FINAL PROJECT

Andrew Castillo

2019-11-14 12:15:04

0A. INTRODUCTION

This paper will examine the impact of macroeconomic recession and growth in the US economy on the volatility of the stock market. Volatility of the market is the degree of variation of a trading price series over time as measured by the standard deviation of logarithmic returns.

Daily open is the volatility at the time the stock market opens (9:30 AM Eastern Standard Time). Daily close is similar, but at closing, which is at 4:00 PM EST.

Daily high and low are the highest and lowest volatility measures on a given day.

Higher volatility means there is an increased risk of fiscal loss when trading. By understanding the impact of the current state of the economy on the stability of the market, investors can make educated choices and take calculated risks when choosing when and where to place their capital.

There are three questions we hope to answer in this analysis:

1. Does recession correlate with higher volatility in stock options?
2. Is the difference between daily opening and closing volatility impacted by recession? If so, what is the impact?
3. Is there a systemic difference in daily high volatility during periods of growth and periods of recession?

These questions have impact on real-time investment of stock options.
We expect recession to correlate with higher volatility, due to an increase in panic-selling. This is a situation where shareholders reach their tolerance for loss during an economic downturn, and sell their positions to minimize further damage to the value of their holdings.

1A. DESCRIPTION OF DATA

The VIX Index is the data set we will use. This data set is sourced from Cboe Global Markets, and measures the "fear" in the market with four metrics: daily high, low, open, and close. Volatility is a unitless measurement.

Volatility is measured by taking the standard deviation of prices all option stocks over the previous three months. It is unitless. The higher the numerical value, the more volatile the market is (often referenced as market 'fear' or 'uncertainty').

To answer the questions presented in the introduction 1), 2), and 3), we'll look at how the volatility of the market changes over time and on a daily basis and the contrast between each cycle of growth and recession. The objective is to find if there is a connection between recession and higher volatility, and similarly growth and lower volatility.

For 1), we'll analyze one-variable statistics.

For 2), we'll look at opening and closing volatility, and compare the difference of their means over each period. A paired sample T-test will be applied. Opening and closing are the measurements of overall market stability at the start of a trading day, and at the end of the same day. We'll take the means of each category over a sample size of n = 252 days for open and close, selected arbitrarily.

For 3), we'll use a bootstrap study with data from the time intervals Sep 1, 2008 to Sep 1, 2009 for recession and Sep 1, 2006 to Sep 1, 2007 for growth, by the National Bureau of Economic Research.

To conclude, we'll summarize the data and examine what possible confounding variables could be.

1B. NECESSARY METADATA

For this analysis, we need to know when the U.S. economy is in periods of recession and growth. According to the National Bureau of Economic Research, the periods of recession are Mar 2001 - Nov 2001, Dec 2007 - June 2009. The growth period we'll be using is Nov 2001 - Dec 2007, by the same source. This data is categorical.

2A. ONE-VARIABLE ANALYSIS

Next, we'll start analyzing the data. The one-variable statistics we'll consider are mean, median, standard deviation, and variance.

The purpose of this section is to present the data as-is, and answer question 1).

The data set is previewed below:

```
##         Date VIX.Open VIX.High VIX.Low VIX.Close
## 1   1/2/2004    17.96    18.68   17.54     18.22
## 2   1/5/2004    18.45    18.49   17.44     17.49
## 3   1/6/2004    17.66    17.67   16.19     16.73
## 4   1/7/2004    16.72    16.75   15.50     15.50
## 5   1/8/2004    15.42    15.68   15.32     15.61
## 6   1/9/2004    16.15    16.88   15.57     16.75
## 7  1/12/2004    17.32    17.46   16.79     16.82
## 8  1/13/2004    16.60    18.33   16.53     18.04
## 9  1/14/2004    17.29    17.30   16.40     16.75
## 10 1/15/2004    17.07    17.31   15.49     15.56
```

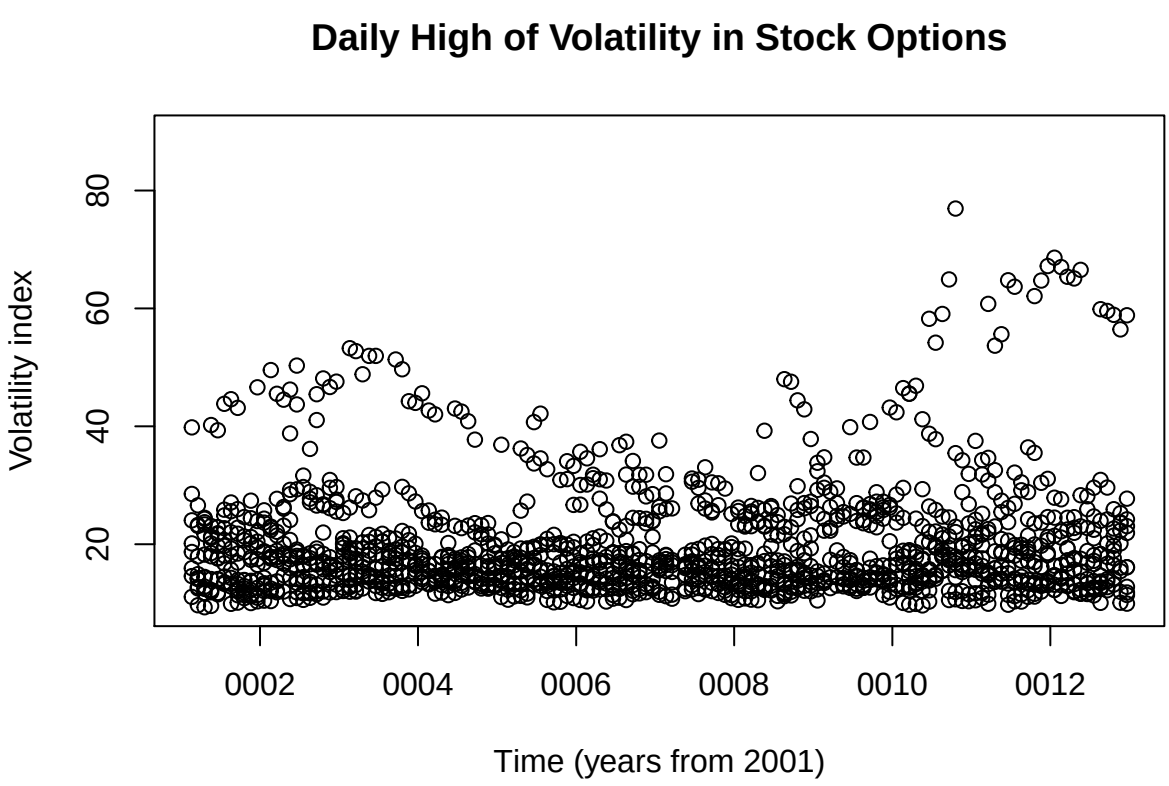
Considering the entire data set, the mean of opening volatility is 18.5019176, the mean of closing is 18.4007529. The mean daily low is 17.648893 and the mean high is 19.3573111. With respect to our chosen time intervals, during recession we have:

Mean opening of 45.6898347, mean closing of 45.6319835, and mean high and low values of 48.5206612, 45.6319835, respectively.

And during growth, we have a mean opening of 15.8828571, mean closing of 15.7938701, and mean high and low values of 16.4756623, 15.7938701, respectively.

The mean of all variables together is 18.4772187.

Next, we'll plot daily high over the entire data set, to view the most intense peaks of fear in the market:



The mean of the daily high over the course of the entire data set is 19.3573111 The standard deviation is 9.3896241.

Variance is computed with the formula

$\sigma_x{}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$

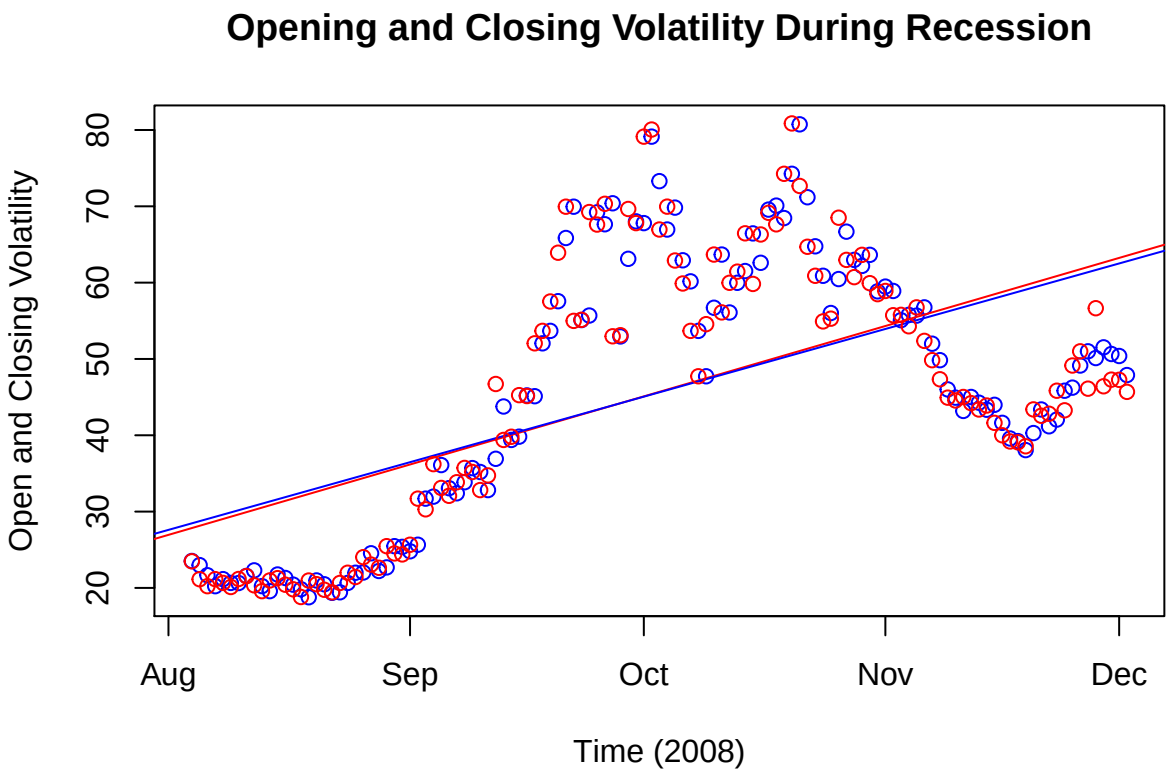
Variance of the volatility during growth (2001 - Sep 2008) for daily high is 25.4652613, and standard deviation (the square root of the variance) is 5.0463117

In recession (2008 - 2009), we have the variance of the high at 358.6690612, and standard deviation is 18.9385602

2B. TWO-VARIABLE ANALYSIS

Our two-variable analysis will include the correlation of daily high to daily low, and a regression line of open and close compared over the time period. The purpose of this section is to answer question 2).

Market open and close during the 2008 financial crisis are plotted against each other here. Open is colored red, and close is colored blue:

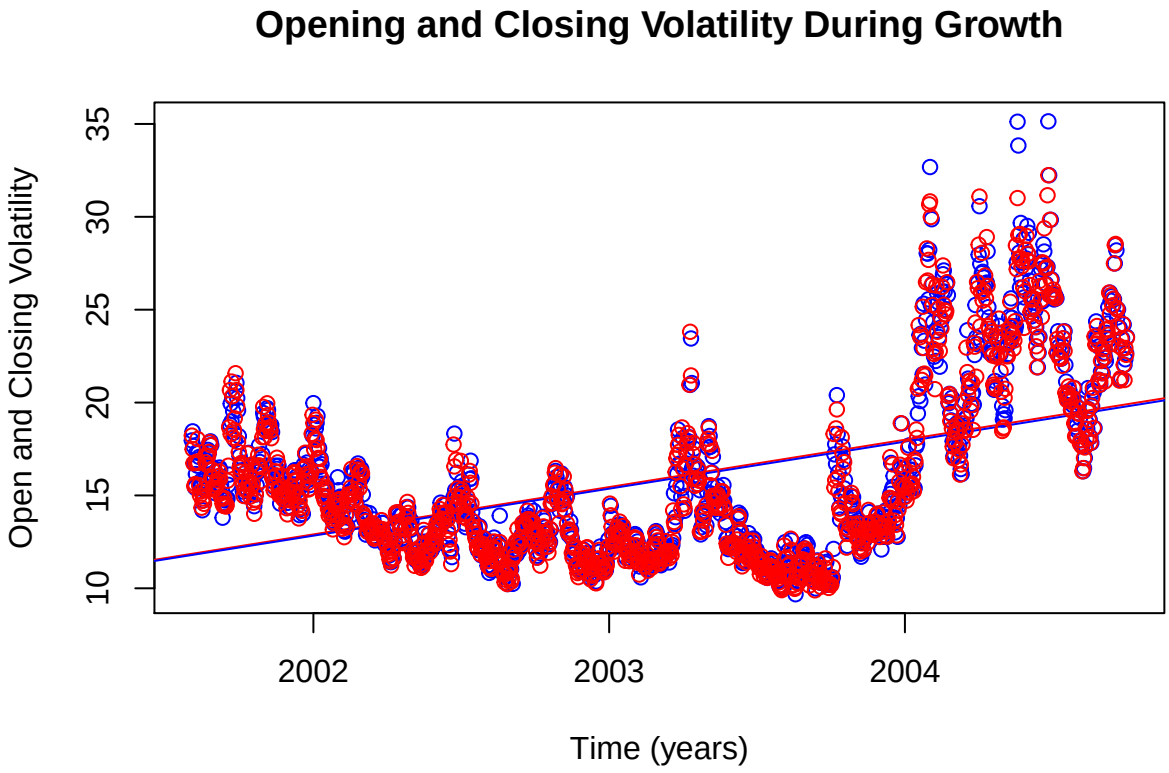


Though slight, note the higher slope value on the red (opening) line. This indicates in general, volatility during recession is higher at open than at close (we'll check later if this is statistically significant).

The jump during September 12 - October 8 is the time period when the housing bubble burst. Over this time period, the mean daily high was 41.0815789 and the mean daily low was 35.5484211, and overall volatility averaged at 38.2386842.

The correlation coefficient is 0.9684801, meaning there is a very strong connection between open and close on any given day during recession, and the trend lines are pictured in the graph.

Pictured below is a plot of the market in a period of growth:



Here, the trend lines both have a shallower slope, and are also more strongly correlated. The correlation coefficient here is 0.9762068

The scaling is different (only going up to 35 on the y-axis, where the recession graph scales to 80).

2C. CONFIDENCE INTERVAL

There will be two confidence intervals we consider- daily open and daily close. Furthermore, we'll plot these for the recession and non-recession periods, and then overall.

An overall confidence interval on daily closing prices is calculated with the equation:

$\sigma_{\bar{x}} = \frac{\sigma}{sqrtn}$

$\bar{x} \pm 1.96 \frac{\sigma}{sqrtn}$

And its z-score, which is calculated by:

$\frac{x-\mu}{\sigma}$

```
##
##  One Sample t-test
##
## data:  gr$VIX.Open
## t = 52.463, df = 251, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  13.54258 14.59901
## sample estimates:
## mean of x
##  14.07079
```

```
##
##  One Sample t-test
##
## data:  gr$VIX.Close
## t = 51.116, df = 251, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  13.47395 14.55383
## sample estimates:
## mean of x
##  14.01389
```

Based on the results of our calculations, a 95% confidence interval of volatility at open during recession is (13.47395, 14.55383), and during growth is (13.54258, 14.59901). This means, during a recession, we can say with 95% certainty the value of volatility will be between (39.25194, 42.63537) at close, and (39.44434, 42.84717) at open.

The recession interval roughly three times wider than the growth interval, and also the values are roughly tripled, with measurements at 13.54258 during growth and 39.25194 during recession on the lower ends of each respective confidence interval.

3A. HYPOTHESIS TEST

Another approach is to use a t-test to examine if the mean difference between opening and closing volatility is statistically significant. Here, we'll apply a paired-sample t-test on open and close data points selected over the course of a year-long interval (n = 251) from Sep 1, 2008 to Sep 1, 2009 for recession and Sep 1, 2006 to Sep 1, 2007 for growth.

The purpose of this section is to answer question 3).

Our null hypothesis is that the mean difference between opening and closing is 0.

Applying our test to the recession time interval, we see:

```
##
##  Paired t-test
##
## data:  re$VIX.Open and re$VIX.Close
## t = 0.97727, df = 251, p-value = 0.3294
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2051898  0.6093961
## sample estimates:
## mean of the differences
##               0.2021032
```

The resulting t-value is 0.97727, and the p-value is 0.97727. Because the p value is greater than 0.05, we accept the null hypothesis, concluding there is a not statistically significant difference between high and low volatility during recession.

Mirroring this test in times of growth, we find:

```
##
##  Paired t-test
##
## data:  gr$VIX.Open and gr$VIX.Close
## t = 0.8234, df = 251, p-value = 0.4111
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.07920407  0.19301359
## sample estimates:
## mean of the differences
##              0.05690476
```

The resulting t-value is 0.8234, and the p-value is 0.4111. Because the p value is still greater than 0.05, we reject the null hypothesis, concluding there is not a statistically significant difference during times of growth either.

3B. SECOND HYPOTHESIS TEST

The second hypothesis test is if daily high and daily low are independent during recession. For this, we'll apply a chi-squared test with the formula:

$\tilde{\chi}^2 = \frac{1}{d}\sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  chirec$X1 and chirec$X2
## X-squared = 2550, df = NA, p-value = 0.0004998
```

With null and alternative hypotheses H_0, H_a.

H_0 : There is no association between the opening and closing volatility during recession.

H_a : There is an association between association between the opening and closing volatility.

The resulting Chi-square value is 2550, and the p-value is 0.0004998. Because the p value is less than than 0.05, we reject the null hypothesis, we conclude daily high and daily low are not independent in times of recession.

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  chirec$X3 and chirec$X4
## X-squared = 2244, df = NA, p-value = 0.4478
```

Similarly, during growth, the Chi-square value is 2244, and the p-value is 0.4488. Because the p value is greater than 0.05, we fail reject the null hypothesis, and conclude the daily high and low are independent in times of growth.

4A. BOOTSTRAP STUDY

Our last analytical piece will be a bootstrap study of the daily high during periods of recession and growth. We'll use our previously selected time intervals of (n = 251) from Sep 1, 2008 to Sep 1, 2009 for recession and Sep 1, 2006 to Sep 1, 2007 for growth.

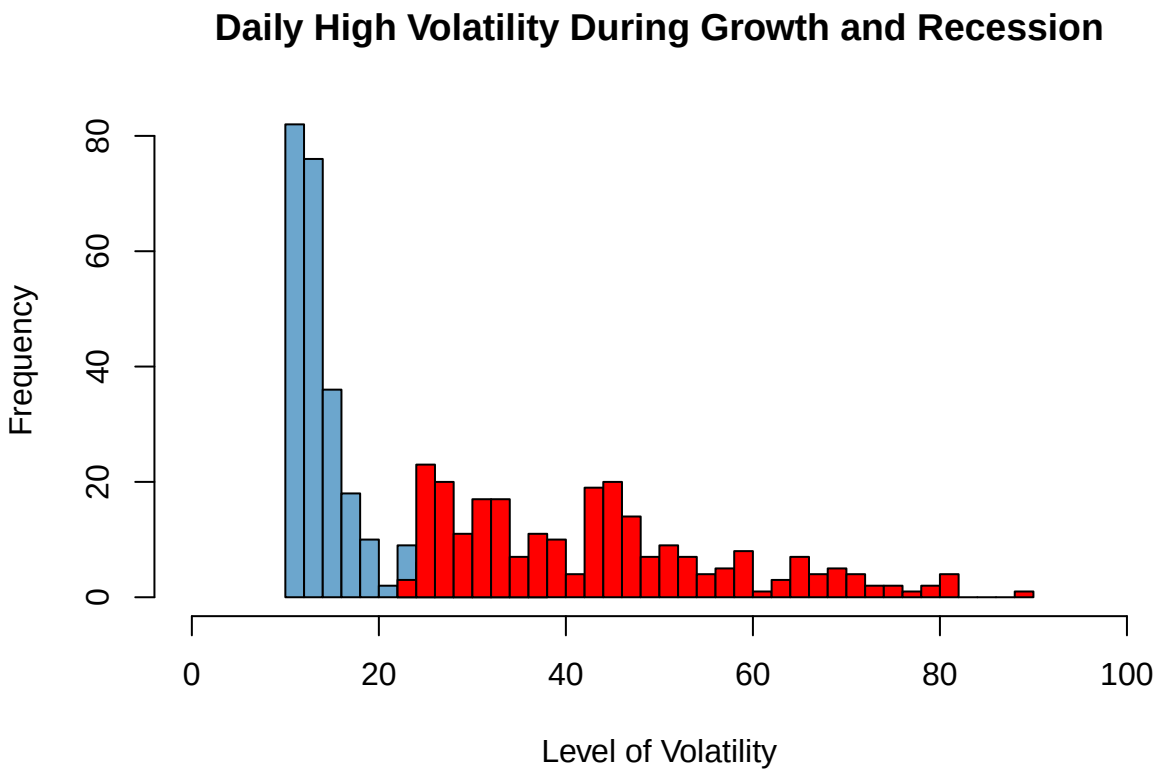
The procedure is as follows:

1. Create a subset of daily high measurements with recession and growth labeled. Show a histogram of each.

2. Compute the t statistic between the two groups, which quantifies how large the difference in means is compared to what you'd expect by chance

3. Shuffle up the recession and growth labels, and recompute the t statistic 10^4 times. The p-value is the proportion of these that are more extreme than our observed value.

(1) First, histograms of the variables are overlaid:



As pictured, the lower end of the volatility scale in the range of (10, 25) is saturated with blue entries, corresponding to growth. The upper end of (25,85) is a more of an even spread of the recession values.

(2) Then, the two-sample t statistic is computed below using t.test:

```
##
##  Welch Two Sample t-test
##
## data:  re$VIX.High and gr$VIX.High
## t = 28.71, df = 301.71, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  26.45184 30.34482
## sample estimates:
## mean of x mean of y
##  43.10274  14.70440
```

Based on our calculations, the t-value is 28.71.

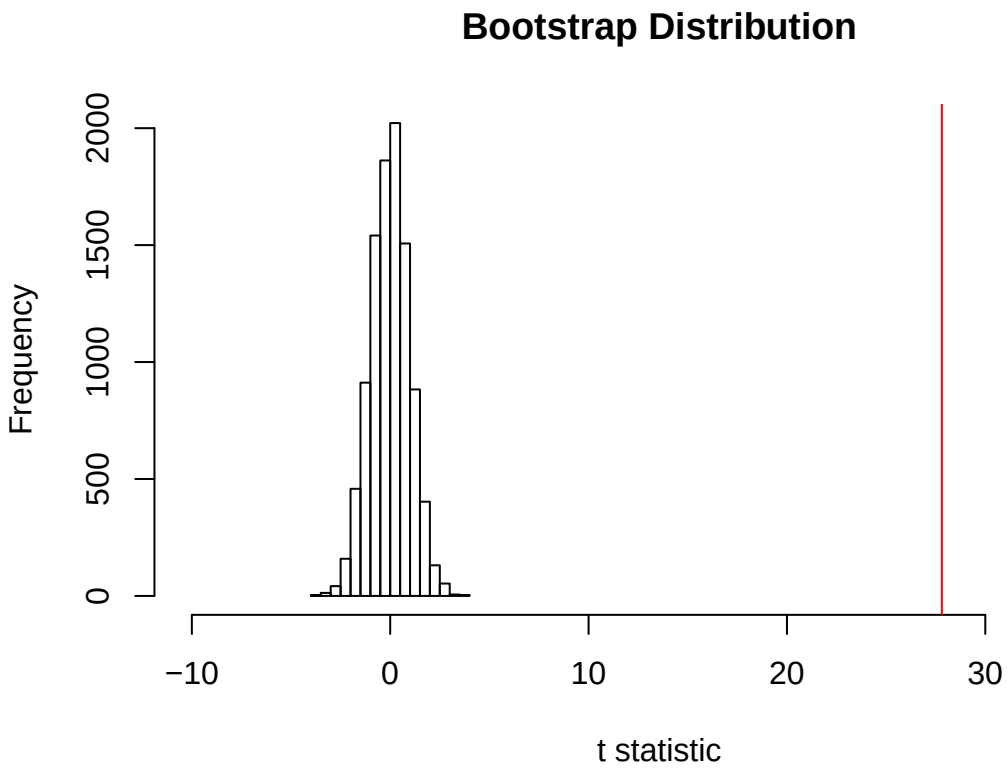
(3) Now, we repeat the code above, but with the the high and low labels shuffeled randomly using sample().

```
##         t
## -0.014118
```

Our rearrangement yields a t-value of -0.014118 pictured above, which is much smaller than the first result of 28.71. If there is no systematic difference between the groups, then rearranging the labels shouldn't affect anything.

(4) Repeat the test 10^4 times, illustrating the result in a histogram:

Then, we find the p-value: the proportion of shuffled t-statistics that are larger in absolute value than the one observed in step (2).



The p-value is approximately zero, which can be seen from the histogram.

In conclusion, we have strong evidence that the average high volatility of periods of recession is greater than the respective high of growth periods. The mean difference is 28.69562 based on our bootstrap test, and this difference is significant ($p \approx 0$, 10^4 samplings).

5A. CONCLUSION

Based on our analysis, the answers to our presented questions are:

1. Recession correlated with a high volatility in stock options, by our one-variable analysis and bootstrap study.
2. The mean difference between daily opening and closing volatility is not impacted by recession. Furthermore, daily high and low are independent in times of growth, but not in times of recession, by our two-variable analysis and hypothesis test.
3. There is a systemic difference between average daily high during recession and during growth.

Overall, the relationship between volatility of stock options and economic growth is a positive relationship. As the economy grows, the market is generally more stable.

Possible confounding variables include:

Change in one specific area of stocks (i.e. dramatic change in oil stocks while the rest of the market is not volatile), if the change in one sector is cancelled by another, current events, and political climate.

Works Consulted:

https://www.federalreserve.gov/econres.htm

http://www.cboe.com/products/vix-index-volatility

https://www.nber.org/