

Homework: Robust Regression And Equity Beta

May 14, 2019

0.1 Data

For each of the exchange-traded funds listed below, find its top 40 US stock holdings (some may not have 40, in which case include as many as you can). Your stock list consists of the size N set of all such companies held by any one of the ETFs.

Obtain 46-trading-day daily adjusted close prices of the stocks and ETFs, plus the QQQ and the SPY exchange-traded funds for January, February and March of this year. Form the corresponding trading-day daily return series.

0.2 Models

For each stock $S = S_1, \dots, S_N$, you will be working with a model for its returns of the form

$$r_S \sim r_{\text{SPY}} + r_{\text{QQQ}} + r_{\text{ETF}_1} + \dots + r_{\text{ETF}_k} \quad (1)$$

where $\text{ETF}_1, \dots, \text{ETF}_k$ are all of the ETFs from our list that the stock is a top-40 member of.

In each case, use the first 30 returns (from the first 31 days' prices), to estimate linear models¹ using

- OLS
- The Huber penalty function

¹In **R** you can use the **MASS** package. In Python try **statsmodels**.

- The Tukey Bisquare penalty function

Make sure your fitting algorithm is setting the scale parameter to something reasonable².

Do this both (A) with intercept forced to zero and (B) with intercept allowed to be nonzero.

0.3 Analysis

Obtain the residuals of each regression for the remaining 15 out-of-sample returns. Normalize the residuals by in-sample residual standard deviation.

Contrast the performance of the 6 estimators both

1. on the $15 \times N$ out-of sample residuals, and
2. on the $15 \times N$ out-of sample residuals normalized by in-sample residual standard deviation.

0.4 Data

ETF list:

1. XLU
2. VOX
3. SMH
4. XBI
5. XLF
6. XOP
7. OIH
8. XRT
9. XLV

²For daily returns, it should be obvious that 1.0 is *not* reasonable.