STAT 434 DATA ANALYSIS - MAX SCHEIBER AND RUSLAN ZAGATSKIY

1. Overview

Our goal was to explore the relationship between the NSA scandals of 2013 and securities of related tech sector stocks.

2. Raw data to polished data

We obtained consolidated trade data from Wharton Research Data Services (WRDS) for Google (GOOG), Apple (AAPL), Microsoft (MSFT), and Facebook (FB) for the January 1st, 2013 through July 31st, 2013 timespan. These consolidated datasets contain average execution prices and total shares volume a few times each second. We chose these stocks because they were the ones identified in the NSA's leaked slide deck.

To make data analysis more manageable, we converted this data into hourly buckets from 9am to 3pm, inclusive. We did this by taking the average execution price over each bucket, equally weighted within the bucket. (For the final presentation, it may be better to volume-weight this mean). We also took volume sums for each bucket.

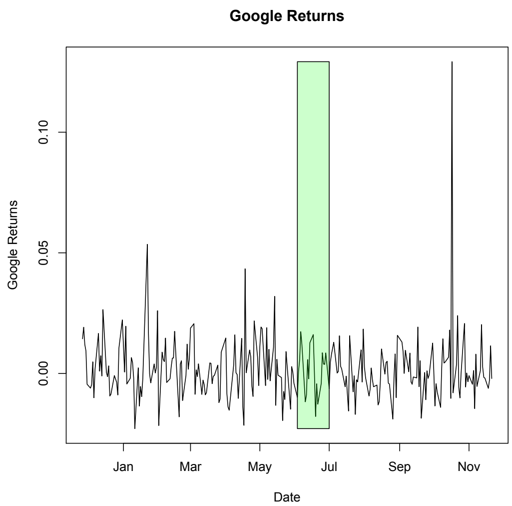
We then created a basket of tech stocks by taking the equal-weighted average of each bucket of all stocks considered. The reason for doing this was to enhance our signal-to-noise ratio; this basket would minimize the effects in securities pricing and volatility of, say, one company's earnings report, while confirming sector-wide trends.

Within this basket, we computed percent returns, which allowed us to compute a scale-independent moving volatility. We currently measure volatility from a given time bucket over the previous 30 time buckets. Given that there are seven buckets per day, we see our volatility measure cover a duration of just over four days. (For the future, we may wish to model volatility with an exponential smooth.)

Date: 11.25.2013.

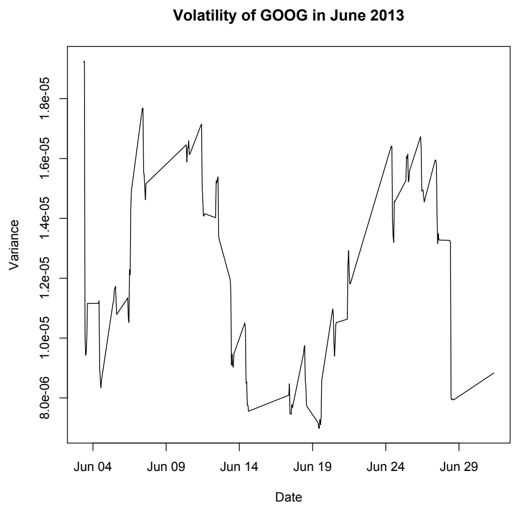
3. Google analysis

Before performing analysis on the basket of stocks, we performed some basic analysis of Google's returns.

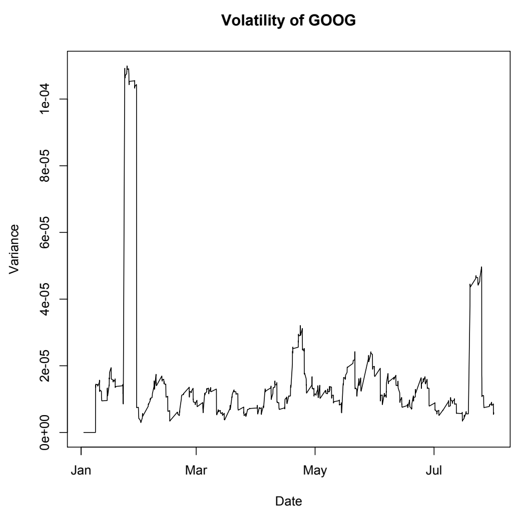


Immediately, we can see that Google did not really have any outliers in its return over the summer. Its main spikes seem to come from its earning reports, implying that the market did not judge the NSA scandals to really affect Google.

We next look at Google's hourly volatility throughout June, which is when the majority of the NSA leaks happened. It initially seems obvious via eyeballing that Google's volatility spiked on June 6th (the date of *The Guardian*'s first leak) and June 20th (the date of *The Guardian*'s second leak).



However, in context, these spikes are extremely minimal. Let us, for example, plot volatility throughout 2013.

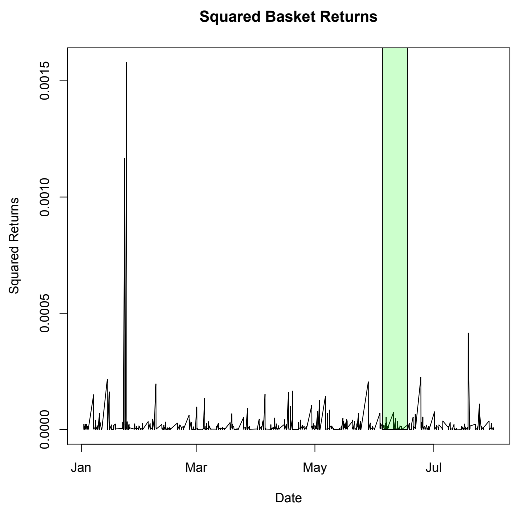


Any June spikes that could be related to the NSA do not look very significant in the context of Google's volatility throughout the year. While it is ostensible that these small June spikes may indeed have occurred as a result of the NSA scandal, this seems to be an early indicator that our hypothesis may not hold.

4. Tech sector basket analysis - full data set

For further confirmation, we now look at the full tech sector basket as discussed in section 2.

To check for any sort of heteroskedasticity, we plot squared returns of the basket. Note that while we also applied our moving volatility model to basket data, the plot of squared returns provides a simpler visual analysis that leads to the same conclusions.



While the basket's 2013 squared returns have some unusual spikes, it looks generally stable. More rigorously, we considered an ARCH(1) model for the basket's returns. The model's first coefficient is statistically insignificant, implying homoskedasticity. Of course, testing just an ARCH(1) is not sufficient. Next we look at an ARCH(8), where we find that the 6th and 7th lags are significant. Finally, we consider a GARCH(1,1), which is essentially equivalent to an infinte order ARCH. The significance of both the ARCH(alpha1) and GARCH(beta1) terms confirms heteroskedasticty in the basket return variance.

ARCH(1)

```
Model:
GARCH(0,1)

Residuals:
      Min        1Q    Median        3Q
-10.94949  -0.45224   0.01911   0.49613   9.4:

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 1.317e-05   1.671e-07   78.793   <2e-16 **
a1 9.939e-03   1.802e-02    0.552    0.581
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.

Diagnostic Tests:
    Jarque Bera Test

data:  Residuals
X-squared = 23056.59, df = 2, p-value < 2.2e-:


    Box-Ljung test

data:  Squared.Residuals
X-squared = 0.0067, df = 1, p-value = 0.9346
```

ARCH(8)

```
Model:
GARCH(0,8)

Residuals:
     Min       1Q   Median       3Q      Max
-6.68405 -0.51203  0.01798  0.56012 10.08654

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 7.765e-06   4.363e-07   17.798  < 2e-16 ***
a1 3.968e-02   2.513e-02    1.579   0.1144
a2 9.101e-03   1.496e-02    0.608   0.5430
a3 5.197e-16   1.063e-02    0.000   1.0000
a4 1.433e-03   6.298e-03    0.228   0.8200
a5 1.336e-03   1.113e-02    0.120   0.9044
a6 8.702e-02   1.404e-02    6.200 5.65e-10 ***
a7 1.521e-01   2.444e-02    6.225 4.83e-10 ***
a8 4.285e-02   2.202e-02    1.946   0.0517 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.

Diagnostic Tests:
    Jarque Bera Test

data:  Residuals
X-squared = 6597.186, df = 2, p-value < 2.2e-16


    Box-Ljung test

data:  Squared.Residuals
```

GARCH(1,1)

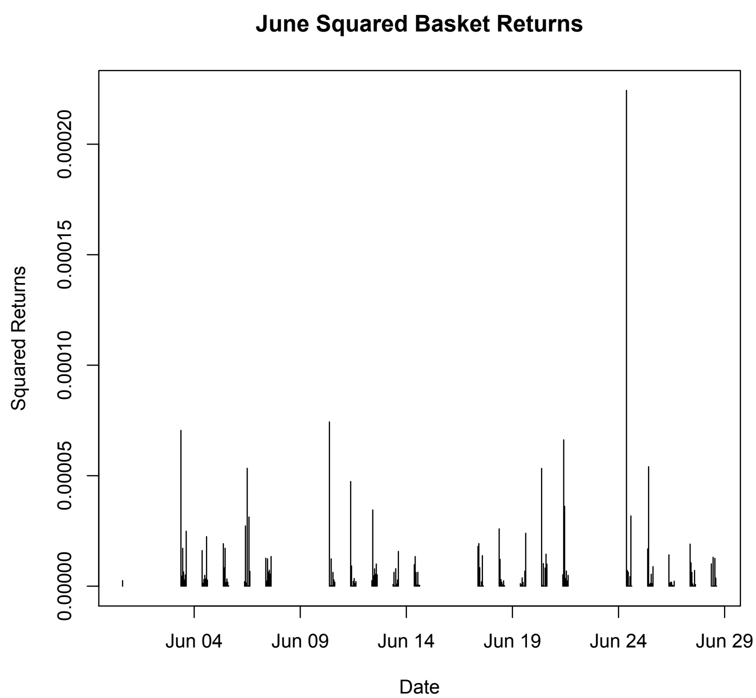
```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model : sGARCH(1,1)
Mean Model  : ARFIMA(1,0,1)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t
mu      0.000088    0.000129  0.682198 0.4951
ar1     0.013567    0.188177  0.072097 0.9425
ma1     0.174019    0.185830  0.936438 0.3490
omega   0.000001    0.000000  3.865306 0.0001
alpha1  0.034592    0.009977  3.467102 0.0005
beta1   0.885028    0.026030 33.999800 0.0000
```

5. June indicator variables

The discovery above is unsurprsing - heteroskedastic variance is a common feature of financial returns. A more interesting question is whether the NSA leaks, independent of other events, impacted volatility.

One approach is to regress using an indicator variable. We set an indicator variable to be true for all buckets on June 6th, the day of the first NSA leak, and we set another indicator variable to be true for all buckets on June 20th, the day of a large second NSA leak. We test the significance of these indicators first for our full data set and find that neither have t-Statistics large enough to reject the null that they have no impact.

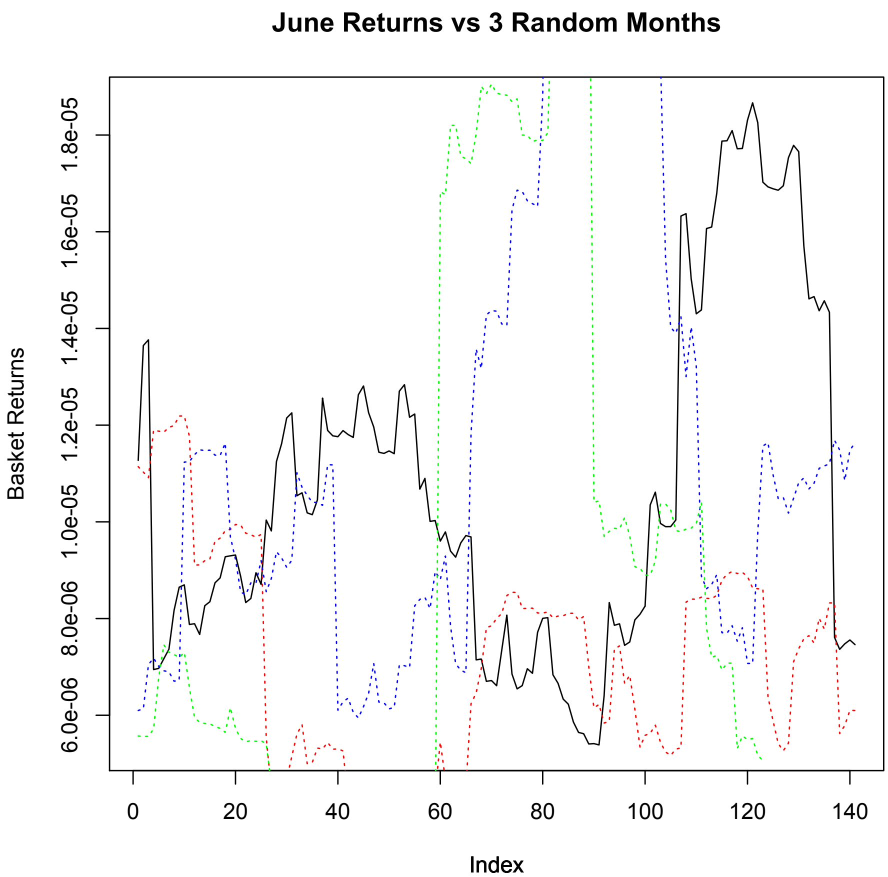
The section above considers long-run impact of the NSA leaks on the volatiltiy of our tech basket. However, it is also worth considering short-run volatility changes centered around the time of the leaks. We first consider the month of June since that is when the leaks began. A plot of squared returns suggests that there is some increased volatiltiy towards the end of June, particularly around June 21st (our model's volatility is lagged by about 3 days). This is confirmed by a GARCH(1,1), which has significant coefficients for lagged variance and lagged squared error terms.



The result of these regressions is that only the June 20th leak is statistically significant for June volatility. The June 6th indicator does not have significant explanatory power for volatility in June or in all of 2013. It is arguable that *The Guardian*'s second leak did spur more volatility in the market, but this washes out in the long run due to other similar events that affect the tech sector. There are a variety of possible explanations for the significance of June 20th versus the null effect of June 6th, but these are purely speculative.

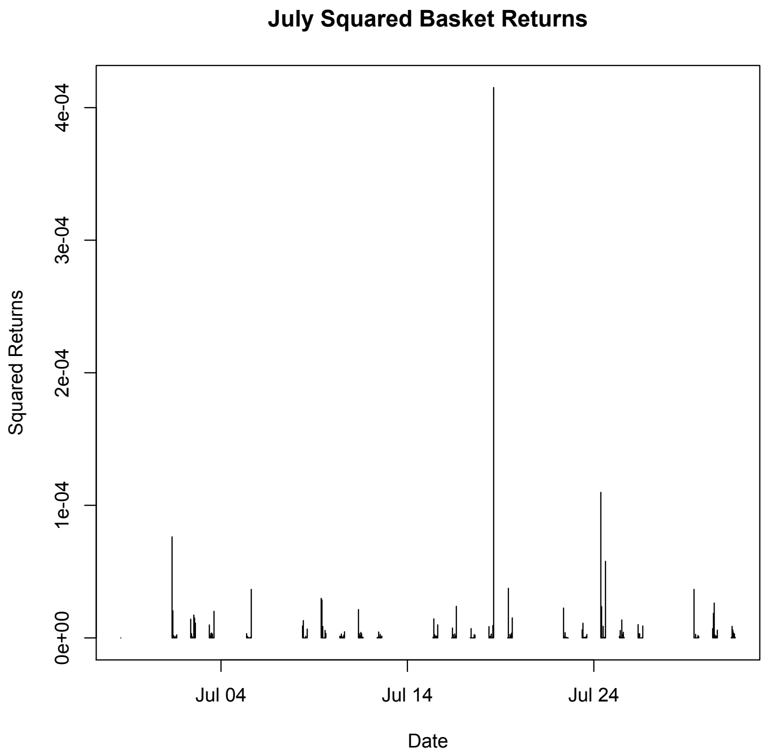
```
Call:
lm(formula = vol[728:868] ~ indic.june20[728:868])

Residuals:
       Min         1Q     Median         3Q        Max 
-5.459e-06 -2.841e-06 -6.594e-07  1.784e-06  7.822e-06 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            1.085e-05  3.096e-07   35.03   <2e-16 **
indic.june20[728:868] -2.973e-06  1.389e-06   -2.14   0.0341 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.584e-06 on 139 degrees of freedom
Multiple R-squared:  0.03189,	Adjusted R-squared:  0.02493 
F-statistic: 4.579 on 1 and 139 DF,  p-value: 0.03412
```

It is worth highlighting that June is not a particularly volatile month. This raises the question of whether these spikes were truly caused by the NSA leak or are just a product of the basket's natural long-run volatility; June's volatiltiy looks fairly tame in the grand scheme of things. This is illustrated in the plot below, which compares June (black) with three random months. If anything, June's volatility is lower than the random months. What this is suggests is that *(*even if) the NSA leak caused a surge in volatility, the change was so small that it is hardly worth mentioning.

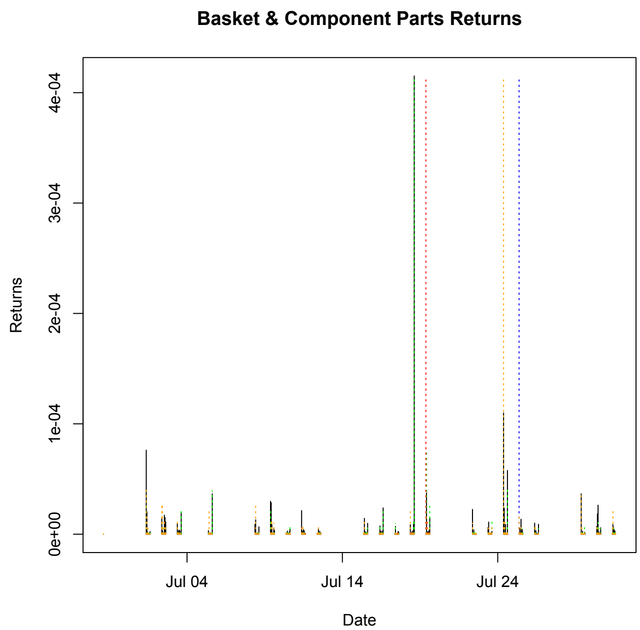


6. July indicator variables

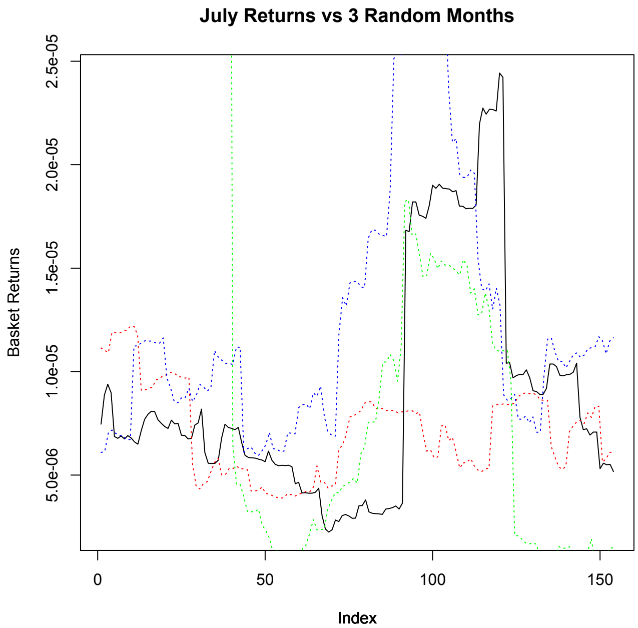
Next, we will consider July. A plot of squared July returns suggests a volatility spike sometime around July 18th. This is confirmed by a basic ARCH(1) model.



While this initially seeming promising, it is worth noting that all of our companies had earnings reports at approximately those dates. Historically, company earnings reports have a large impact on volatility; investors are eager to adjust their portfolios based on the release of new public information. The plot below shows the squared returns of our four companies (colored) and the basket squared returns. The spikes in the company volatities are likely casued by their respective earnings releases; these combine to see the volatility spike in the basket.



Finally, we can note that July volatility was not particularly large relative to other months in 2013. A plot comparing July volatility to the volatility in 3 random months is shown below. The results here are similar to those of June: even if there is a story to tell, it is not particularly significant in the grand scheme of things.


July Returns vs 3 Random Months
Basket Returns
5.0e-06
1.0e-05
1.5e-05
2.0e-05
2.5e-05
0
50
100
150
Index

7. To come

Since last time, we've made progress on an analysis of volume and are working on combining this with the rest of our results. Similarly, a news impact curve for June has been constructed, and we are considering how to best include this in our story.

Overall, we have concluded that there is little evidence that the NSA leaks had an effect on the volatility of the companies our basket. The only date left to seriously consider is June 20th.