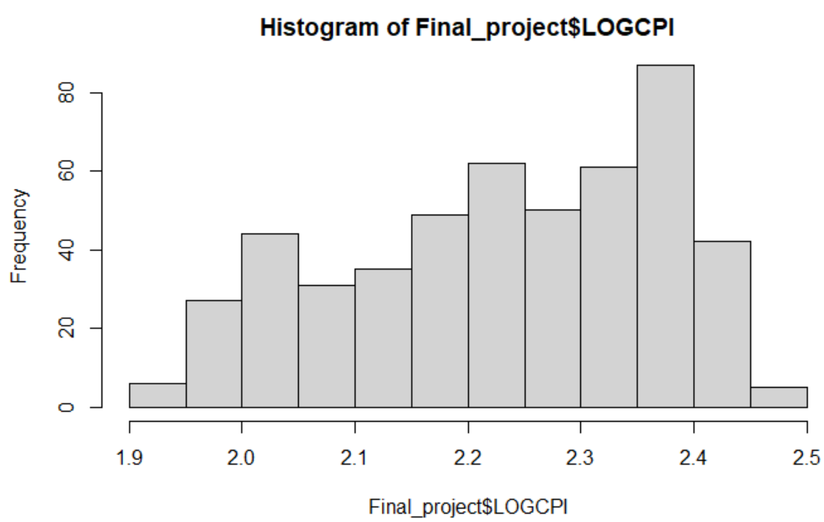
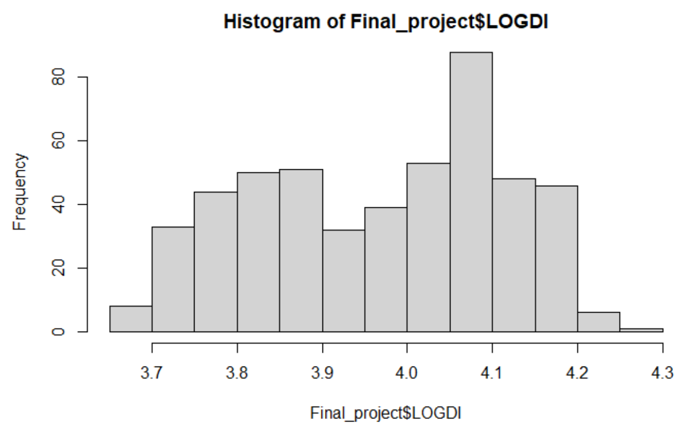
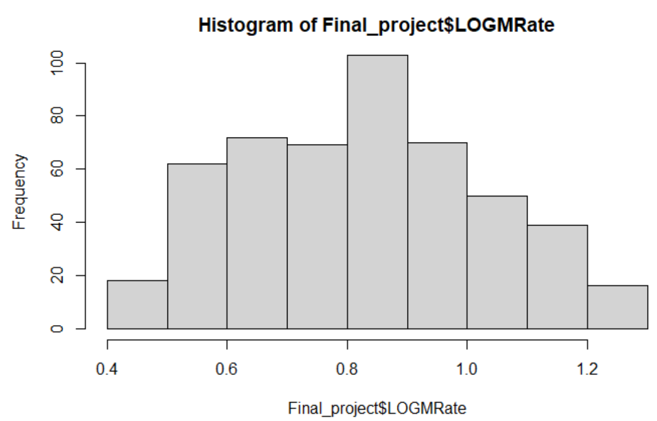
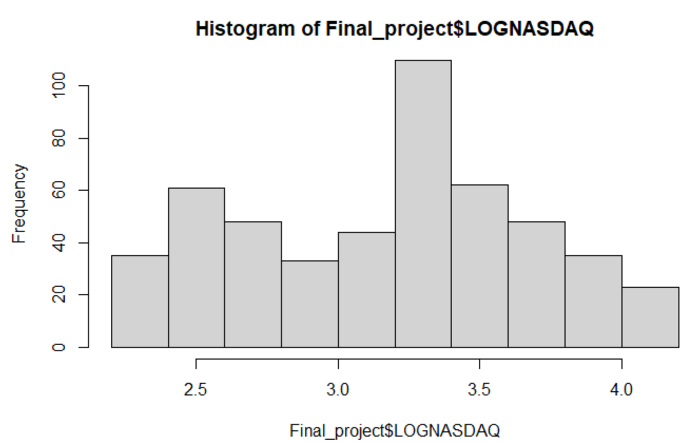
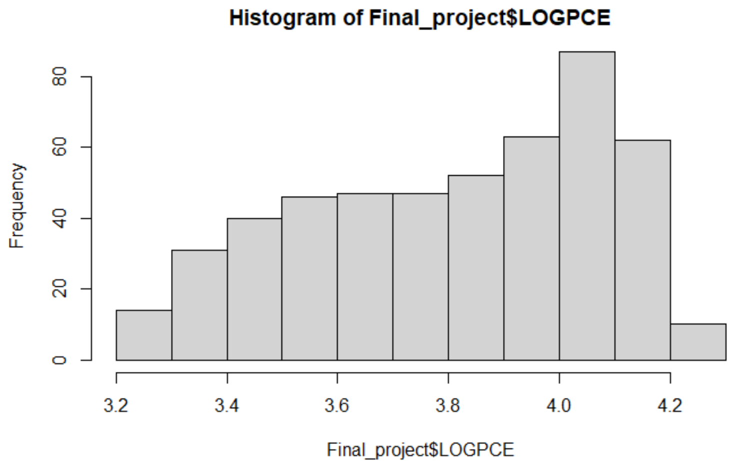
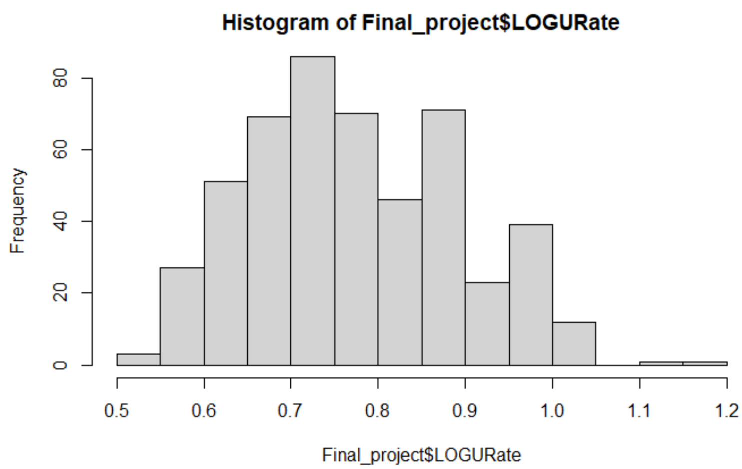
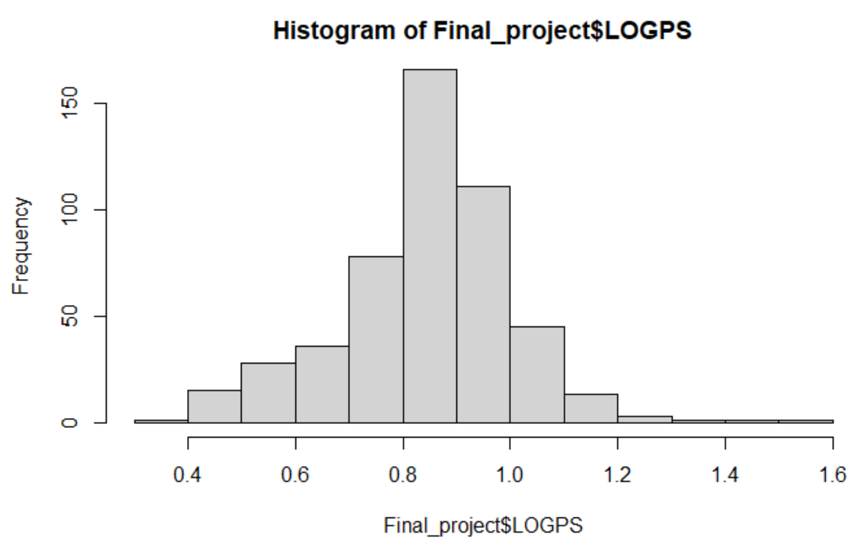
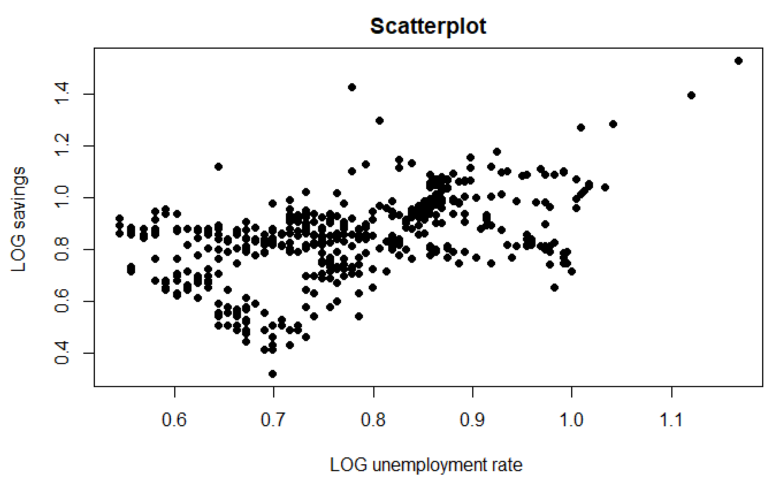
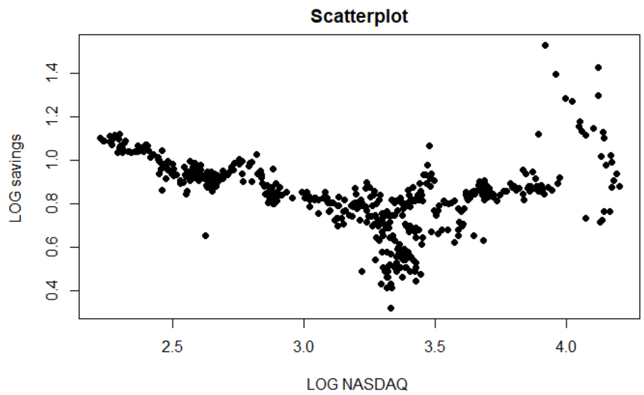
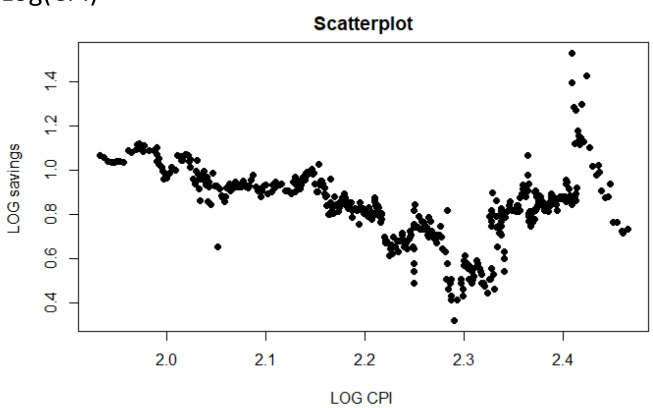
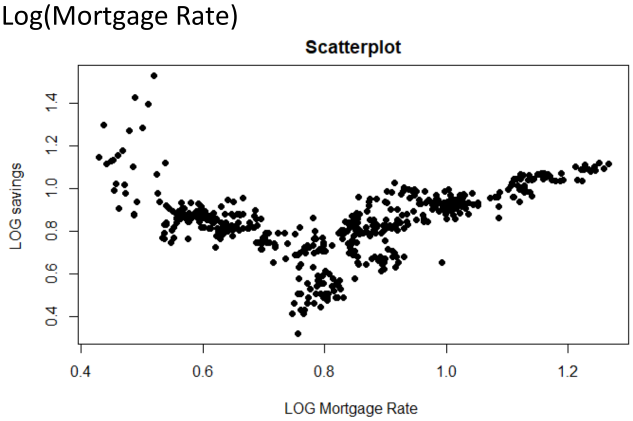
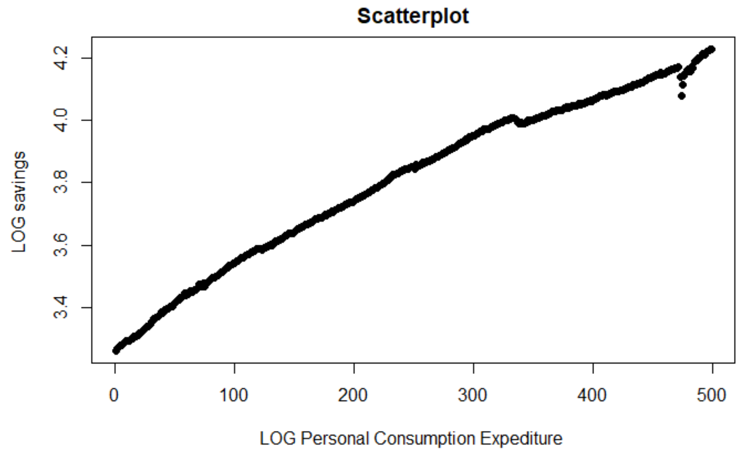
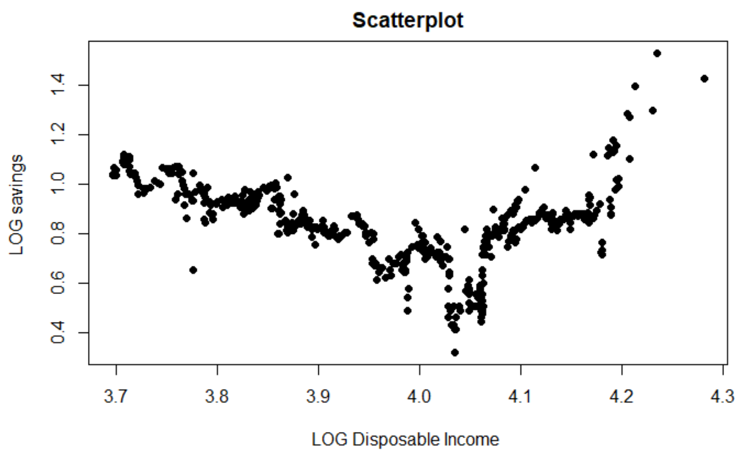
Residual standard error: 1.539 on 492 degrees of freedom
Multiple R-squared:  0.7426,    Adjusted R-squared:  0.7394
F-statistic: 236.5 on 6 and 492 DF,  p-value: < 2.2e-16
````

Model #2 is the best performing model with an Adjusted R^2 value of 0.7394. Model #2 includes all variables except the binary recission field. The Y intercept is negative, CPI, Mortgage rate, Unemp_rate, NASDAQ, disposable_income all of which have positive coefficients with Personal_consumption_expenditure being the only coefficient with a negative value. All variables are statistically significant at the 0.05 level. The model in its entirety is statistically significant at the 0.05 level.

```
```{r linear model3}
fit3<-lm(personal_savings ~ Mortgage_rate+Unemp_rate+NASDAQ+disposable_income+Personal_consumption_expenditure,
summary(fit3)
```
```

```
Call:
lm(formula = personal_savings ~ Mortgage_rate + Unemp_rate + 
    NASDAQ + disposable_income + Personal_consumption_expenditure, 
    data = Final_project)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7879 -1.3013  0.3507  1.2477  5.4262 

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      -1.331e+01  1.218e+00  -10.93   <2e-16 ***
Mortgage_rate                     4.870e-01  5.248e-02    9.28   <2e-16 ***
Unemp_rate                        8.536e-01  4.291e-02   19.89   <2e-16 ***
NASDAQ                            9.487e-04  4.805e-05   19.74   <2e-16 ***
disposable_income                 2.889e-03  1.909e-04   15.13   <2e-16 ***
Personal_consumption_expenditure -2.491e-03  1.414e-04  -17.62   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.58 on 493 degrees of freedom
Multiple R-squared:  0.7281,	Adjusted R-squared:  0.7254 
F-statistic: 264.1 on 5 and 493 DF,  p-value: < 2.2e-16
```

Model #3 is the worst performing model which excludes recession and CPI fields. Model #3 has an Adjusted R^2 of 0.7254. All variables are statistically significant at the 0.05 level. The model in its entirety is statistically significant at the 0.05 level.

VIF test:

- For fit1:

```
                             CPI                    Mortgage_rate                       Unemp_rate                           NASDAQ 
                      194.933230                        15.295371                         1.174107                         5.508326 
               disposable_income Personal_consumption_expenditure                   Recession_categ 
                       76.535557                       192.181077                         1.340669 
```

The VIF test is for collinearity. I used greater than ten as a high collinearity. In this test, we see that there is high collinearity with CPI, Mortgage rate, disposable income, and personal expenditure, as they are all over ten by a great amount. If it were to keep the variables, we would have to account for collinearity.

- For fit2:

```
                             CPI                    Mortgage_rate                       Unemp_rate                           NASDAQ 
                      177.600366                        11.887017                         1.173342                         5.451846 
               disposable_income Personal_consumption_expenditure 
                       75.282907                       179.687401 
```

In the second test, we still see a high collinearity with the same four variables. Again, we would have to take into account the collinearity if this test is used as the model.

- For fit3:

```
                   Mortgage_rate                       Unemp_rate                           NASDAQ                disposable_income 
                        6.555423                         1.166086                         4.548331                        73.329974 
Personal_consumption_expenditure 
                       66.753763 
```

In this test, there are only two variables with high collinearity; disposable income and personal consumption expenditure. This model would be the best use to get rid of collinearity. However, using this model I found that the R2 would be reduced by too much to make it effective.

Correlation matrix:

	v1_CPI	v2_MortgageRate	v3_UnempRate	v4_NASDAQ	v5_DisposableIncome	v6_PCE	v7_Recession	v8_PersonalSavings
v1_CPI	1.0000000	-0.9191833	-0.2629231	0.8033689	0.9895345	0.9910961	-0.1101652	-0.1994624
v2_MortgageRate	-0.9191833	1.0000000	0.2899592	-0.6421761	-0.8953269	-0.8774375	0.2412681	0.3110060
v3_UnempRate	-0.2629231	0.2899592	1.0000000	-0.2950891	-0.2556699	-0.2675020	0.1708873	0.5007660
v4_NASDAQ	0.8033689	-0.6421761	-0.2950891	1.0000000	0.8345709	0.8499857	-0.1356519	0.1477876
v5_DisposableIncome	0.9895345	-0.8953269	-0.2556699	0.8345709	1.0000000	0.9915729	-0.1073278	-0.1123876
v6_PCE	0.9910961	-0.8774375	-0.2675020	0.8499857	0.9915729	1.0000000	-0.1105069	-0.1472672
v7_Recession	-0.1101652	0.2412681	0.1708873	-0.1356519	-0.1073278	-0.1105069	1.0000000	0.1143346
v8_PersonalSavings	-0.1994624	0.3110060	0.5007660	0.1477876	-0.1123876	-0.1472672	0.1143346	1.0000000

Residual Plots



The residual cloud could be more evenly distributed in a cloud shape. There also is a significant pull in the red line which represents nonlinear relationships. The red line is trending upwards which shows heteroscedasticity aka a non-constant relationship. Heteroscedasticity does not appear to be an issue until 15.



The residuals do not look to be normally distributed from -3 to -1 and then again at 1 to 3. This observation is made by looking at how the tales skew in the lower and upper portions of the graph. From looking at the graphs shape it seems that the data is right skewed as both lower and upper portions are peeling in an upwards direction.



There is obvious nonlinearity shown in the Scale-Location plot. We have identified nonlinearity by observing the redline decreasing, then rapidly dipping, to then finally start to increase at multiple points. There is also an un-even distribution of plots which signals heteroscedasticity (non-constant relationships).



Plot 474 is well past cook's distance and is an outlier in my dataset. The redline is also showing a non – liner pattern as it dips down multiple times and then shoots upwards, ideally it would be straight.

APPLIED LOG TRANSFORMATION ON VARIABLES:

Histograms for transformed Variables:

- CPI:



- Disposable Income:



- Mortgage Rate



- NASDAQ:



- Personal Consumption Expenditure:



- Unemployment Rate:



- Personal Savings



After conducting a linear transformation (log) our data is much less skewed and better represents a normally distributed data set. There is still skewness present in our PCE, and CPI fields but overall, the histograms look much better.

- CPI: The logged version of the CPI histogram is somewhat evenly distributed with a slight left skew.
- Mortgage Rate: The logged version Mortgage Rate histogram is much better than the version which did not have a linear transformation applied as the graph looks normally distributed.
- Unemployment Rate: The logged version of the Unemployment Rate histogram is more evenly distributed than the original version which did not have a linear transformation applied.
- NASDAQ: The logged version of the NASDAQ histogram now shows a somewhat evenly distributed graph.
- Disposable Income: The logged version of the Disposable Income histogram is more evenly distributed than the original version which did not have a linear transformation applied.
- Personal Consumption Expenditure: The logged version of the Personal Consumption Expenditure histogram is somewhat evenly distributed with a slight left skew.
- Personal Savings: The histogram of Personal Savings skewness has been mostly eliminated and the logged versions histogram shows a normal distribution.

Scatter plots for transformed Variables:

- Log(Unemployment Rate)



The transformed (log) scatterplot for unemployment rate shows a flattened line from 0 to 1 and then has a couple of points that go up after 1.0.

- Log(NASDAQ)



The log scatterplot of NASDAQ shows a negative linear relationship from 0 to about 3.25. After 3.25, the plot becomes more erratic and confidence levels decrease.

- Log(CPI)



The log scatterplot of CPI shows a negative trend line going from 0 to 2.3. After 2.3, the plot shows a dramatic rise and becomes more erratic.

- Log(Mortgage Rate)



The log scatterplot of mortgage rate shows an erratic patterning of plots from 0 to 0.5. After 0.5, there seems to be a negative linear relationship until 0.8 then becomes a positive linear relationship for the rest of the graph. At 0.8, there are many plots that make that point more unreliable.

- Log(Personal Consumption Expenditure)



This log scatterplot of personal consumption expenditure shows a strong linear correlation between the two variables. There is a definite linear line from 0 to 500.

- Log(Disposable Income)



The scatterplot for the log of disposable income shows a negative linear relationship from 3.7 to about 4.0. After 4.0, the plots become more erratic and confidence levels go down, however, the seems to be a little positive relationship.

Linear Regression Analysis on transformed Variables:

```
```{r regression - tranformed variables}
Log_fit1<-lm(LOGPS ~ LOGMRate+LOGURate+LOGNASDAQ+LOGDI+LOGCPI+LOGPCE+Recession_categ, data = Final_project)
summary(Log_fit1)
```

```

```
Call:
lm(formula = LOGPS ~ LOGMRate + LOGURate + LOGNASDAQ + LOGDI +
    LOGCPI + LOGPCE + Recession_categ, data = Final_project)

Residuals:
     Min   1Q   Median   3Q   Max
-0.58075 -0.03578 0.00862 0.04045 0.15472

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)   -14.16893   0.77509 -18.280 < 2e-16 ***
LOGMRate   -0.87758   0.06760 -12.981 < 2e-16 ***
LOGURate   -0.29821   0.04381 -6.808 2.91e-11 ***
LOGNASDAQ   -0.10979   0.03148 -3.487 0.000532 ***
LOGDI   8.57705   0.30492 28.129 < 2e-16 ***
LOGCPI   11.17763   0.52910 21.126 < 2e-16 ***
LOGPCE   -11.19690   0.36332 -30.818 < 2e-16 ***
Recession_categ -0.04652   0.01124 -4.138 4.12e-05 ***

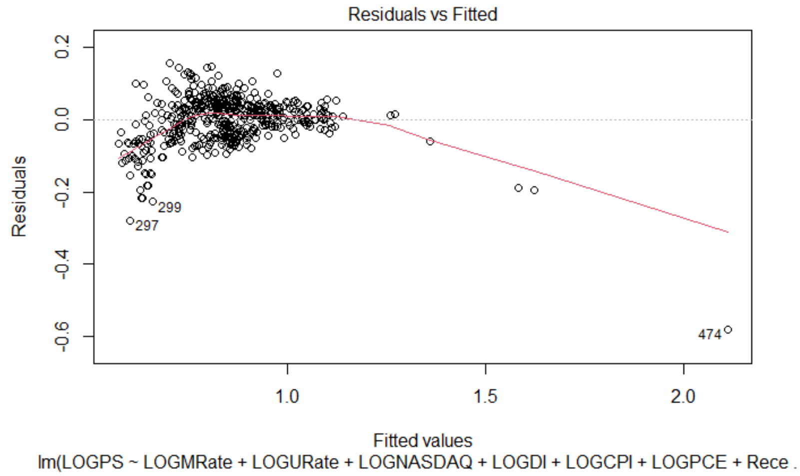
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06712 on 491 degrees of freedom
Multiple R-squared: 0.8283,   Adjusted R-squared: 0.8259
F-statistic: 338.5 on 7 and 491 DF, p-value: < 2.2e-16
```

Model #4 includes all fields of the data set all of which have gone through a liner transformation. The p-value of the entire model is under 0.05 and is statistically significant. All variables are statistically significant at the 0.05 level. The Adjusted R^2 value is 0.8259 which is the highest R^2 we have achieved. Model #4 represents our parsimonious model.

Correlation Matrix for the transformed Variables:

```
```{r}
cor(Final_vector1)
```

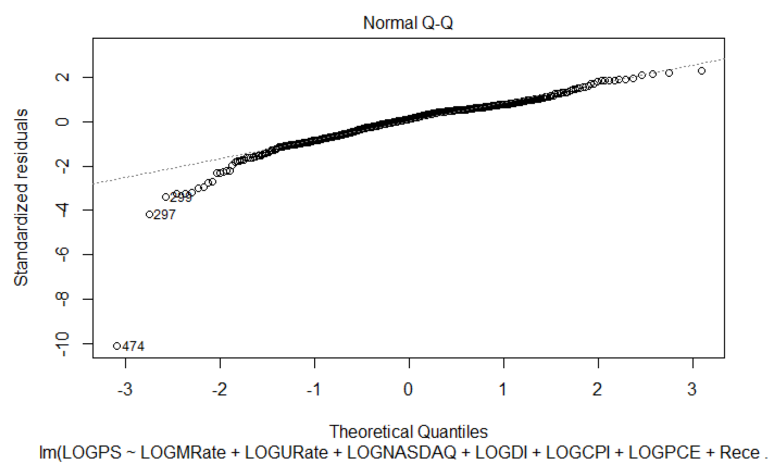
```

```
                       v1_CPI v2_MortgageRate v3_UnempRate v4_NASDAQ v5_DisposableIncome   v6_PCE v7_Recession v8_PersonalSavings
v1_CPI   1.0000000   -0.9616515   -0.3495828 0.9608305   0.9925776 0.9983878 -0.12194696   -0.35329558
v2_MortgageRate   -0.9616515   1.0000000   0.2392242 -0.9263085   -0.9602482 -0.9591049   0.19871165   0.21080780
v3_UnempRate   -0.3495828   0.2392242   1.0000000 -0.4625635   -0.3532591 -0.3731790   0.15898556   0.46375246
v4_NASDAQ   0.9608305   -0.9263085   -0.4625635 1.0000000   0.9672418 0.9635448 -0.18452895   -0.30221360
v5_DisposableIncome 0.9925776   -0.9602482   -0.3532591 0.9672418   1.0000000 0.9954867 -0.12161643   -0.31564095
v6_PCE   0.9983878   -0.9591049   -0.3731790 0.9635448   0.9954867 1.0000000 -0.13251466   -0.37331814
v7_Recession   -0.1219470   0.1987117   0.1589856 -0.1845289   -0.1216164 -0.1325147   1.00000000   0.05751202
v8_PersonalSavings -0.3532956   0.2108078   0.4637525 -0.3022136   -0.3156410 -0.3733181   0.05751202   1.00000000
```

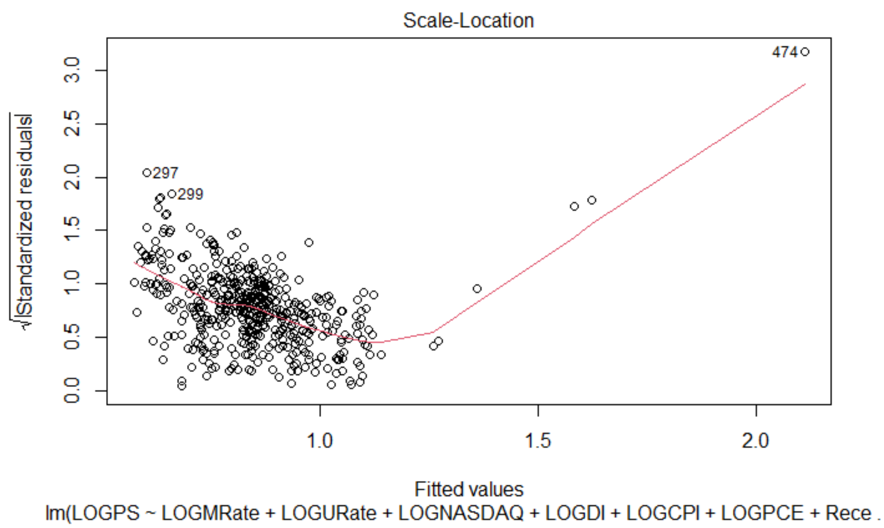
Residual plots:



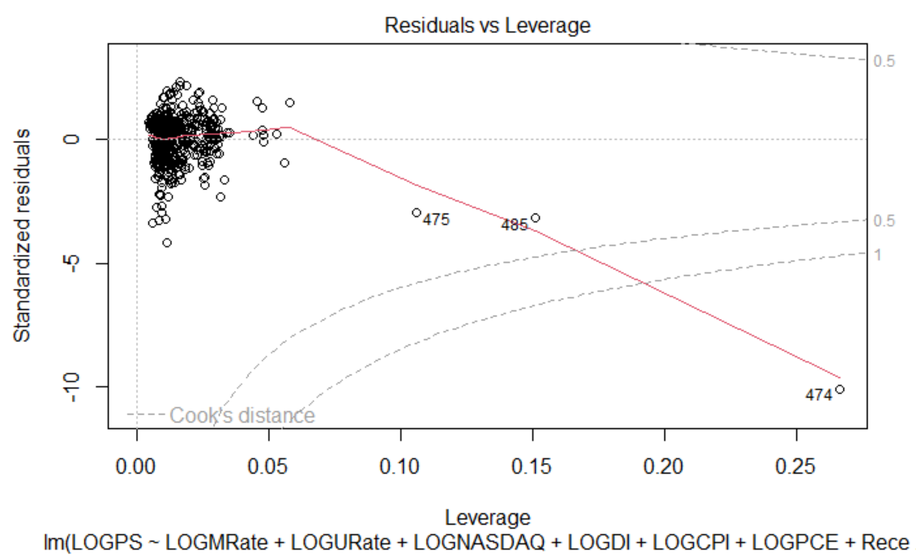
The residual cloud could be more evenly distributed across the X axis in a cloud shape pattern. The red line is trending up from 0.0 to 0.5 and is only straight from 0.5 to 1.5 showing a non-constant relationship between variables and then starts to trend down showing heteroscedasticity. The clump of points is more concentrated when compared to the previous version which did not include any linear transformations.



The Normal Q-Q graph of logged variables is an improvement from the version which did not include any liner transformations and displayed skewness. The residuals don't look normally distributed from -3.5 to -2.5 and then again at 2.5.



This graph shows strong Heteroscedasticity, the plots are non-constant as the red line trends up and also show signs of non-linearity with the dips and increases of the red line trajectory.



The red line is straight from 0.00 to 0.05 but then dips downwards. Plots 474 continues to be an outlier as it is past Cook's distance.

References

1. Largest countries in the world (by area). Worldometer. (n.d.). Retrieved December 6, 2022, from https://www.worldometers.info/geography/largest-countries-in-the-world/
2. Mirashi, S. (2022, July 29). US Macroeconomic Data. Kaggle. Retrieved October 20, 2022, from https://www.kaggle.com/datasets/sarthmirashi07/us-macroeconomic-data
3. Most populous countries in the world (2022). Worldometer. (n.d.). Retrieved December 6, 2022, from https://www.worldometers.info/population/most-populous-countries/
4. Team, T. I. (2022, June 16). A review of past recessions. Investopedia. Retrieved December 6, 2022, from https://www.investopedia.com/articles/economics/08/past- recessions.asp