STAT 771 Project - Draft 2

Steven Moen

Background and Motivation

My M.S. thesis at the University of Chicago developed a novel method to understand and measure value-at-risk, a commonly accepted method to measure downside risk. The metric is understood as follows: a one-day 1% VaR of -10 million dollars for a portfolio means that the portfolio will lose at least 10 million dollars of its value on the 1% worst trading days. A major advantage of VaR is that it distills a distribution of returns into one number. As such, VaR is often used in stress testing by regulatory agencies (Holton 2014).

There have been many popular approaches in the literature such as modeling the total distribution of returns (Longerstaey and Spencer 1996), and approaches using a semiparametric or a nonparametric historical simulation (Richardson, Boudoukh, and Whitelaw 2005). While modeling the entire distribution is likely too simplistic, Engle and Manganelli in a 2004 paper (Engle and Manganelli 2004) argue that nonparametric methods in the other camp are usually chosen for "empirical justifications rather than on sound statistical theory". To balance these approaches, they propose a framework called CAViaR that directly forecasts the VaR quantile using a conditional autoregressive quantile specification. This approach builds upon the statistical literature that extends linear quantile models to settings amenable to financial modeling, such as with heteroskedastic and nonstationary error distributions (Portnoy 1991).

My thesis extends the model beyond a univariate setting into a multivariate setting using the diffusion index model, originally developed by Stock and Watson for predicting conditional means (Stock and Watson 2002b, 2002a). My model uses exchange-traded funds (ETFs) as explanatory variables that are combined into principal component vectors at the forecast origin. Combining these principal component vectors with transformations of lagged autoregressive response variables produces similar predictive accuracy during periods of relatively low volatility (when compared to the CAViaR model) along with more insight into the drivers of the changes in the response variable.

Intended Work

As encouraging as the results were from my thesis, two important questions remained unanswered. The first is whether some sort of mixture model would be appropriate, that is, aiming to use the basket of ETFs during good times, and use the CAViaR ARMA specification during bad times. The approach of using ETFs allows a prediction based on forward-looking expectations of fundamental factors. Indeed, ETFs are just baskets of individual stocks or bonds, and those securities are (in theory) based on rational expectations about future resources, market conditions, etc - the microfoundations of what drives our economy. The ARMA specification, while practically and statistically sound, is contradicted by economic theory and practice - the weak form of the efficient market hypothesis states that it is impossible to forecast future values of asset prices using past values. But perhaps this view is incomplete.

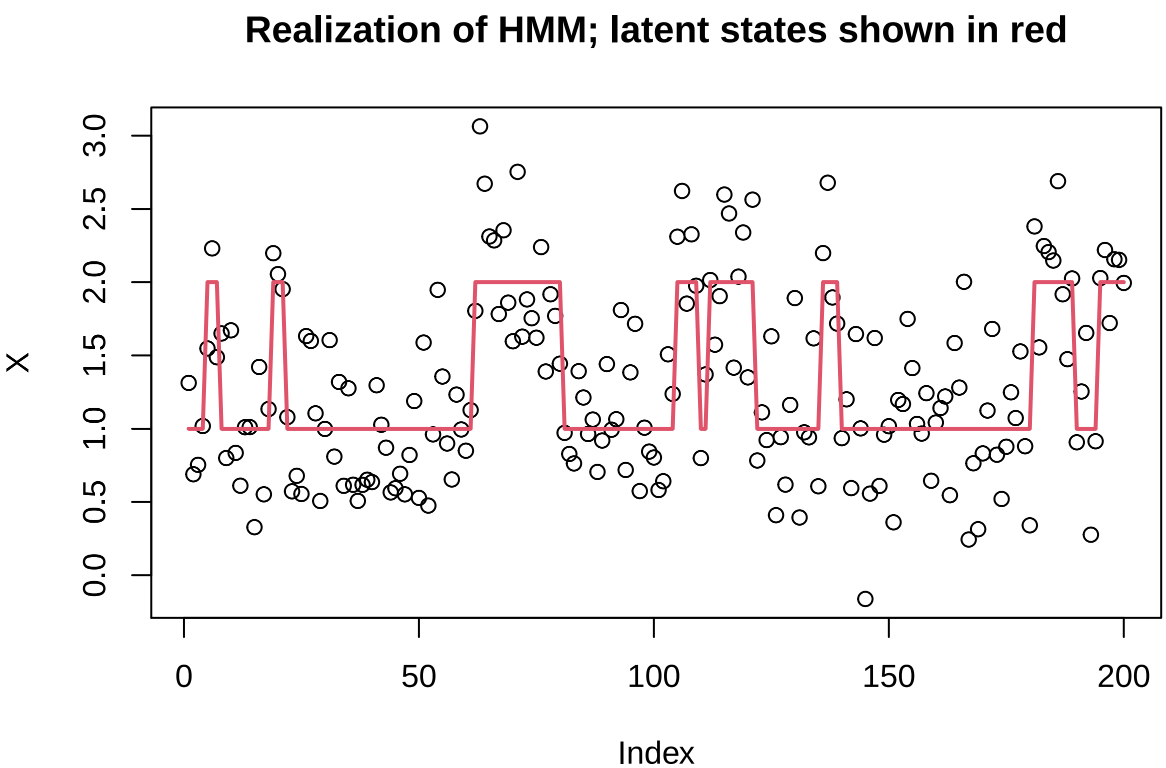
To combine these ideas, I would fit a Hidden Markov Model to infer the state of the world - the "rational" one, or the "irrational" one. Given the highly non-normal nature of financial data, I suspect there would be many interesting statistical and computational challenges that would arise with this approach. In addition, it is likely worth exploring alternative ensemble methods to further probe into the seemingly enigmatic nature that pervades financial time series.

The idea is based on a homework assignment I did for Professor Stephen's Dataset. See below. I'll replace the toy dataset with my results. Remember that the red line will not exist in my dataset since it represents the "true" model.

When I implement this, the "rational" state of the world will refer to the predictions of one of the multivariate CAViaR models where as the "irrational" state of the world will refer to one of the predictions from the univariate CAViaR models. One way to possibly implement this is to say that if the losses from the multivariate models are lower, then the HMM will lean towards the rational state of the world, otherwise, it will lean towards the other state.

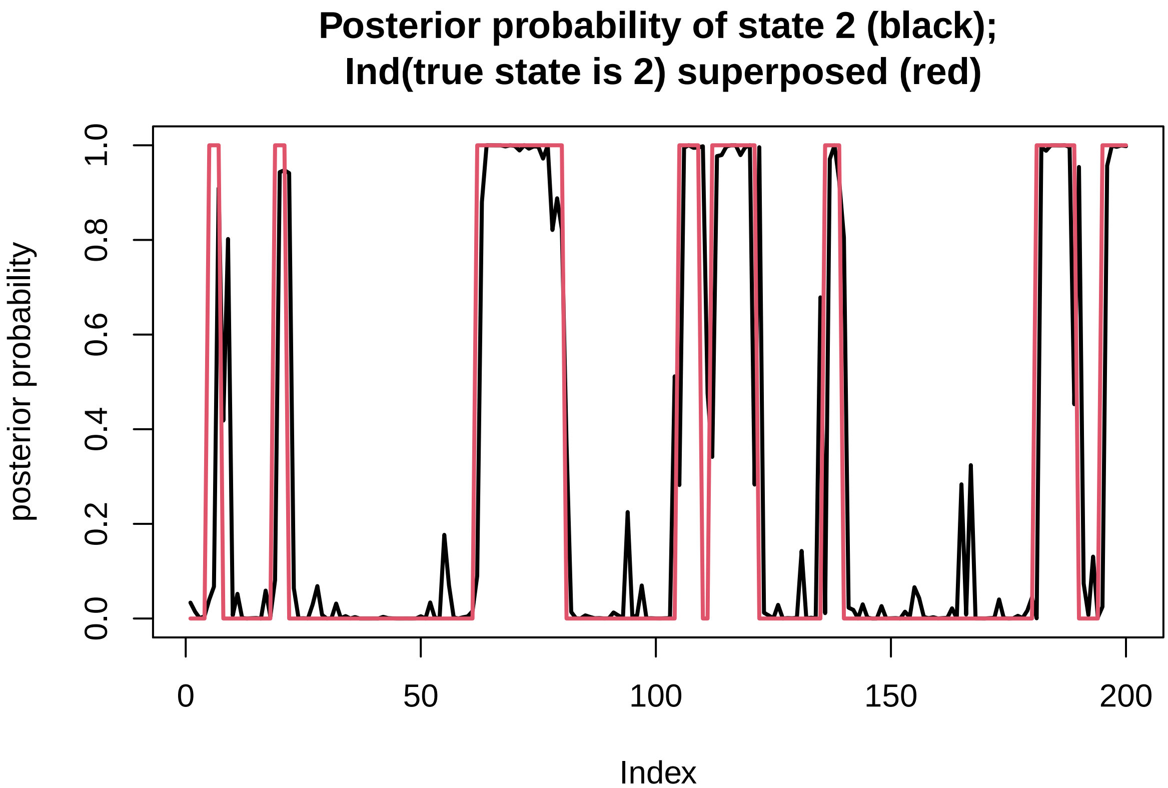
Example of HMM

Below are simulations from an HMM where there is a 90% probability of staying in the current state and a 10% chance of switching. Later, I use a forwards-backwards algorithm to estimate the probability of being in any state of the world.

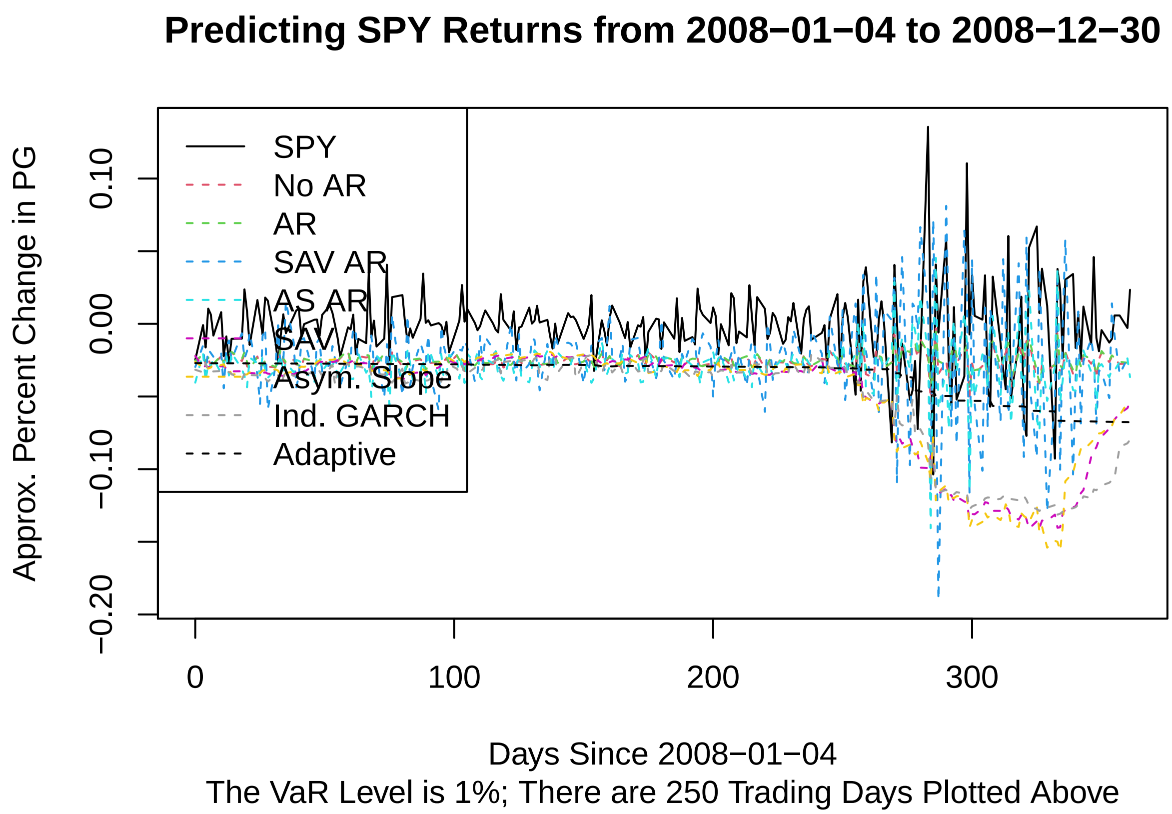


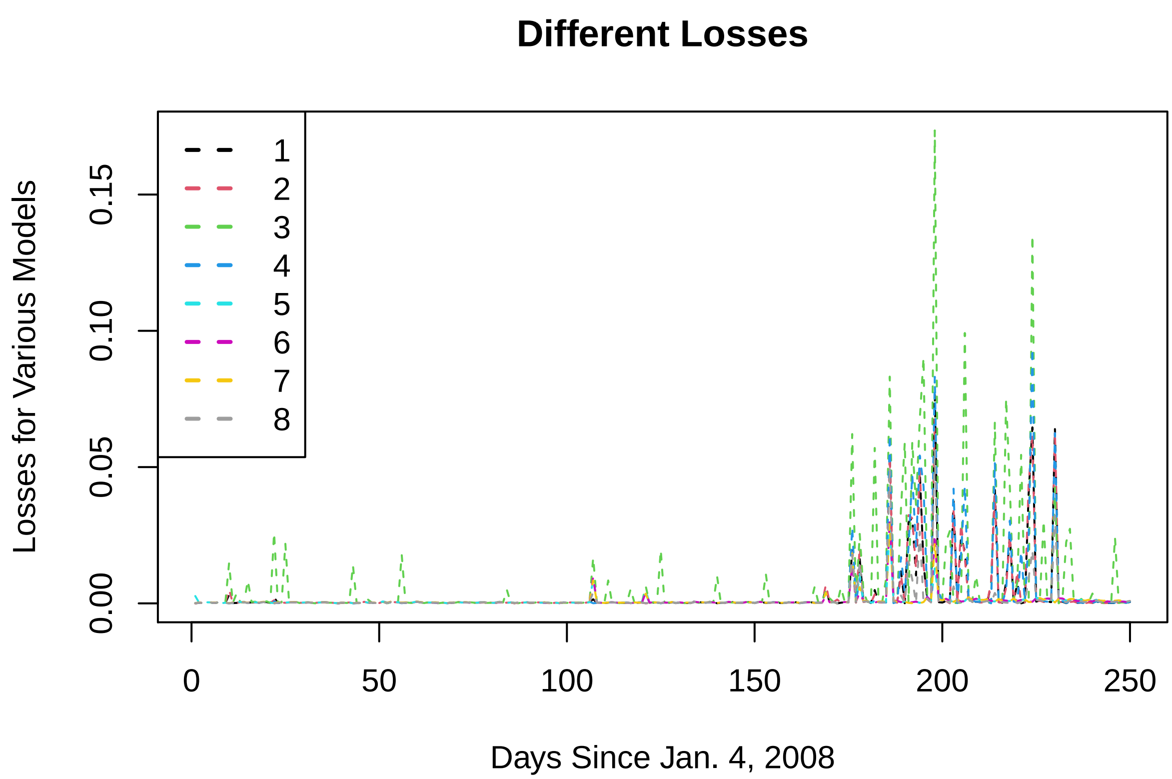

```
## [1] TRUE

## [1] 1.995051

## [1] 1 1

## [1] 0.04067454 0.09972794

## [1] 0.04519393 0.99727936

##              [,1]       [,2]
## [195,]         NA         NA
## [196,]         NA         NA
## [197,]         NA         NA
## [198,]         NA         NA
```

```
## [199,] 0.1404025 0.9020708
## [200,] 1.0000000 1.0000000

##               [,1]      [,2]
## [195,]          NA        NA
## [196,]          NA        NA
## [197,]          NA        NA
## [198,]          NA        NA
## [199,] 0.04519393 0.9972794
## [200,] 1.00000000 1.0000000
```



Loss Function Work on 10.27.2020

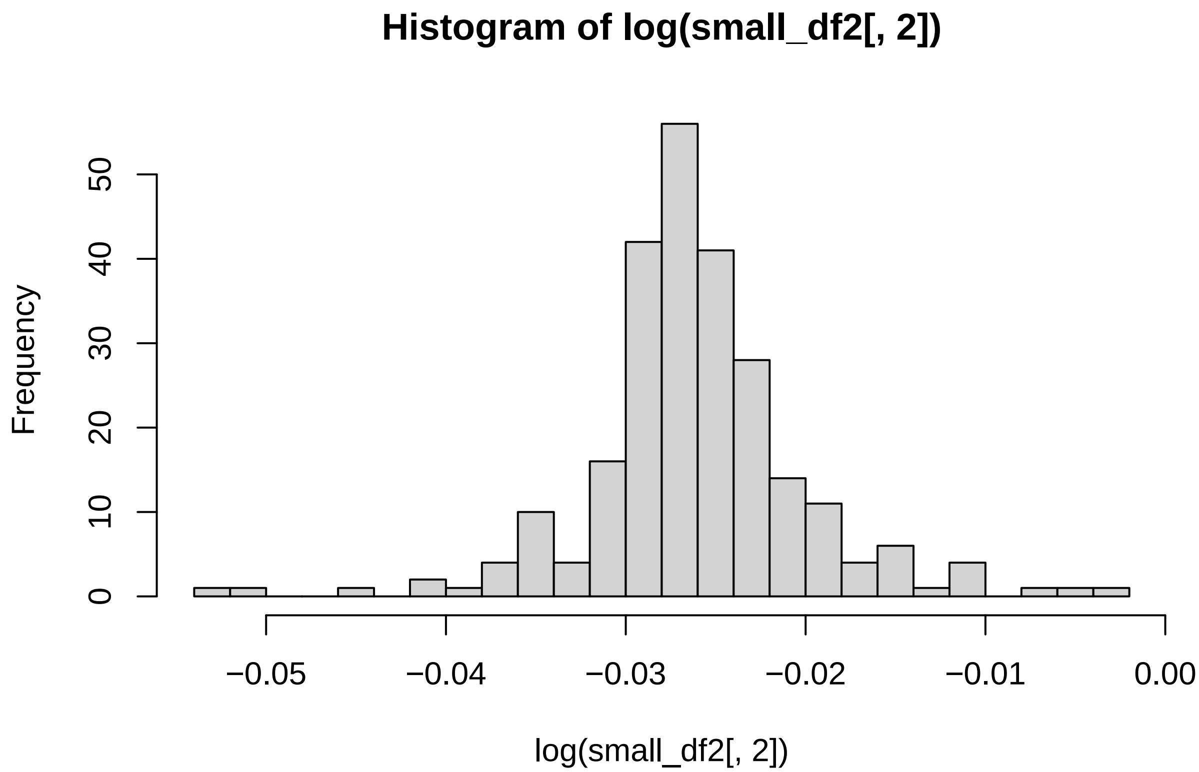


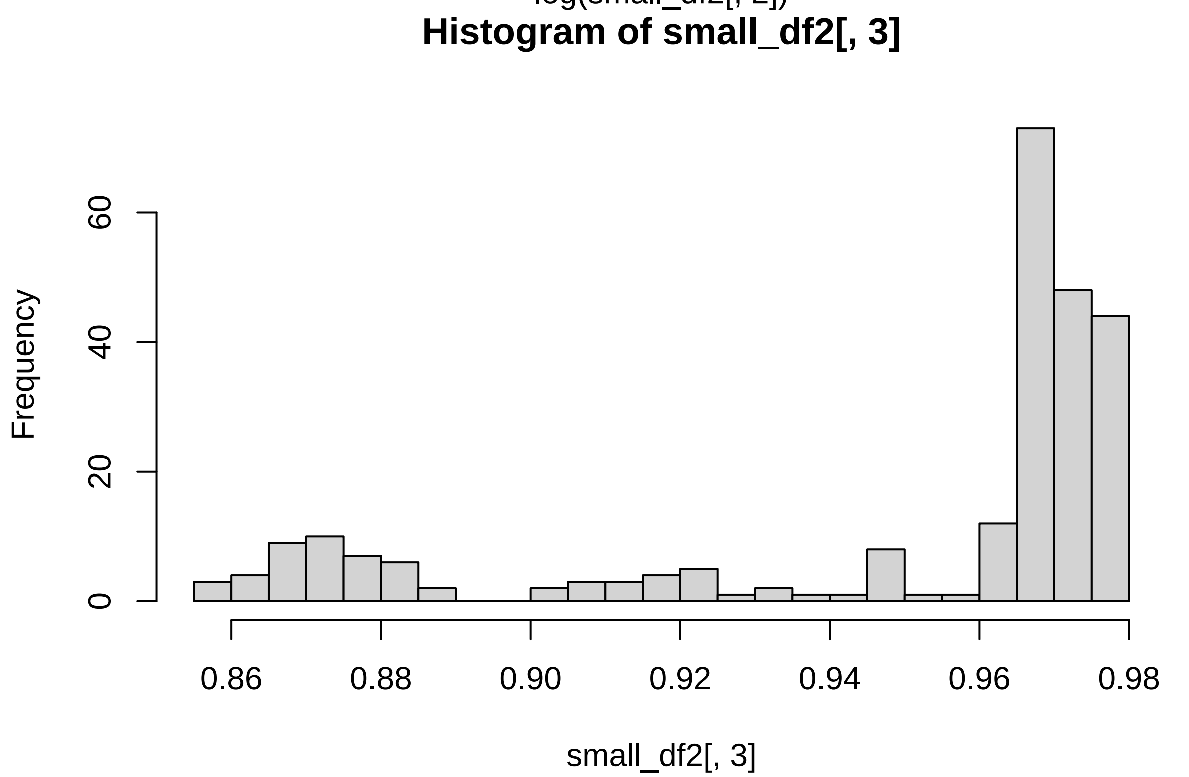


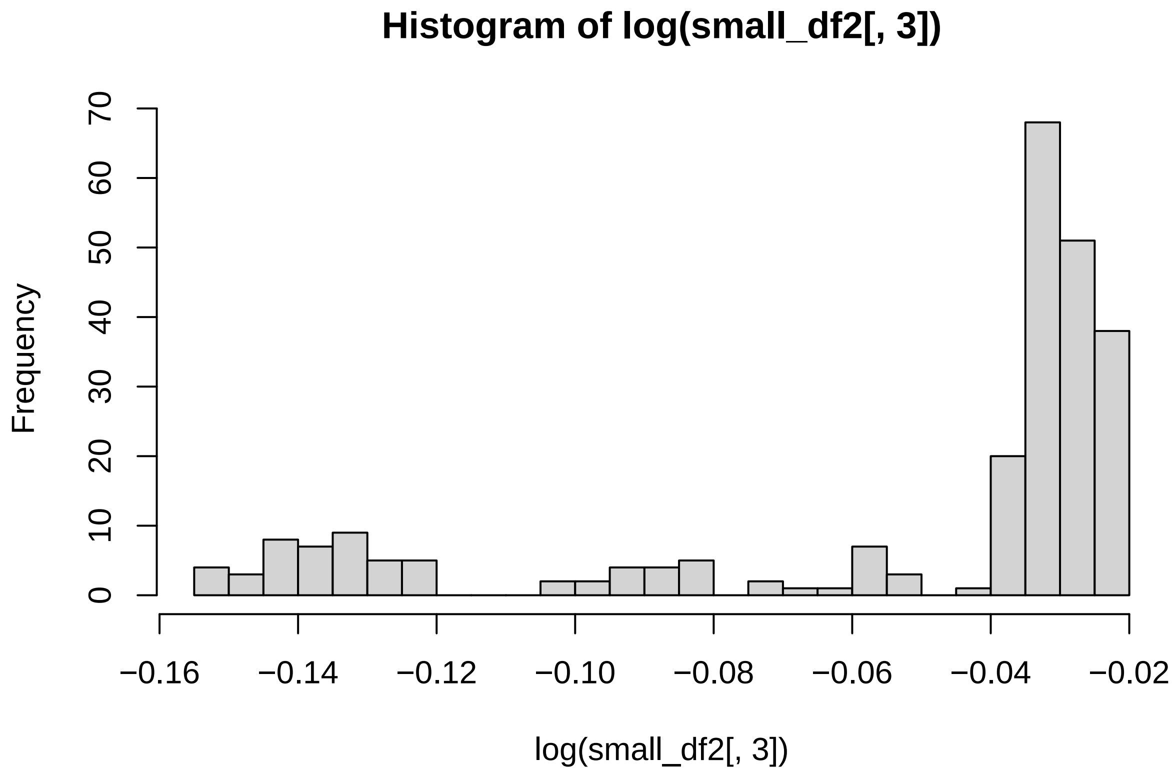

```
## [1] 0.7362329 0.7368474 1.7328787 0.8630283 0.2076144 0.2127849 0.2188360
```

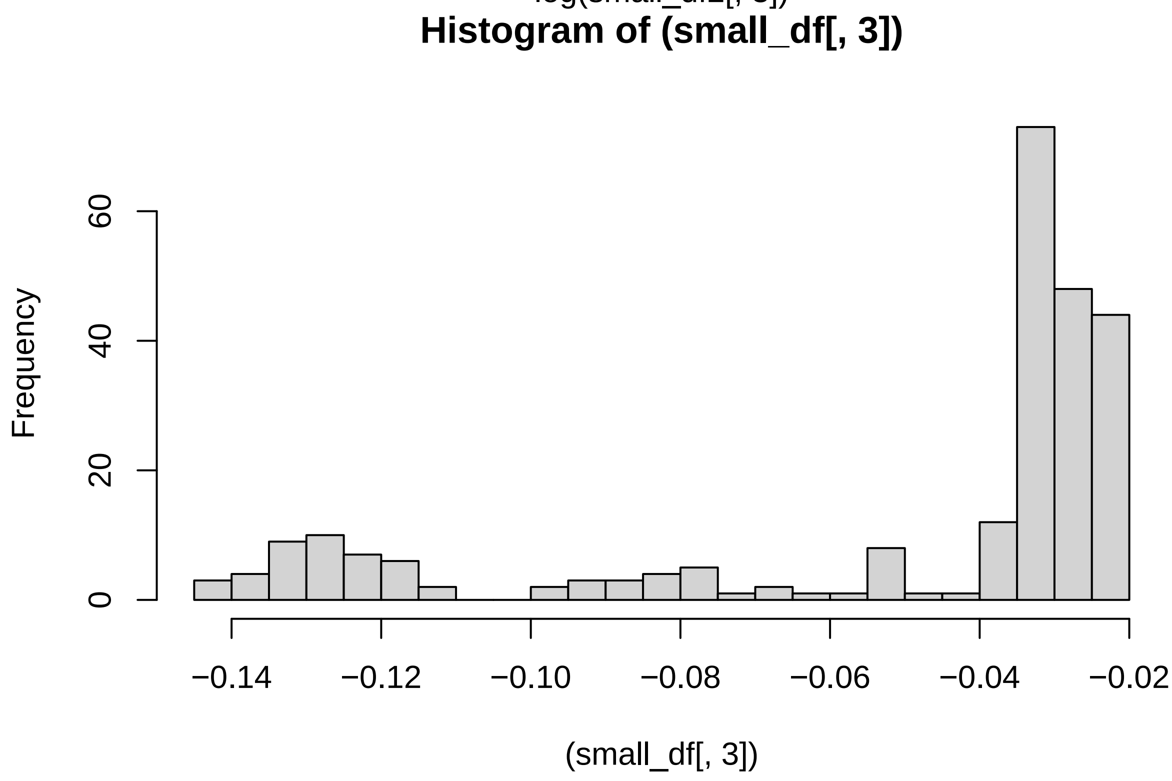
```
## [8] 0.3554456
```

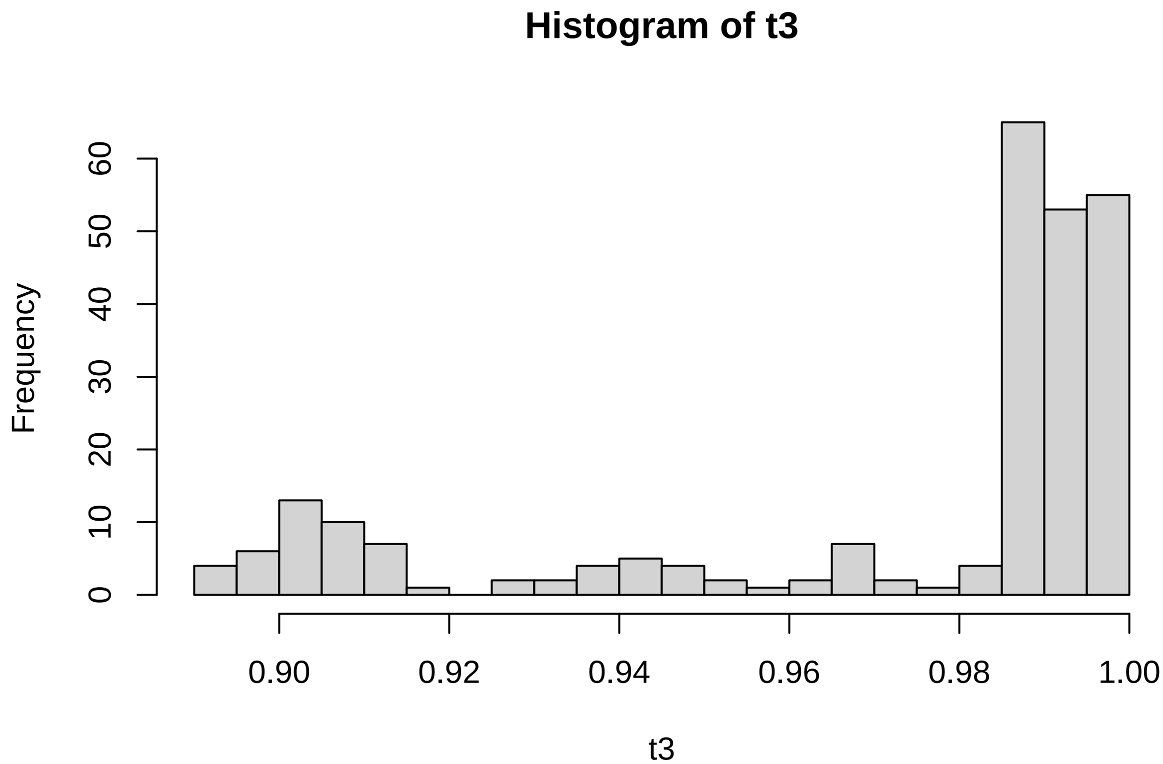
Based on the losses for each model during the last 250 trading days in 2008, it looks like the best options among the candidate model classes [LIST] are model 1 and model 5 specified below [SPECIFY]. Coincidentally, these are the simplest models available. A natural criticism of this approach is that the losses are lower for the CAViaR specifications without lagged predictors. This is a fair point, however, the period of 2008 is a period of extreme crisis, and a simpler, ARMA-style model might seem to work better.


Histogram of log(small_df2[, 2])
Frequency
0
10
20
30
40
50
−0.05
−0.04
−0.03
−0.02
−0.01
0.00
log(small_df2[, 2])


Histogram of small_df2[, 3]
Frequency
0
20
40
60
0.86
0.88
0.90
0.92
0.94
0.96
0.98
small_df2[, 3]

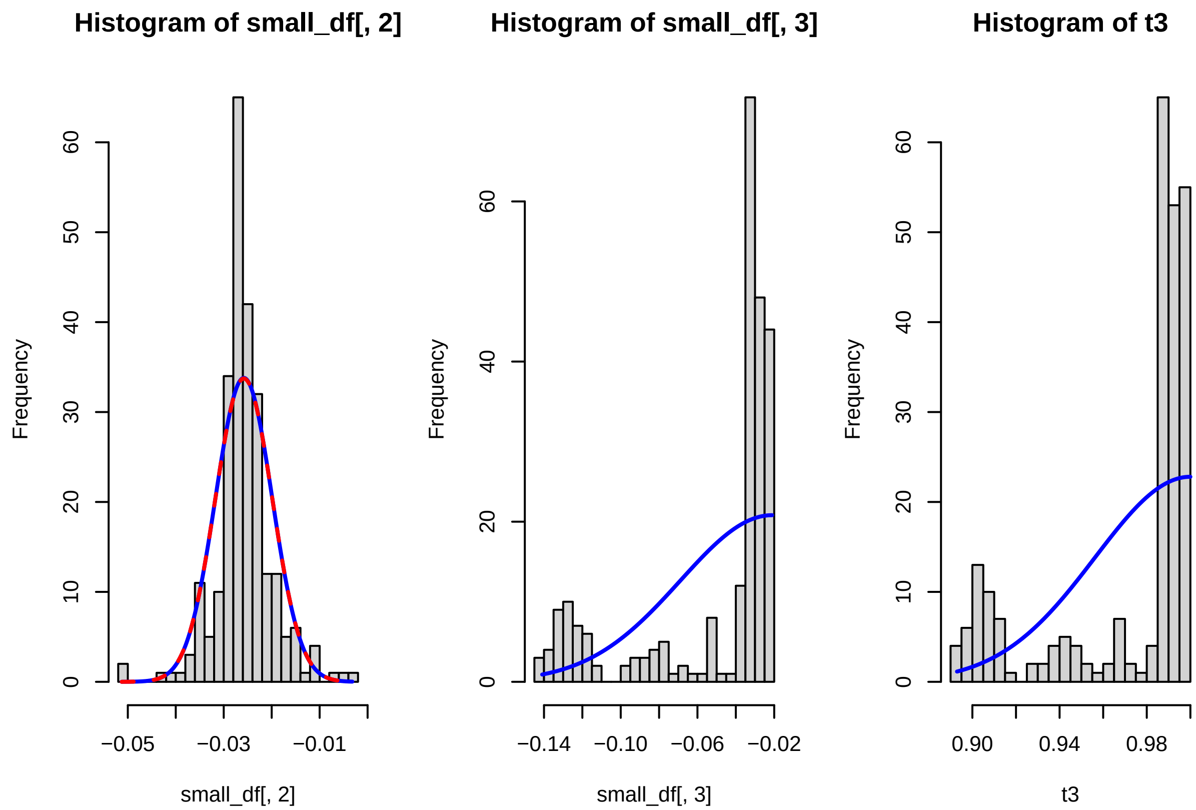

Histogram of log(small_df2[, 3])
Frequency
0
10
20
30
40
50
60
70
−0.16
−0.14
−0.12
−0.10
−0.08
−0.06
−0.04
−0.02
log(small_df2[, 3])


Histogram of (small_df[, 3])
Frequency
0
20
40
60
−0.14
−0.12
−0.10
−0.08
−0.06
−0.04
−0.02
(small_df[, 3])



It looks like we're in a bit of trouble. Let's see what happens.

```
## $par
##        mean          sd          xi
## 0.965101377 0.026367066 0.004662492
##
## $objective
## [1] -601.3692
##
## $convergence
## [1] 1
##
## $iterations
## [1] 38
##
## $evaluations
## function gradient
##       83      119
##
## $message
## [1] "false convergence (8)"
```



It looks like the normal approximation is good enough for us for the first model.

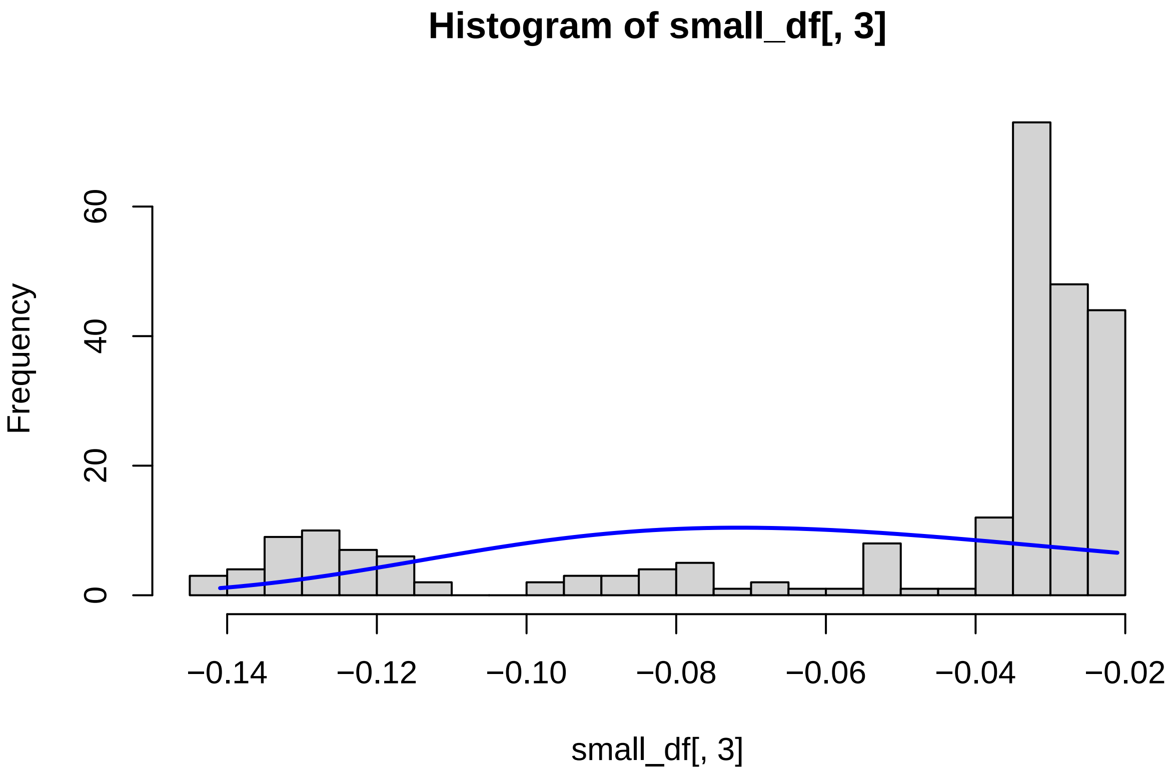
We will also try fitting a Gumbel, Fréchet, and a Weibull distribution to the data.

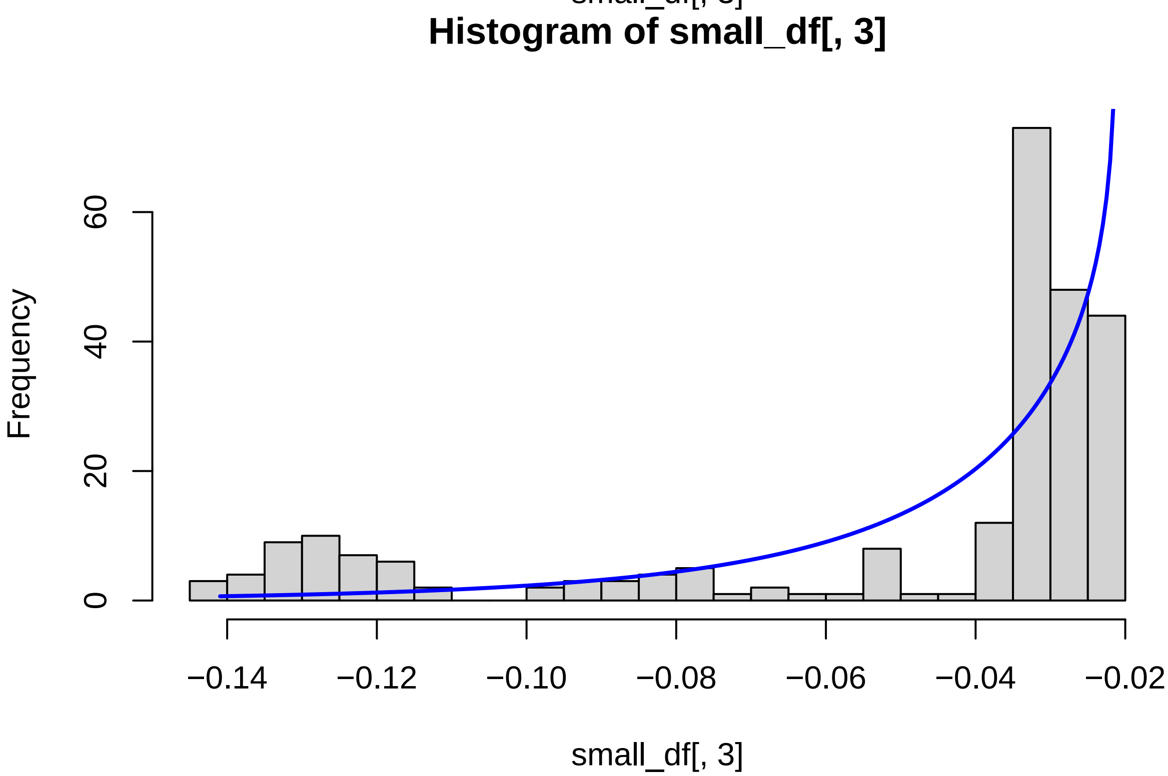
```
## $conv
## [1] 0
##
## $nllh
## [1] -411.77
##
## $mle
## [1] -0.07157664  0.04409873
##
## $se
## [1] 0.002975693 0.001966694
```

```
## [1] -0.07157664  0.04409873
```

```
## $conv
## [1] 0
##
## $nllh
## [1] -647.1938
##
## $mle
## [1] -0.04463765  0.02769730 -1.17441110
##
## $se
## [1] 1.999319e-06 1.999325e-06 8.703099e-04
```

```
## [1] -0.04463765  0.02769730 -1.17441110
```





Now let's compare the K-L divergences.

Table 1: Comparing K-L Divergences By Model Fits

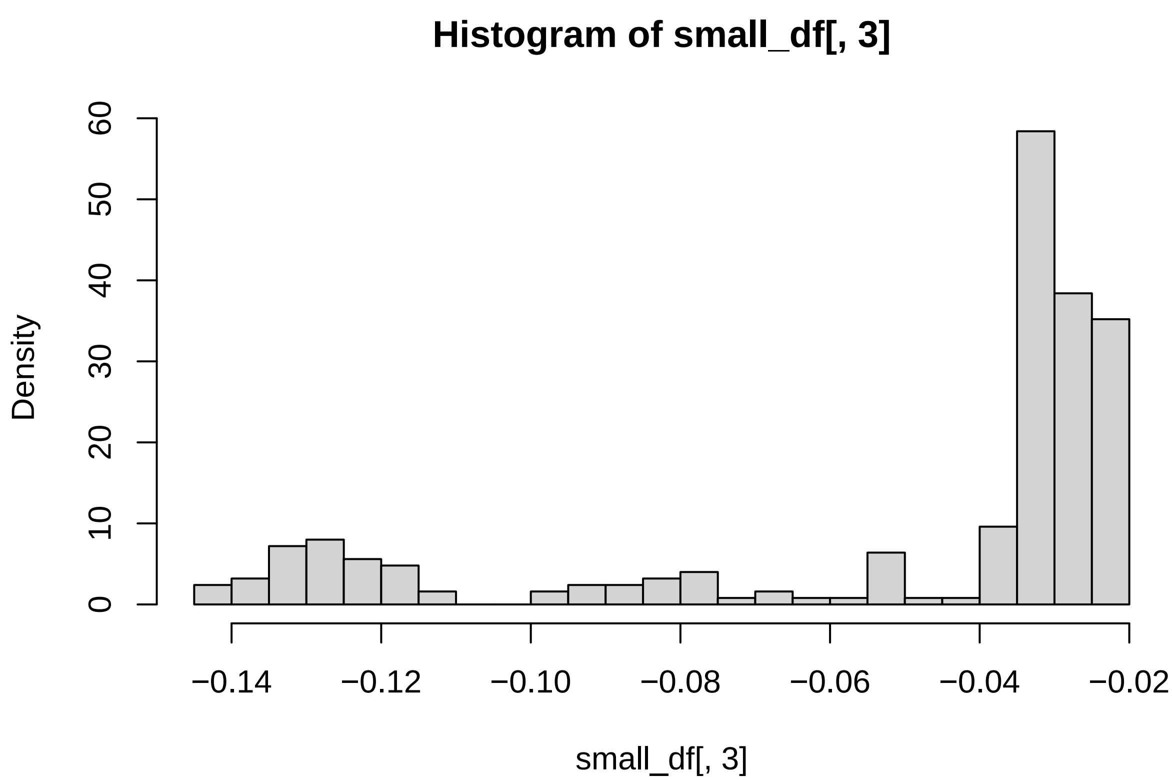
	Skew-Normal	Gumbel	GEV
Mean Sum K-L Divergence	22.52	25.4	17.71

Note:
The Parameters for Each of these Models Were Estimating Using Maximum Likelihood

Table 2: Optimal Parameters for the Normal Distribution

	Mean	Standard Deviation
Optimal Parameters	-0.03	0.01

Note:
The Mean Was Estimated Using the Sample Mean, SD was Estimated Using Sample Std. Dev.




```
##  [1]  2.4  3.2  7.2  8.0  5.6  4.8  1.6  0.0  0.0  1.6  2.4  2.4  3.2  4.0  0.8
## [16]  1.6  0.8  0.8  6.4  0.8  0.8  9.6 58.4 38.4 35.2
```

```
##         mean           sd           xi
## -0.059291805  0.028889558  0.002408128
```

Based on these results, we will use the generalized extreme value distribution with the following parameters.

```
##         mean           sd           xi
## -0.059291805  0.028889558  0.002408128
```

```
## [1] -0.07157664  0.04409873
```

Now we will use this in our modeling below.

Table 3: Optimal Parameters for the Skew Normal Distribution

	Location	Scale	Shape
Optimal Parameters	-0.0593	0.0289	0.0024

Note:
Estimated Using Maximum Likelihood

Table 4: Optimal Parameters for the Gumbel Distribution

	Location	Scale
Optimal Parameters	-0.0716	0.0441

Note:
Estimated Using Maximum Likelihood

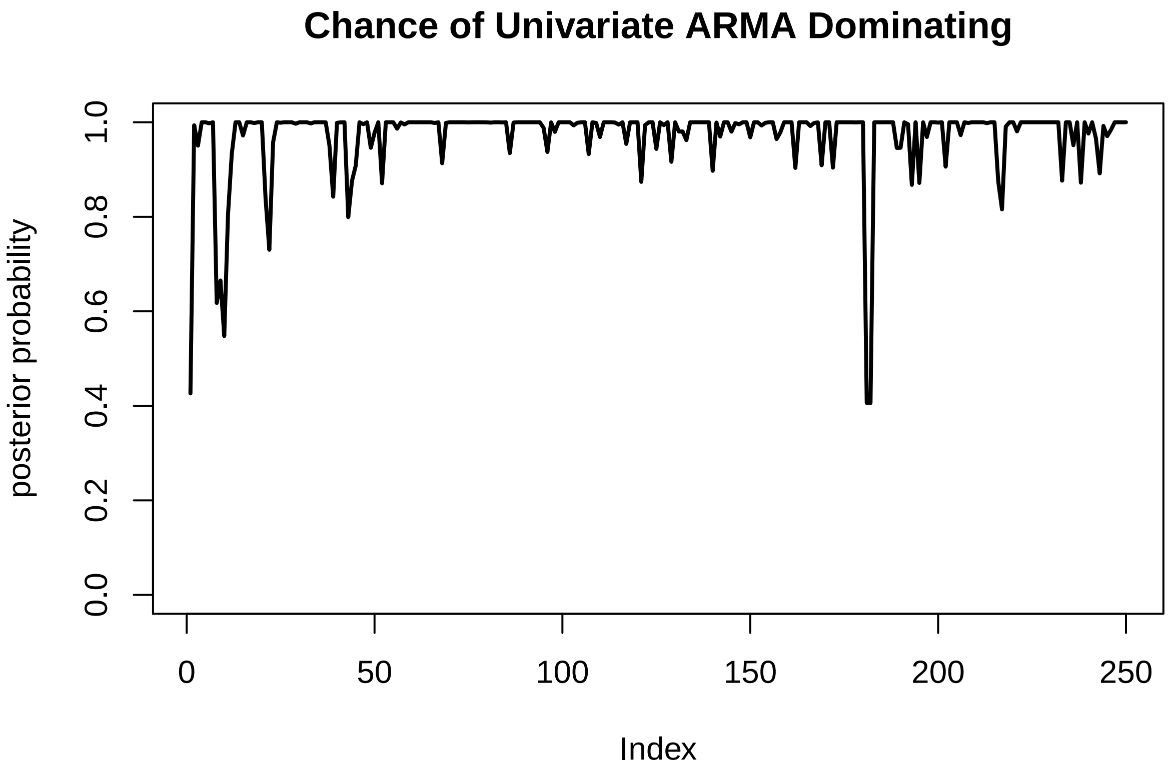
HMM Problem Solving

```
##          [,1]     [,2]
## [1,] 33.25742 2.777143
## [2,]       NA       NA
## [3,]       NA       NA
## [4,]       NA       NA
## [5,]       NA       NA
## [6,]       NA       NA
```

Table 5: Optimal Parameters for the Generalized Extreme Value Distribution

	Location	Scale	Shape
Optimal Parameters	-0.04	0.03	-1.17

Note:
Estimated Using Maximum Likelihood



For future work, it might be worth exploring what happens if the HMM were fit to all 8 candidate models.

Changepoint Detection

The second question is to understand shifts in the economy using a changepoint detection algorithm:

1. Using a set of ETFs, perform Principal Component Analysis at T many points for M many factors - $f_{m,t}$
2. At each time point, add the vectors together to get a resultant: $\sum_{m=1}^{M} f_{m,t} = r_t$, giving $r_1, r_2, ..., r_T$.
3. Starting with an arbitrary reference point t_0 with associated r_0 resultant, measure the angle between resultants calculated at different time steps r_t

$$\theta_t = \arccos\left(\frac{r_0 \cdot r_t}{||r_0||||r_t||}\right)$$

The angle θ could be plotted over time, and changepoints could be detected using Monte Carlo simulation, because PCA transformations are non-linear, so calculating an analytical density from the transformed data is intractable. Moreover, the data fed into the PCA transformation is non-normal, which further supports the notion of using Monte Carlo simulation to establish reasonable estimates of uncertainty for detected changepoints. As with the first line of reasoning, there would certainly be interesting challenges, particularly in creating crisp null and alternative hypotheses.

Data Used

The response variable used in this analysis is SPY, which is an exchange-traded fund that aims to track the performance of the S&P 500, which is discussed above. It is broadly used as a bellwether of the U.S. economy, and has the advantage of avoiding survivorship bias - while an individual stock might go bankrupt or merge with another, it is reasonable to assume that these issues do not apply with an ETF.

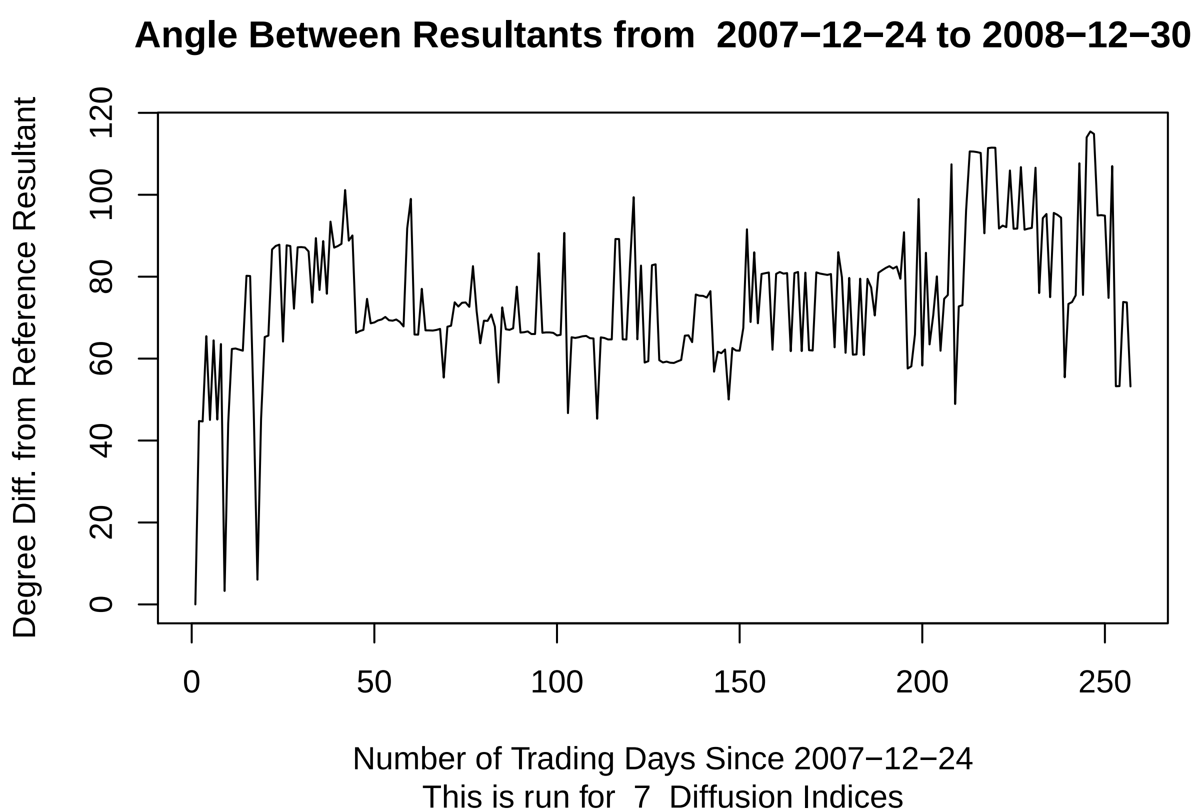
Following this logic, there are several classes of response variables used in this analysis. The first group is a set of U.S. sector ETFs obtained from Seeking Alpha (NA 2020). As with the response variable, these ETFs were publicly traded throughout the Great Recession of 2008.

a. Utilities (XLU)
b. Consumer Staples (XLP)
c. Healthcare (XLV)
d. Technology (XLK)
e. Consumer Discretionary (XLY)
f. Industrial (XLI)
g. Financial Services (XLF)
h. Basic Materials (XLB)
i. Energy (XLE)

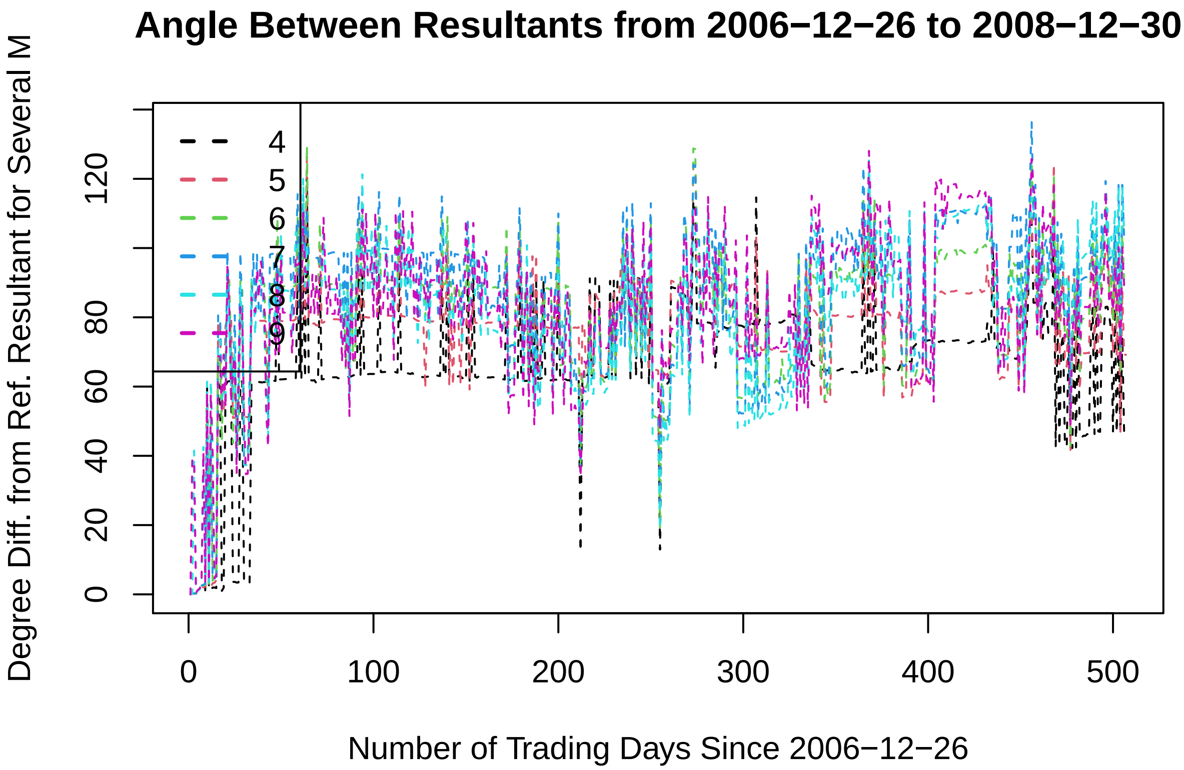
The second group for this analysis is bond ETFs. Like the previous two groups, these ETFs potentially contain forward-looking information about the stock market. These ETFs were chosen because they were the first fixed-income ETFs available in the United States, and had enough history for this paper (NA 2017).

a. iShares 1-3 Year Treasury Bond Fund (SHY)
b. iShares 7-10 Year Treasury Bond Fund (IEF)
c. iShares 20+ Year Treasury Bond Fund (TLT)
d. iShares iBoxx $ Investment Grade Corporate Bond ETF (LQD)

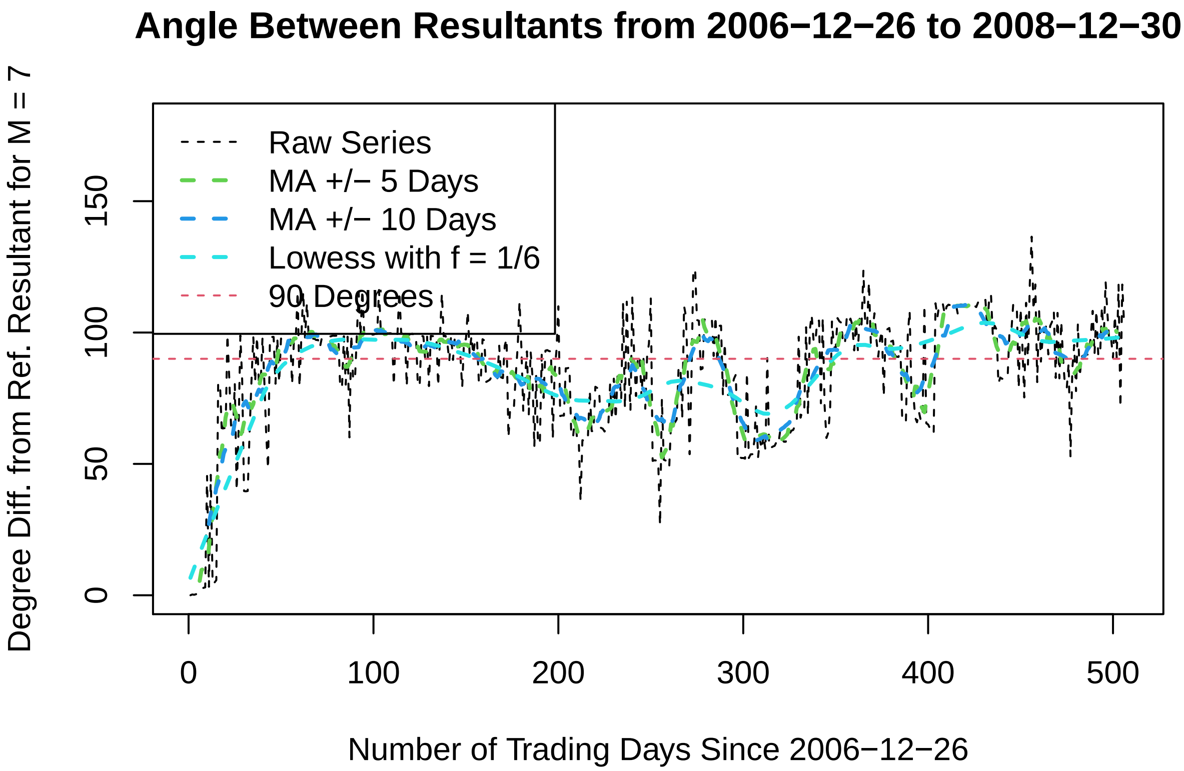
Results



U.S. ETFs Results for 2007 - 2008



A natural criticism of the above fits are that the data is noisy. A complication of this analysis is fact that there isn't necessarily anything to "pin" the data do, because the problem is unsupervised. As such, I think a decent way to picking the wheat from the chaff (or the signal from the noise) is to apply some smoothing filters to the above data. Two options to do so are the moving average smoother and the Lowess smoothers.



In the above plot, I applied a few filters, namely the moving average filter with equal weights for plus/minus 5 days and 10 days as well as the Lowess with a weight of $f = 1/6$ for higher levels of precision. Below are the specifications for moving average with 5 and 10 days, where x_t is the angle between the resultants.

$$m_{t,5} = \sum_{j=-5}^{5} \frac{1}{11} x_{t-j}$$

$$m_{t,10} = \sum_{j=-10}^{10} \frac{1}{21} x_{t-j}$$

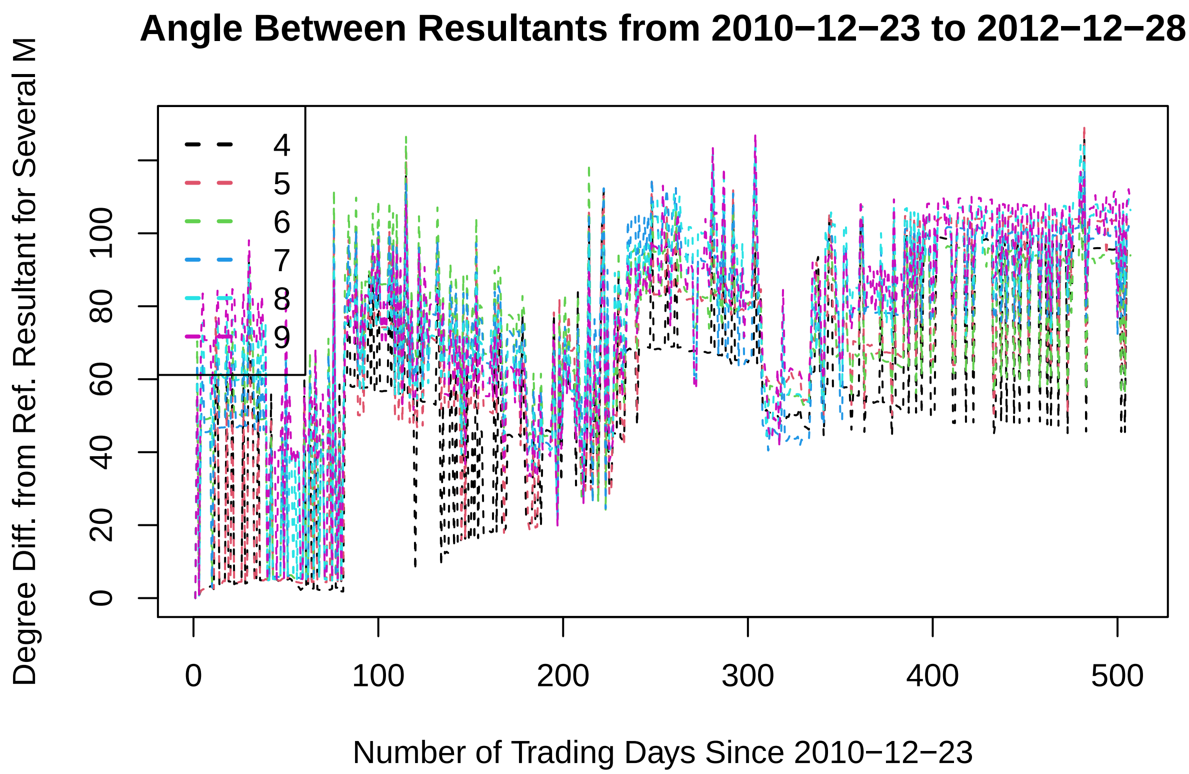
The Lowess smoother is a smoother that per Shumway and Stoffer (Shumway and Stoffer 2016), is a technique " based on k-nearest neighbors regression, where in one uses only the data $[x_{t-k/2}, ..., x_t, ..., x_{t+k/2}]$ to predict the true value of x_t. Based on a visual inspection of the data, it appears that a more precise estimator was in order, so the Lowess function only uses 1/6th of the data.

There are some interesting trends that bear discussion. While some of macroeconomics may appear to be little more than a crystal ball, or a science based upon strong priors, I don't believe that to be the case here. In the above graph is that the angle between resultant vectors for baskets of equity exchange-traded funds (ETFs) for different sectors of the U.S. economy point, almost as a leading indicator, towards the once-in-a-lifetime tumult that was about the grip the U.S. and global financial markets. Indeed, on the graph in the report, the angle of the resultant seems to stabilize for the month of August 2008, which corresponds roughly to the $400 - 425$ trading days since the reference point at the end of 2006. This is before Lehman Brothers declared bankruptcy, before AIG was bailed out, and before Congress passed TARP. Now it remains to be seen whether this is coincidental or whether there really is a leading indicator here, but I believe the question is worth asking and worth exploring further.

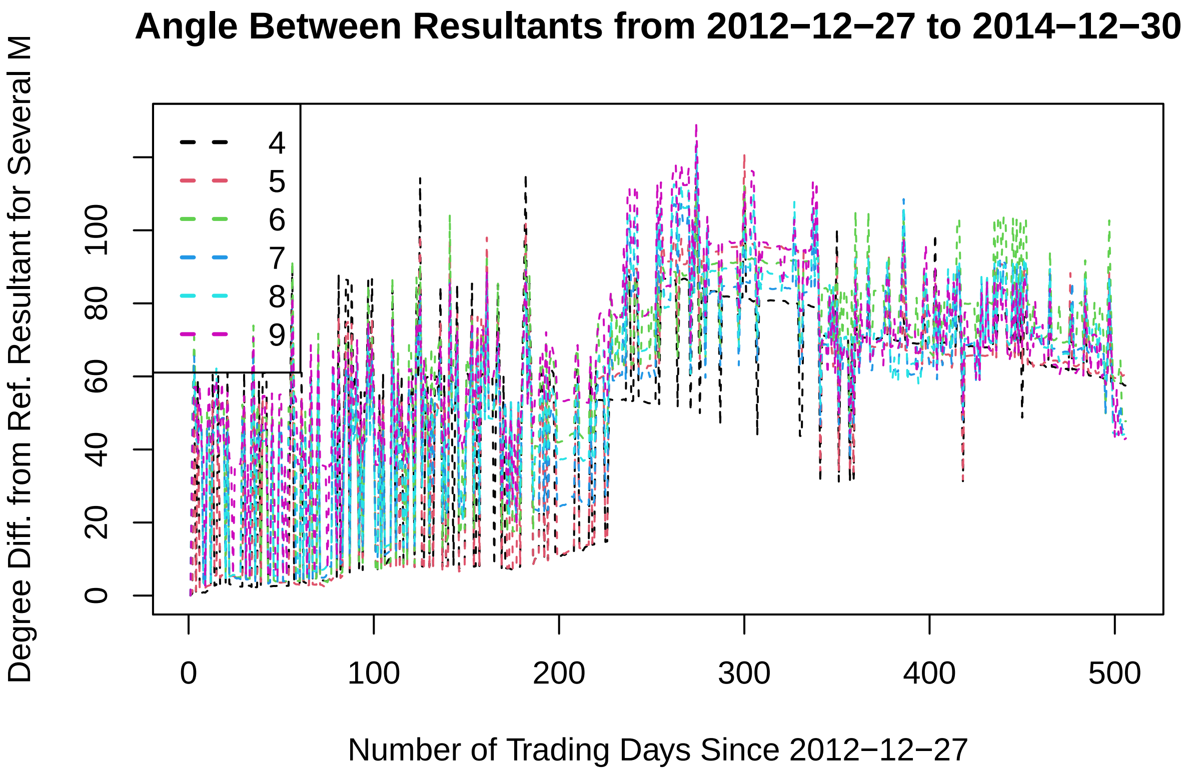
From a statistical perspective, the most important fact about these trends is that the algorithm that produces them, namely PCA, is *an unsupervised learning algorithm.* At no point in this work is there a model of any kind trying to predict gyrations in the S&P500 or the Dow Jones or the Real Economy. Indeed, remember that these principal components are based upon sector ETFs, things like Energy, Consumer Staples, and Utilities, which in turn are based upon individual stocks - companies like Exxon, Capital One, General Electric, and Amazon. These stock prices may seemingly be enigmatic and noisy, but ironically they look this way because according to economic theory, they contain all the relevant information about their company as determined by market participants (Malkiel 2003). What if, in all of their foresight, they saw the most cataclysmic economic event of our time before it happened, without even aiming to do so?

Appendix

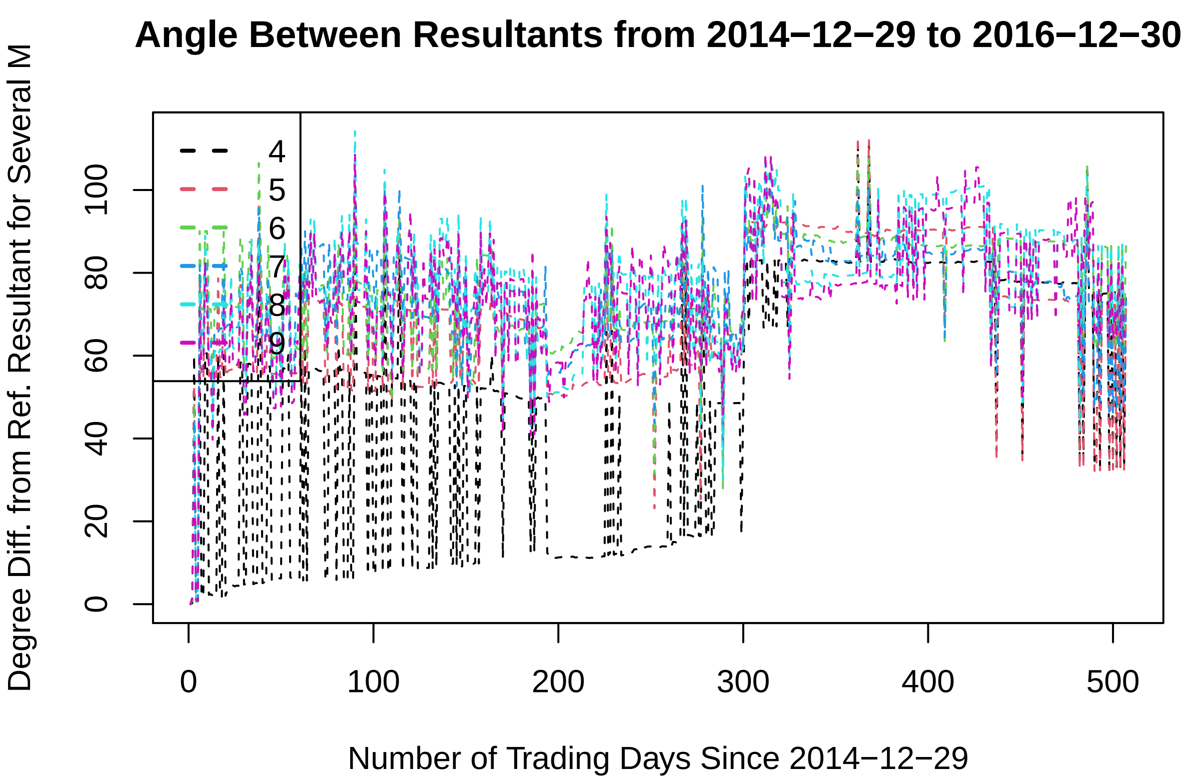
U.S. ETFs Results for the 2011 - 2012



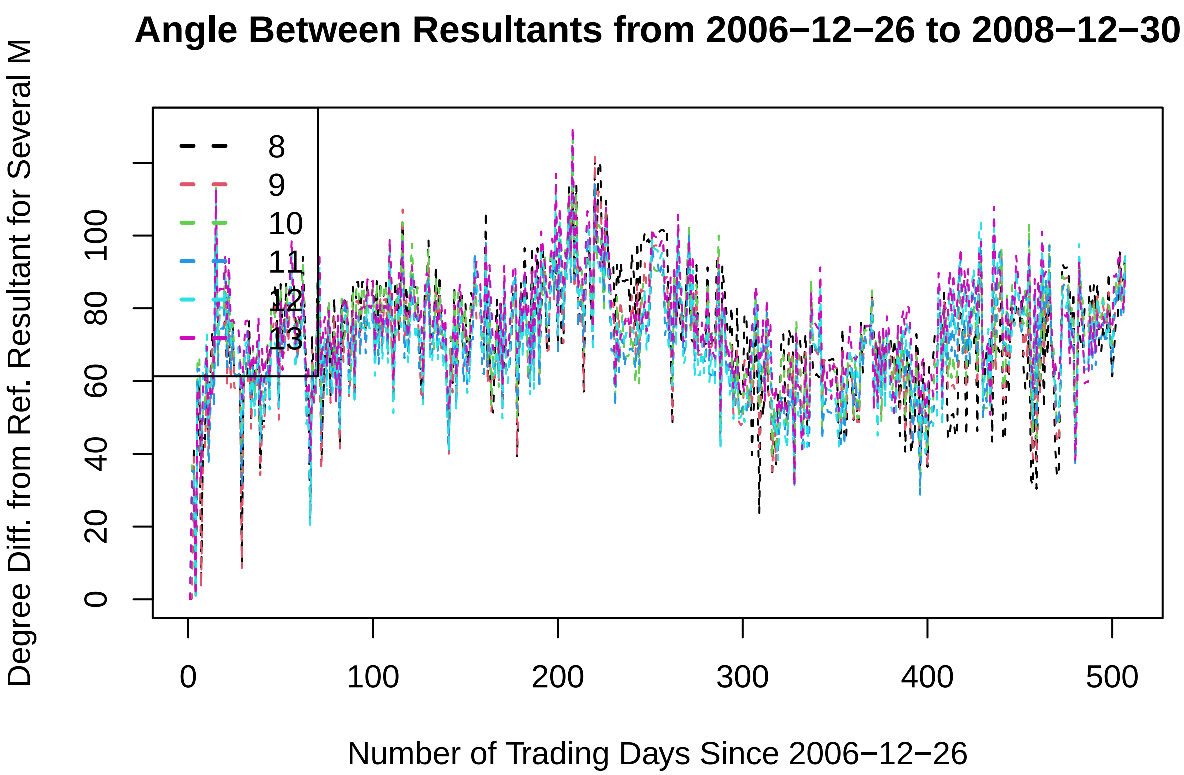
U.S. ETFs Results for the 2013 - 2014



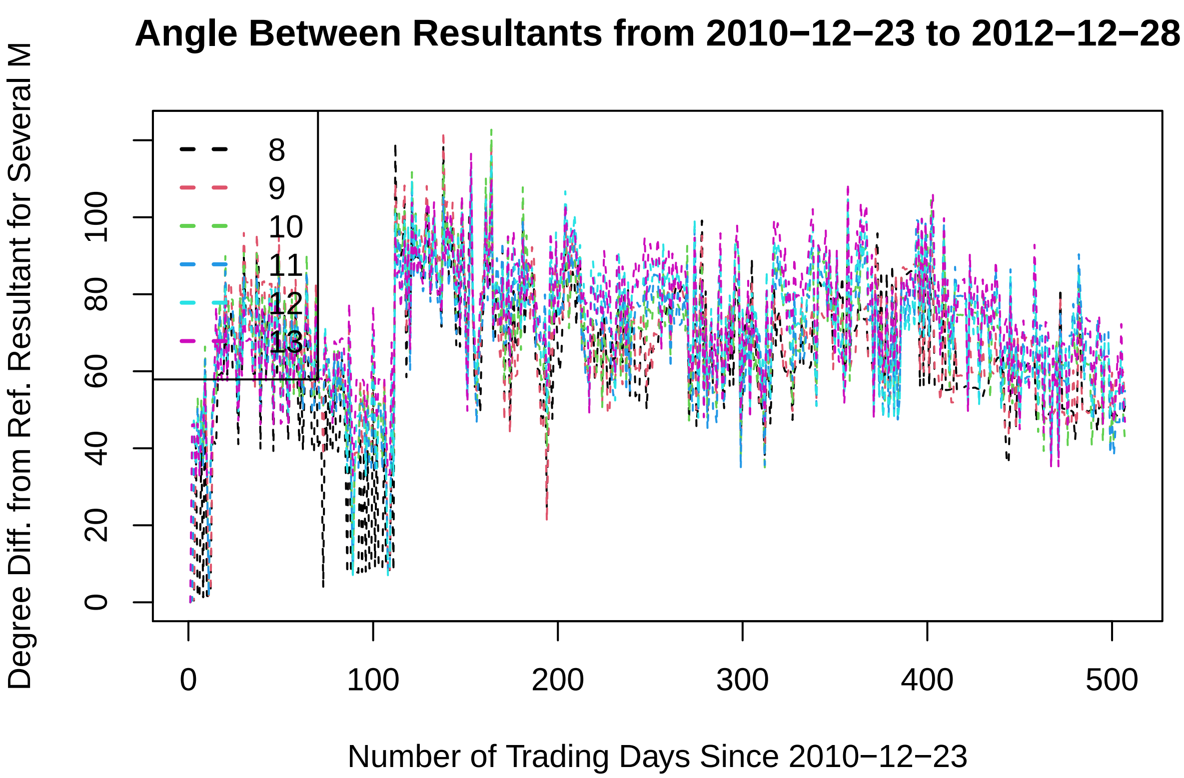
U.S. ETFs Results for the 2015 - 2016



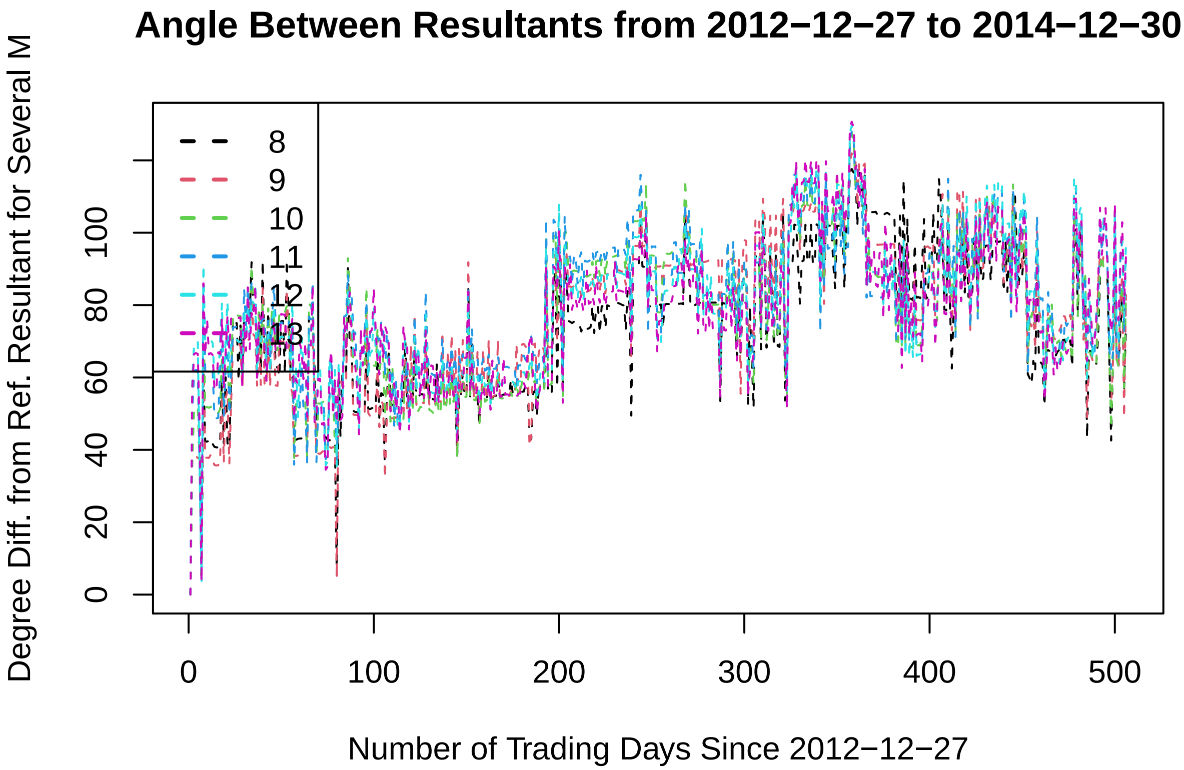
U.S. + Bond ETF Results for 2007 - 2008



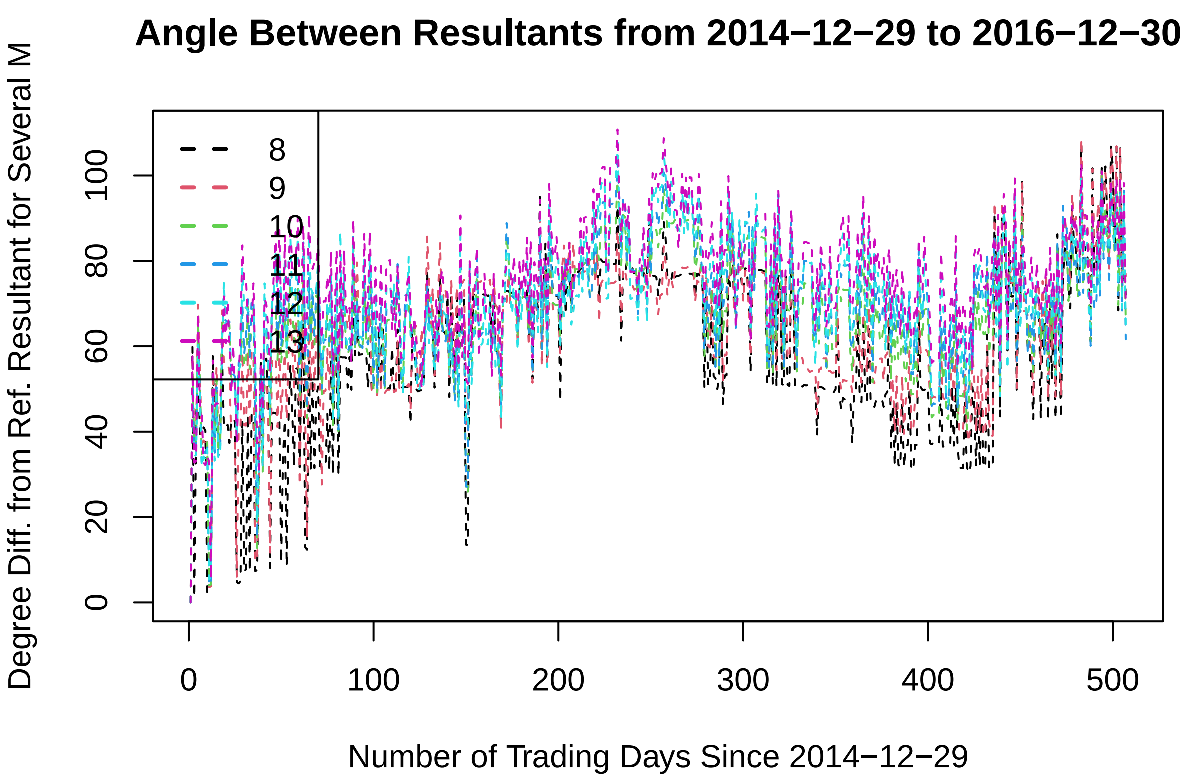
U.S. + Bond ETF Results for the 2011 - 2012



U.S. + Bond ETF Results for the 2013 - 2014



U.S. + Bond ETF Results for the 2015 - 2016



Code

The code can be found at the location listed below in the "STAT_771_Class_Project.Rmd" file.

https://github.com/stevenmoen/stat_771_final_project

Big HMM Function

Literature Cited

Engle, Robert F, and Simone Manganelli. 2004. "CAViaR." *Journal of Business & Economic Statistics* 22 (4). Taylor & Francis: 367–81. https://doi.org/10.1198/073500104000000370.

Holton, Glyn A. 2014. "History of VaR - Value-at-Risk: Theory and Practice." https://www.value-at-risk.net/history-of-value-at-risk/.

Longerstaey, Jacques, and Martin Spencer. 1996. "RiskMetrics - Technical Document." J.P. Morgan/Reuters. http://www.jpmorgan.com/RiskManagement/RiskMetrics/RiskMetrics.html.

Malkiel, Burton Gordon. 2003. *A random walk down Wall Street : the time-tested strategy for successful investing*. New York: W.W. Norton.

NA. 2017. "iShares Institutional Guide to Bond ETFs." https://www.complianceweek.com/download?ac=5780.

———. 2020. "Sector ETFs | Seeking Alpha." https://seekingalpha.com/etfs-and-funds/etf-tables/sectors.

Portnoy, Stephen. 1991. "Asymptotic behavior of regression quantiles in non-stationary, dependent cases." *Journal of Multivariate Analysis* 38 (1): 100–113. https://doi.org/https://doi.org/10.1016/0047-259X(91)90034-Y.

Richardson, Matthew P., Jacob (Kobi) Boudoukh, and Robert F Whitelaw. 2005. "The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.51420.

Shumway, Robert, and David Stoffer. 2016. *Time Series Analysis and Its Applications With R Examples Fourth Edition*. Fourth. Springer. https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf.

Stock, James H, and Mark W Watson. 2002a. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics* 20 (2). Taylor & Francis: 147–62. https://doi.org/10.1198/073500102317351921.

———. 2002b. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97 (460). [American Statistical Association, Taylor & Francis, Ltd.]: 1167–79. http://www.jstor.org/stable/3085839.