

Mini Research Project

HEYANG CAO

2018/6/20

Introduction

A individual return can be attributed to firm-specific characteristics, like financial status, management team capability, etc., and market factors such as market betas, value, growth, etc..

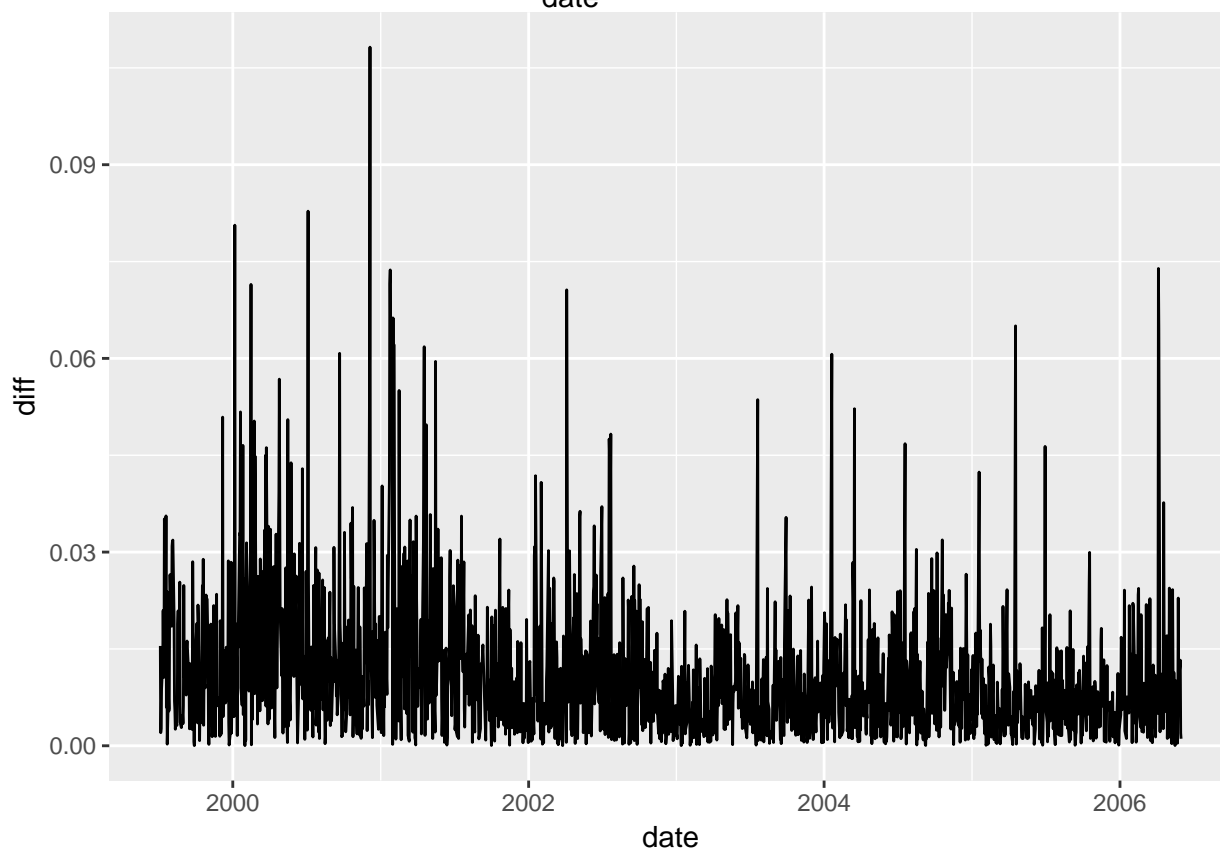
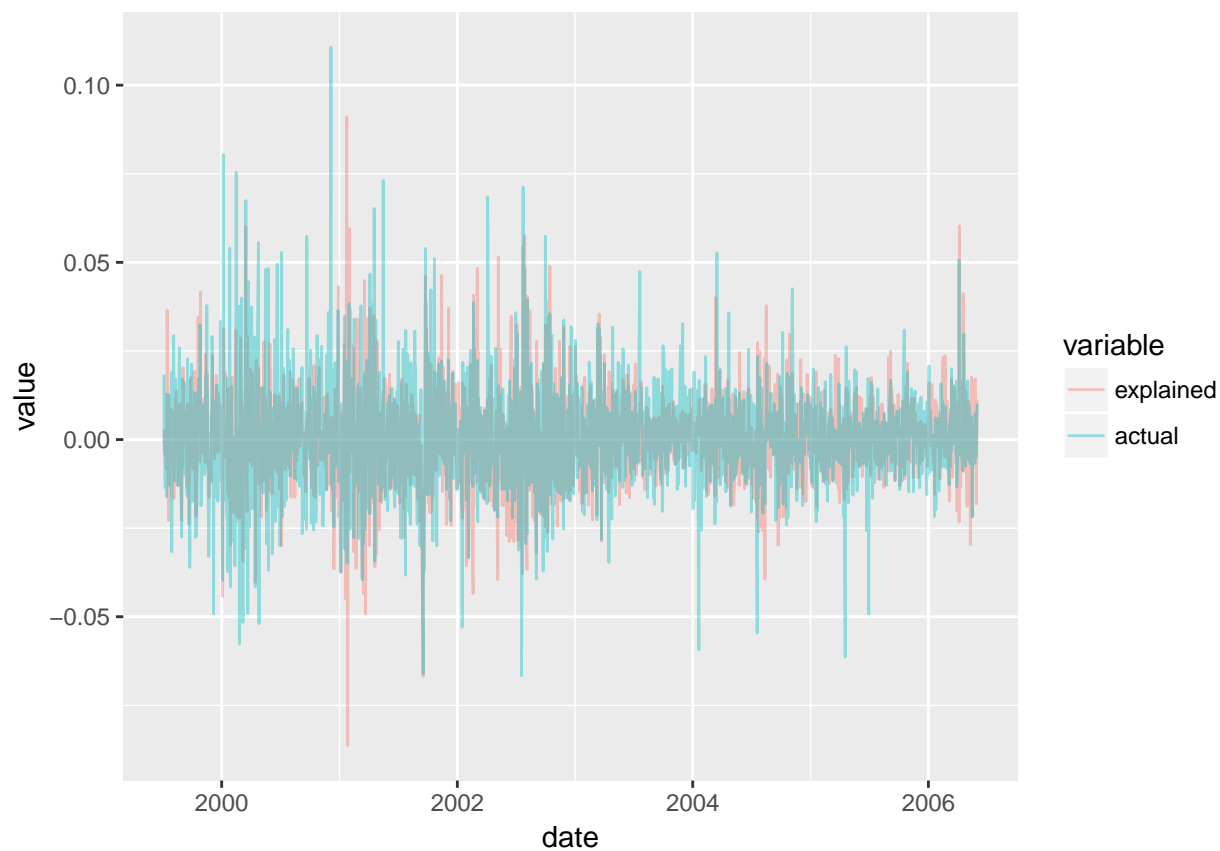
Regardless of market factors, some firm specific characteristics can be shared by similar firms. So maybe we can explain one firm's return by its peers. I think two method can be used based on my assumption. One is that we first calculate the historical correlations and picked stocks with top ranks according some threshold or limited number to pick and regress on contemporaneous returns of these stocks. The other method is that we apply LASSO on contemporaneous returns of all other stocks and use cross validation to find the best number to pick. I implemented the LASSO method because it is more data-driven.

The second model is quite simple. First I tried to use S&P 500 equal weight return as the factor to explain the firm return. Secondly, I used a Fama-French 3 Factor model to decompose the firm return. Both methods are done by time series regressions.

LASSO

It turns out that peers are not providing enough explanation power of the firm returns as shown below:

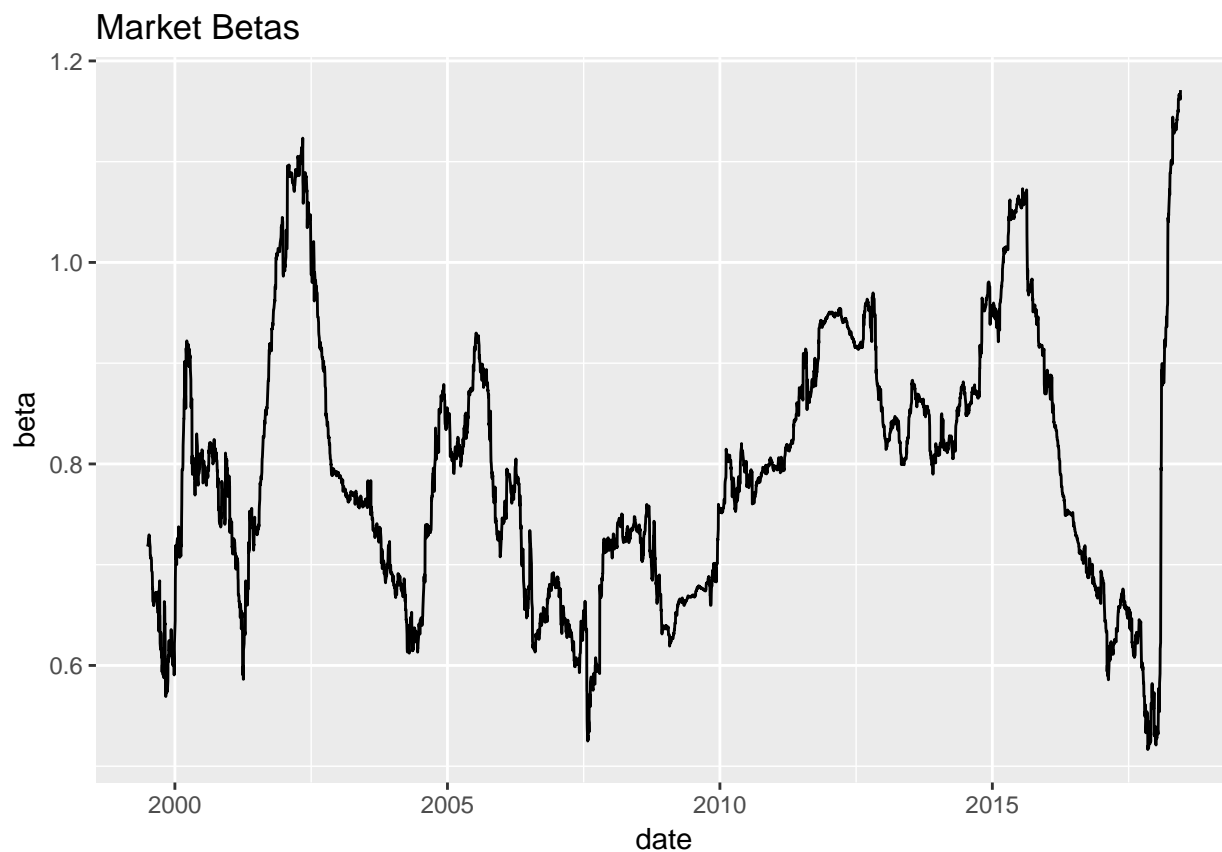
AverageExplainedReturn	AverageActualReturn	AverageDeviation
0.0006071	0.0006337	0.0109423



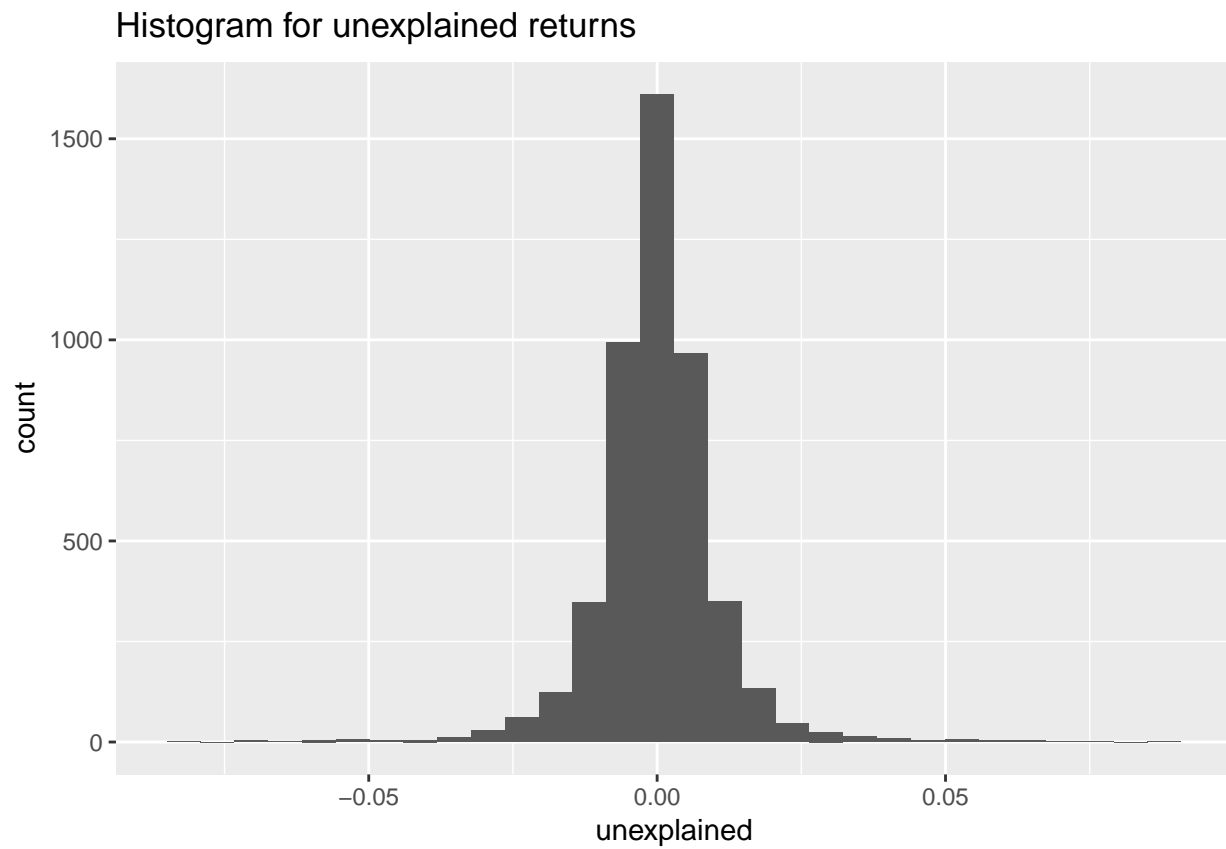
Factor Model

One Factor Model

First I used the market beta to explain the individual returns. I regress MMM daily returns on the return by equal-weighting the whole universe of S&P500 stocks, which can be a delegate of market. It does not provide a good explanation of the individual returns as you can see below. I would argue that there are other factors and firm-specific characteristics affecting stock prices.

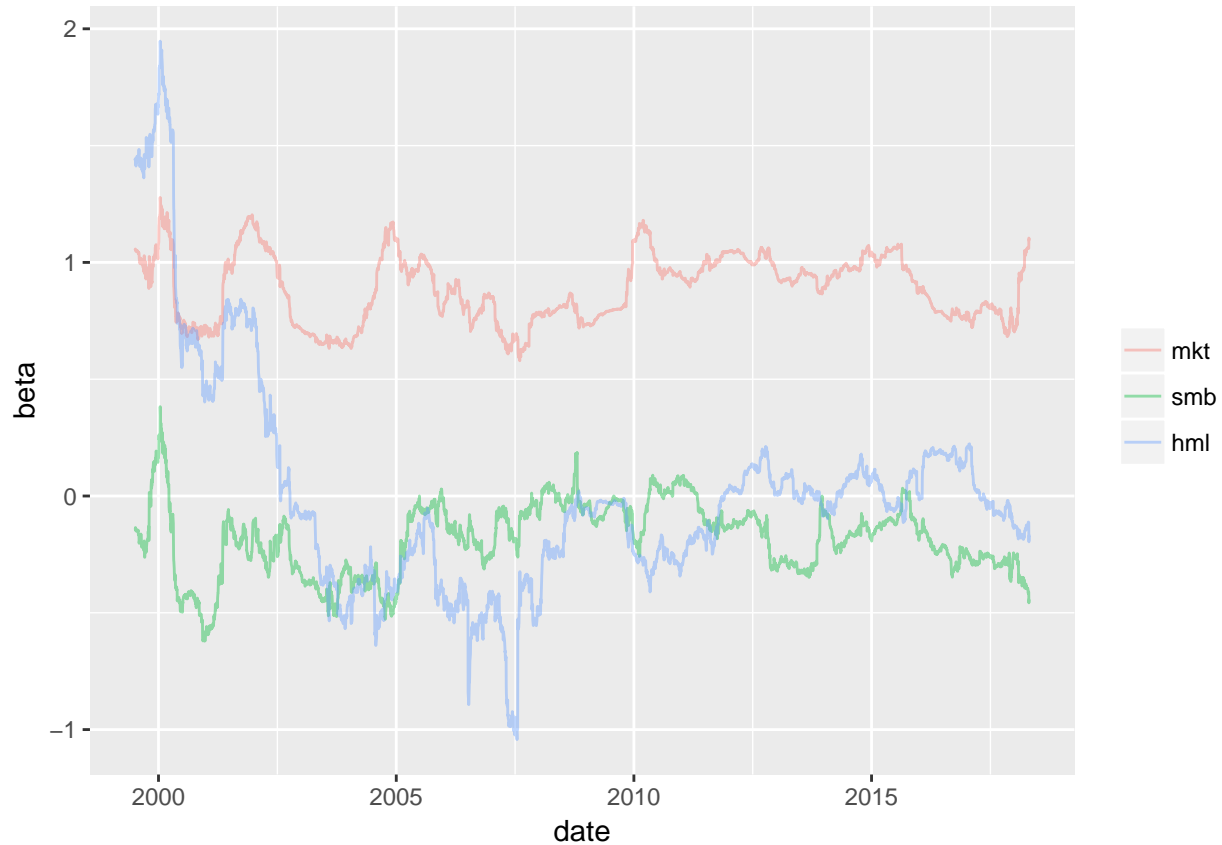


```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Three Factor Model

Here I used Fama-French 3-factor model to try to explain the individual returns. I didn't construct the factors by my own but used the available factors on Prof. French's website.



	AverageExplainedReturn	AverageActualReturn	AverageDeviation
1-Factor	0.0005496	0.0005448	0.0069212
3-Factor	0.0005821	0.0005357	0.0067636

Next Step

The reason why I don't want to use OLS regression is that it requires very long historical data because of the high dimensions eating up the degree of freedom. Two possible solutions came up to my mind. The first one was some typical dimension-reduction method, for example, PCA, DCT and Wavelet transformation. But they are pure data driven and I cannot find sound fundamental justification for them.

Then I came up with the Kalman filter, which is a dynamic linear MMSE estimator. It does not have the restriction on prior data. In fact, I am trying to implement the Kalman filter. I have not shown it in this summary because I currently have problem with the covariance matrix, which should be converging to a steady state according my experience. In my case, however, it diverges. I need some time to figure out how it happens since it is really time-consuming when dealing with large matrix (500-by-500) operations. I am not requiring additional time. I just want to make it perfect. So let me know if you are interested, I would keep it updated on my Git.