Discussion of:

Dissecting Characteristics Nonparametrically

by:
Joachim Freyberger, Andreas Neuhierl and Michael Weber

Kent Daniel[†]

[†] Columbia Business School & NBER

Texas Finance Festival
22 April, 2017

Columbia Business School

The State of Asset Pricing Research

- Harvey, Liu, and Zhu (2015): And the Cross-Section of Expected Returns." (314 anomalies)

  > *. . . Echoing a recent disturbing conclusion in the medical literature, we argue that most claimed research findings in financial economics are likely false.*

- McLean and Pontiff (2016): "Does Academic Research Destroy ... Predictability" (97 anomalies)
  - anomaly portfolio returns are 26% lower out-of-sample, and 58% lower post-publication.
- Linnainmaa and Roberts (2016): "The History of the Cross-Section of Returns"
  - accounting anomalies 1918-1960s using Moody's manuals.

  > *. . . most accounting-based return anomalies are spurious. When examined out-of-sample by moving either backward or forward in time, anomalies' average returns decrease, and volatilities and correlations with other anomalies increase.*

Characteristics & Returns

- Clearly we need to better understand:
  - Are there *any* stable relationships between characteristics and average returns?
  - Which characteristics matter?
  - What is the right functional form?
    - sparsity, interactions, time-variation
- Given the large no. of characteristics that seem to forecast returns (437,HXZ), the possible interactive effects, the nonlinearities (14 knots), and then possibility of time variation (53 years), the number of potential parameters is big:

$$\frac{437 \cdot 438}{2} \times 14 \times 53 \approx 7.1 \times 10^7$$

- In addition to this paper, there are a bunch of recent papers, each with slightly different approaches to tackling this problem.
  - Kozak, Nagel, and Santosh (2017) – "Shrinking the Cross-Section"
  - Feng, Giglio, and Xiu (2017) – "Taming the Factor Zoo"
  - Fama and French (2017) – "Choosing Factors"

Approach

- The estimation here uses a LASSO approach:

$$\tilde{\beta}_t = \underset{b_{s,k}}{\text{argmin}} \sum_{i=1}^{N} \left(R_{i,t} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{s,k} \cdot p_k(\tilde{c}_{i,s,t-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(\sum_{k=1}^{L+2} b_{s,k}^2 \right)^{1/2}$$

where $\tilde{c}_{i,s,t-1}$ is a normalized characteristic on $[0, 1]$ and where:

$$p_1(c) = 1, \;\; p_2(c) = c, p_3(c) = c^2$$

and

$$p_k(c) = \max\{c - t_{k-3}, 0\}^2 \;\; \text{for} \;\; k = 4, \ldots, L+2$$

Approach

- The estimation here uses a LASSO approach:

$$\tilde{\beta}_t = \underset{b_{s,k}}{\operatorname{argmin}} \sum_{i=1}^{N} \left(R_{i,t} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{s,k} \cdot p_k(\tilde{c}_{i,s,t-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(\sum_{k=1}^{L+2} b_{s,k}^2 \right)^{1/2}$$

where $\tilde{c}_{i,s,t-1}$ is a normalized characteristic on $[0, 1]$ and where:

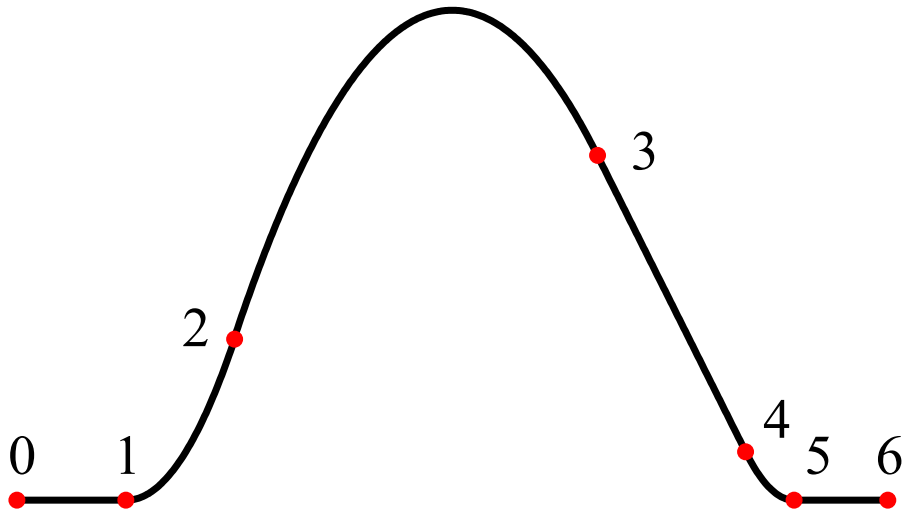
$$p_1(c) = 1, \;\; p_2(c) = c, p_3(c) = c^2$$

and

$$p_k(c) = \max\{c - t_{k-3}, 0\}^2 \;\text{ for }\; k = 4, \ldots, L+2$$

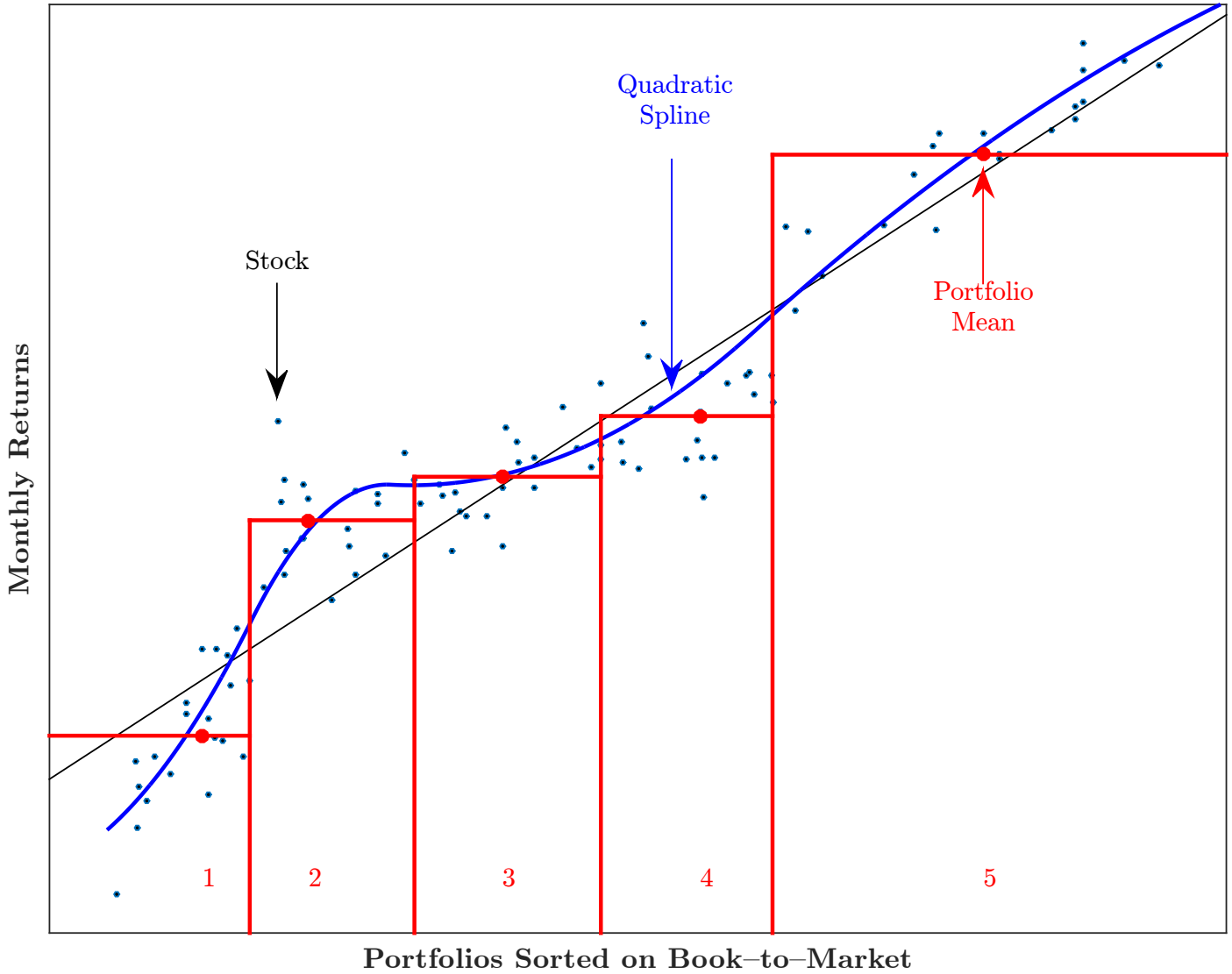
  - Note that this is just a Fama and MacBeth (1973) regression, with
    - Nonlinear (quadratic spline) terms and
    - a LASSO-penalty applied to the sum of the Euclidean norms of the sets of characteristic coefficients

Quadratic Spline

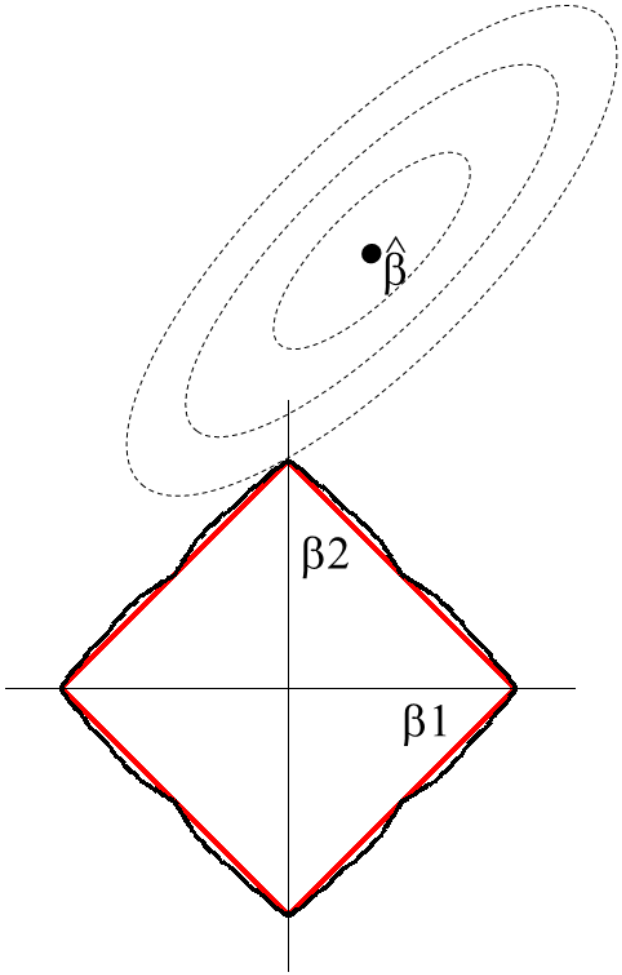


- Note that the second derivative of the resulting function is piecewise-linear.

Quadratic Spline



Lasso Method



see https://en.wikipedia.org/wiki/Lasso_(statistics)

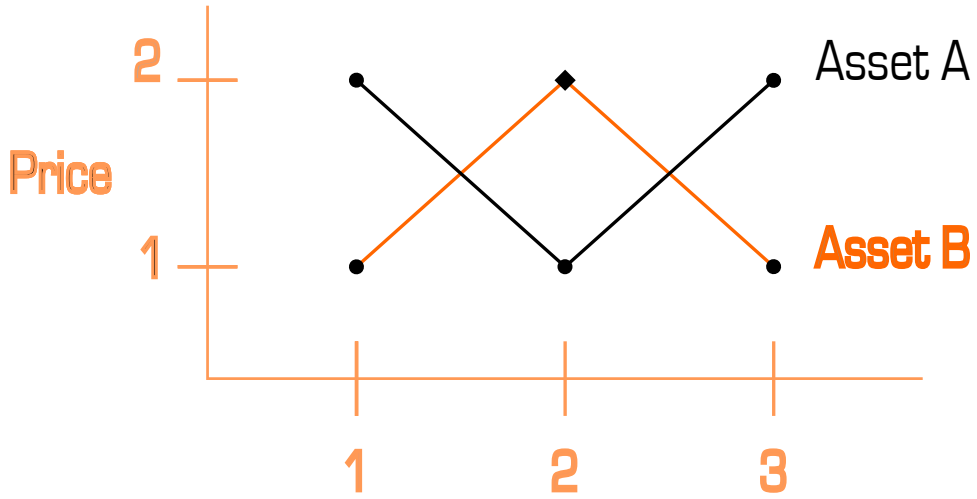
Results

- This approach developed here eliminates a bunch of characteristics based on a LASSO approach.
  - Out of the initial 36 characteristics, only 15 variables are found to have independent explanatory power.
- It works well out-of-sample:
  - The examine a long-short (decile, EW) portfolio, based their NP model.
  - The NP model generates an annualized Sharpe-ratio of 3.42 out-of-sample (1991-2014), compared to 2.26 for the linear model, a 50% increase.
    - Note that VW portfolios the SR is 1.24, as compared to 1.01 for the linear model.

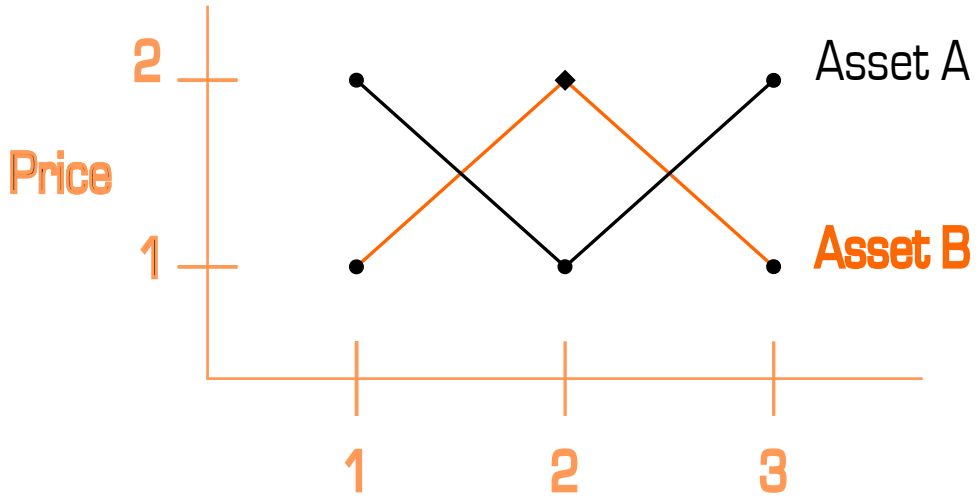
The Portfolio Approach – FF(93)

- This paper explores three approaches to estimating the expected return function:
  1. portfolio sorts.
  2. Fama and MacBeth (1973) regressions.
  3. non-parametric estimation plus a LASSO penalty.
- The portfolio approach has been the preferred approach in the finance literature.
  - e.g, Fama and French's papers rely almost exclusively on the value-weighted portfolio approach.
  - Certainly one reason for this, as noted here, is that it is inherently non-parametric
  - However, I think another reason is the robustness.

EW Portfolio Return Bias

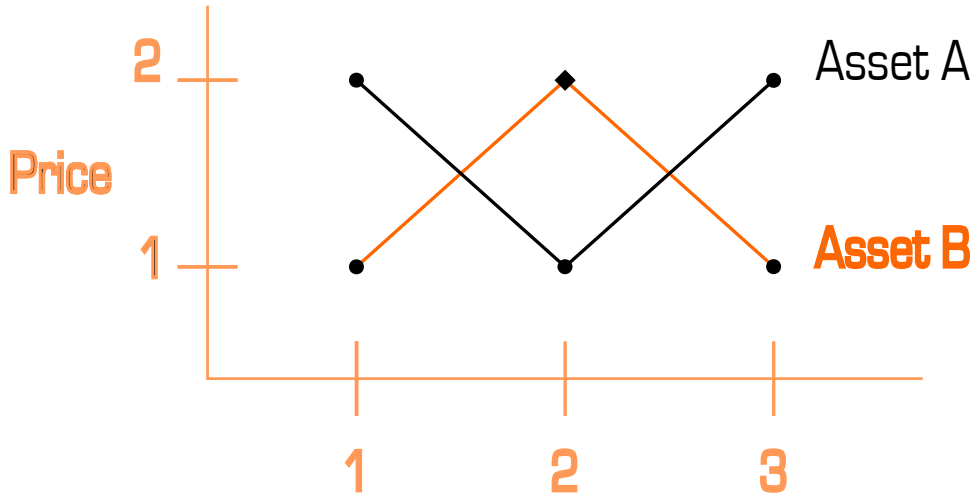


EW Portfolio Return Bias



$$
\begin{aligned}
R_{EW,2} &= (1/2)*(-50\%) &+(1/2)*(+100\%) &= 25\% \\
R_{EW,3} &= (1/2)*(+100\%) &+(1/2)*(-50\%) &= 25\%
\end{aligned}
$$

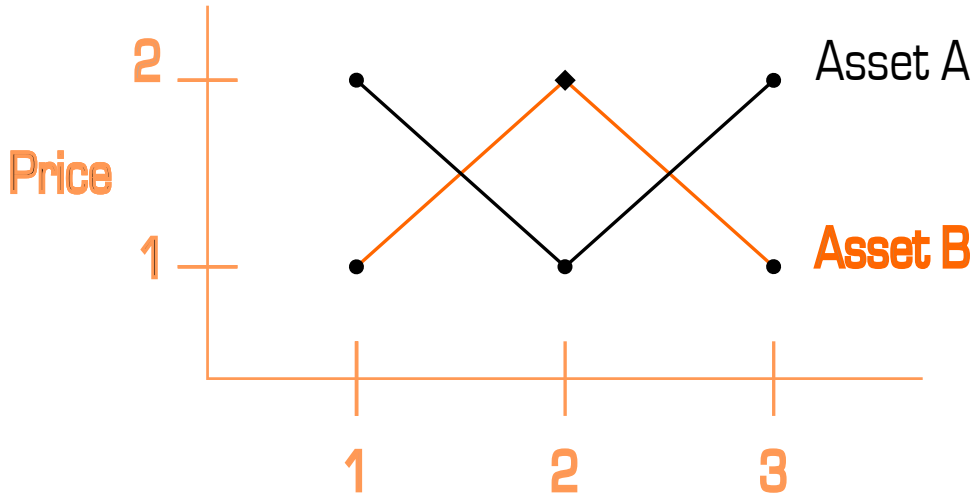
EW Portfolio Return Bias



$$
\begin{aligned}
R_{EW,2} &= (1/2)*(-50\%) &&+(1/2)*(+100\%) &&= 25\% \\
R_{EW,3} &= (1/2)*(+100\%) &&+(1/2)*(-50\%) &&= 25\%
\end{aligned}
$$

$$
\begin{aligned}
R_{EW,2} &= \frac{Gain}{InitialCost} &&= \frac{(1/4)*(-1)+(1/2)*(+1)}{(1/4)*2+(1/2)*1} &&= 25\% \\
R_{EW,3} &= \frac{Gain}{InitialCost} &&= \frac{(1/2)*(+1)+(1/4)*(-1)}{(1/2)*1+(1/4)*2} &&= 25\%
\end{aligned}
$$

EW Portfolio Return Bias

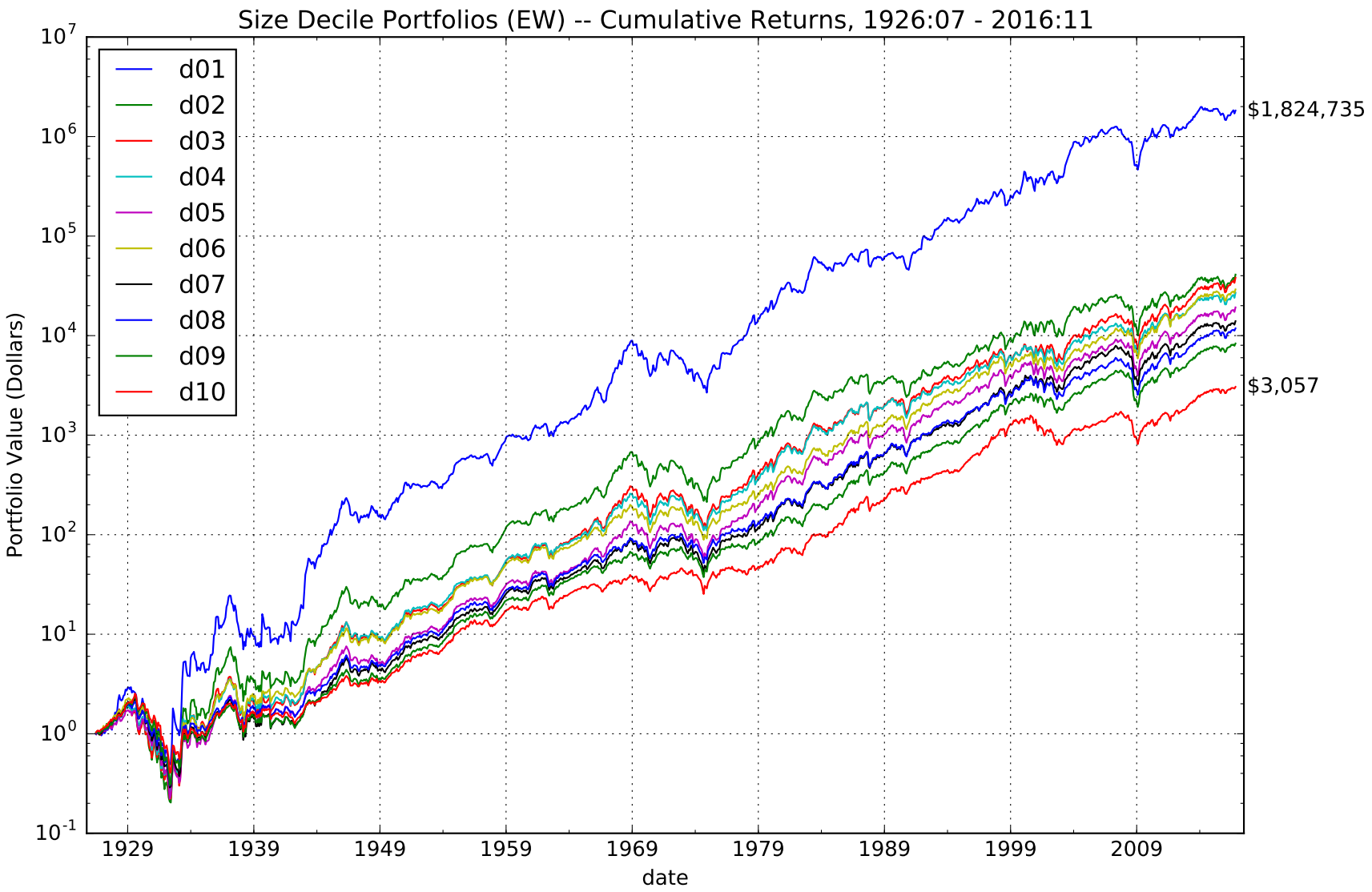


$$
\begin{aligned}
R_{EW,2} &= (1/2)*(-50\%) &&+(1/2)*(+100\%) &&= 25\% \\
R_{EW,3} &= (1/2)*(+100\%) &&+(1/2)*(-50\%) &&= 25\%
\end{aligned}
$$

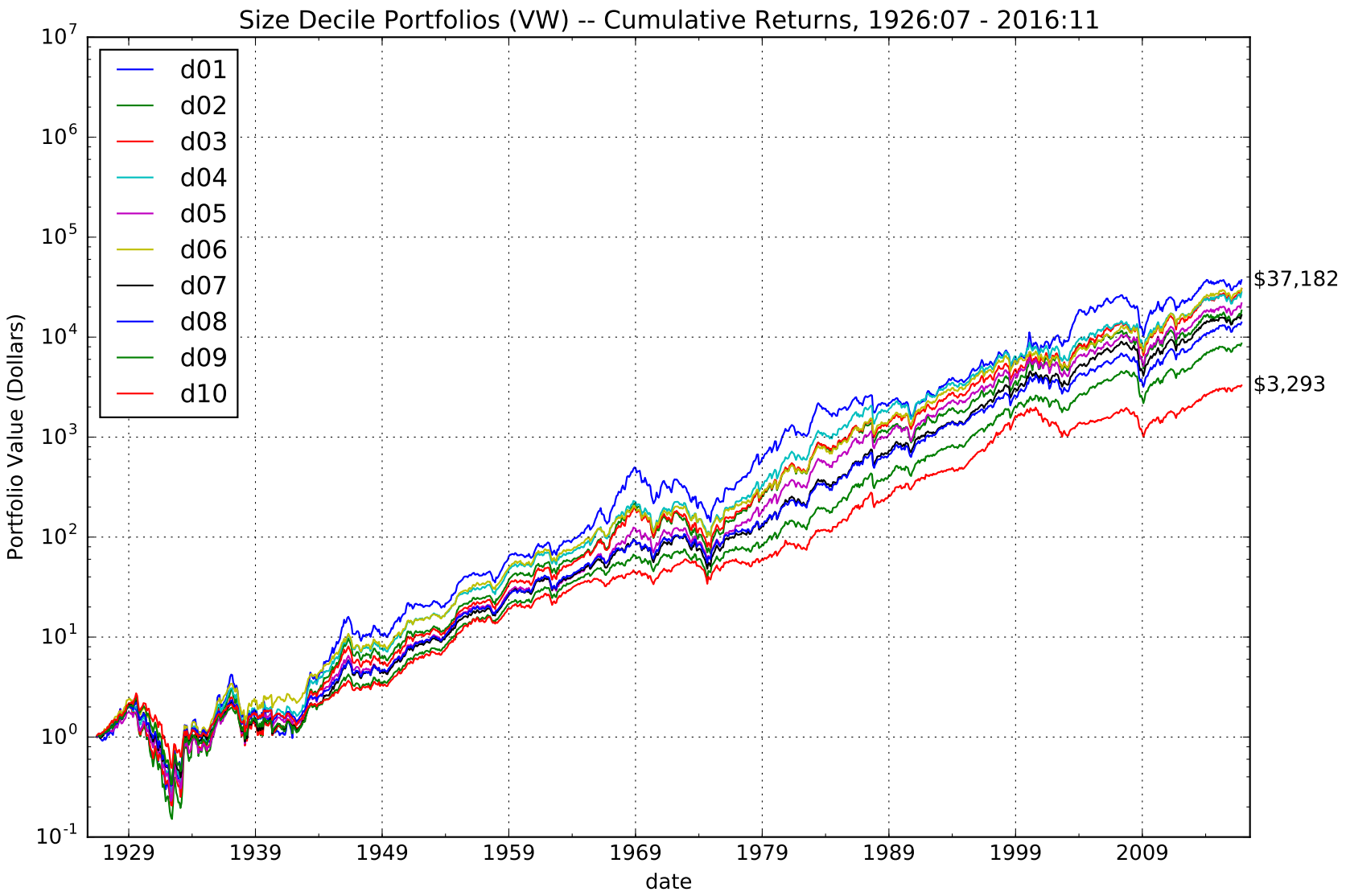
$$
\begin{aligned}
R_{EW,2} &= \frac{Gain}{InitialCost} = \frac{(1/4)*(-1)+(1/2)*(+1)}{(1/4)*2+(1/2)*1} &&= 25\% \\
R_{EW,3} &= \frac{Gain}{InitialCost} = \frac{(1/2)*(+1)+(1/4)*(-1)}{(1/2)*1+(1/4)*2} &&= 25\%
\end{aligned}
$$

- *This bias will arise with any constant weight portfolio approach, including Fama-MacBeth, or the NP version of FM used here.*

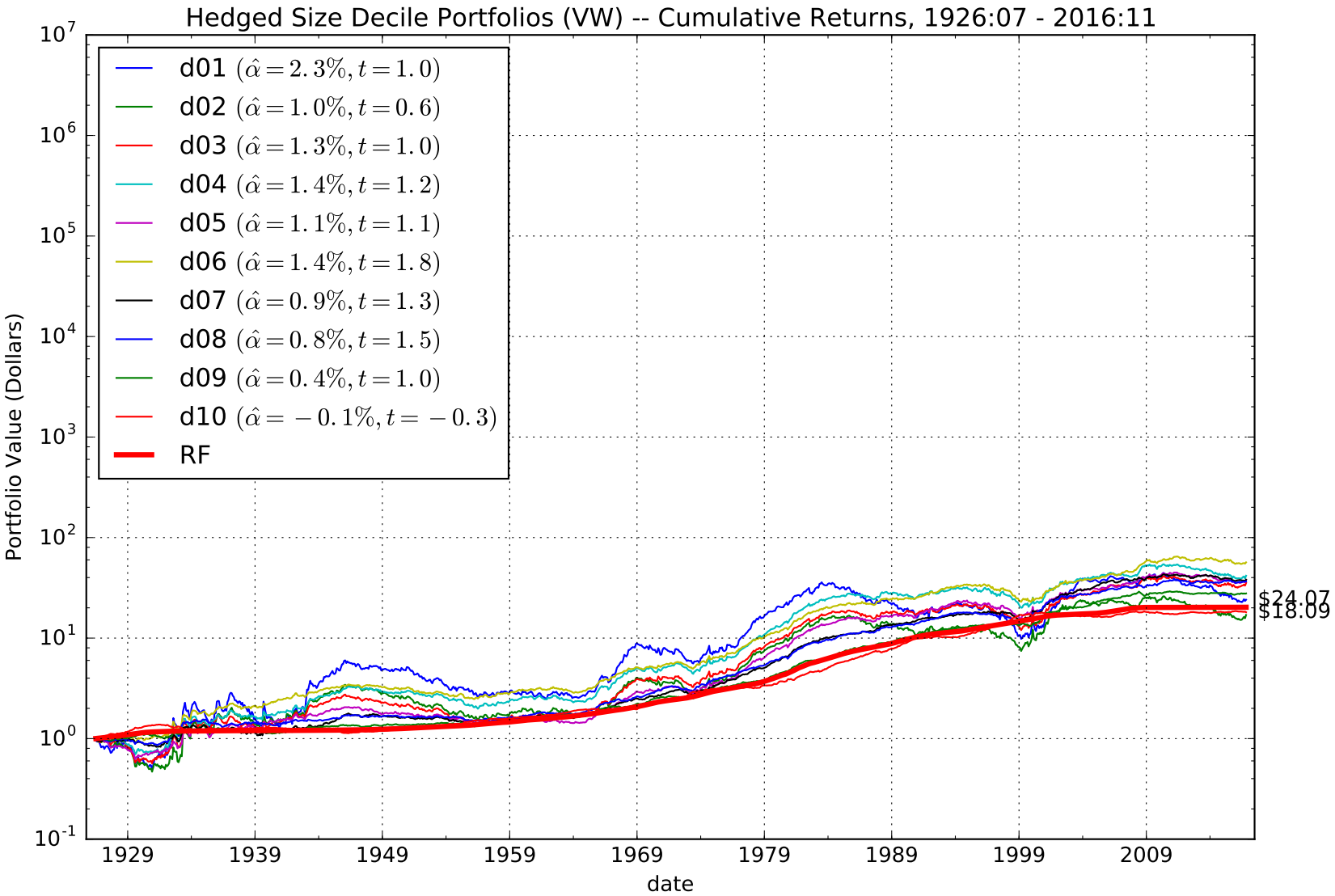
EW Size Decile Portfolio Returns



VW Size Decile Portfolio Returns



Size Decile Portfolio Returns



Adjusted FM Regressions

- Again, the same bias arises with *any* approach where the portfolio weights are relatively constant

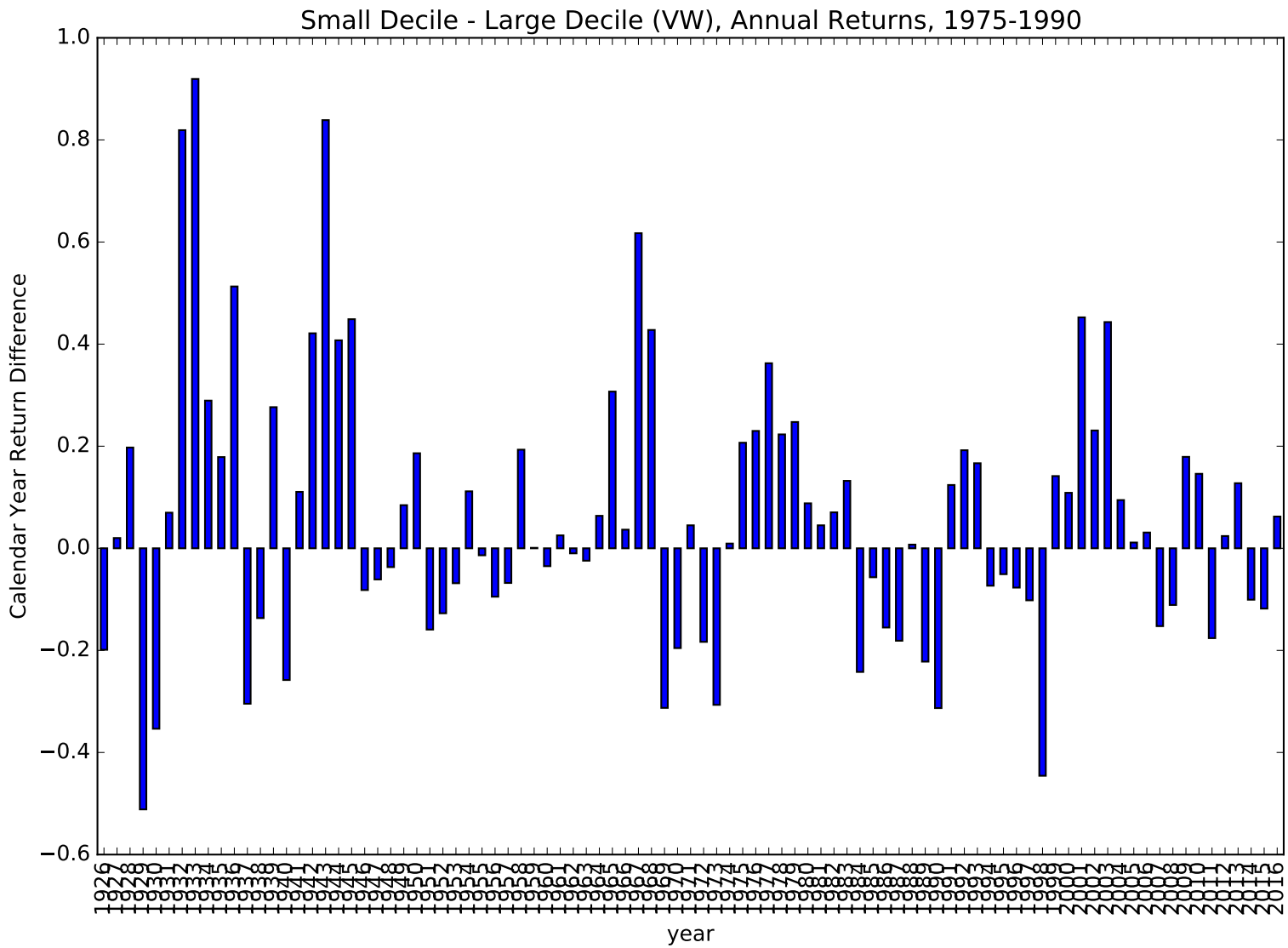
$$\begin{aligned} \beta_{t+1} &= \left(\mathbf{X}_t \mathbf{X}_t'\right)^{-1} \mathbf{X}_t \mathbf{R}_{t+1} \\ &= \mathbf{w}_t^{\mathrm{FM}'} \mathbf{R}_{t+1} \end{aligned}$$

  - See, e.g., Asparouhova, Bessembinder, and Kalcheva (2010, 2013).
- In Daniel and Titman (2006), we propose a "value-weighted" FM portfolio that alleviates this bias:

$$\mathbf{w}_t^{\mathrm{VW}} = \frac{\mathbf{w}_t^{\mathrm{FM}+} \cdot \mathbf{ME}_t^{+}}{ME_t^{+}} + \frac{\mathbf{w}_t^{\mathrm{FM}-} \cdot \mathbf{ME}_t^{-}}{ME_t^{-}}$$

  - where $\mathbf{ME}_t^{+,-}$ is the N$\times$1 vector of market capitalizations at time t, for firms with a $+/-$ weight in $\mathbf{w}_t^{\mathrm{FM}}$, and $ME_t^{+} = \mathbf{ME}_t^{+'}\mathbf{1}$

Size Decile Portfolio Returns



Sentiment Interaction

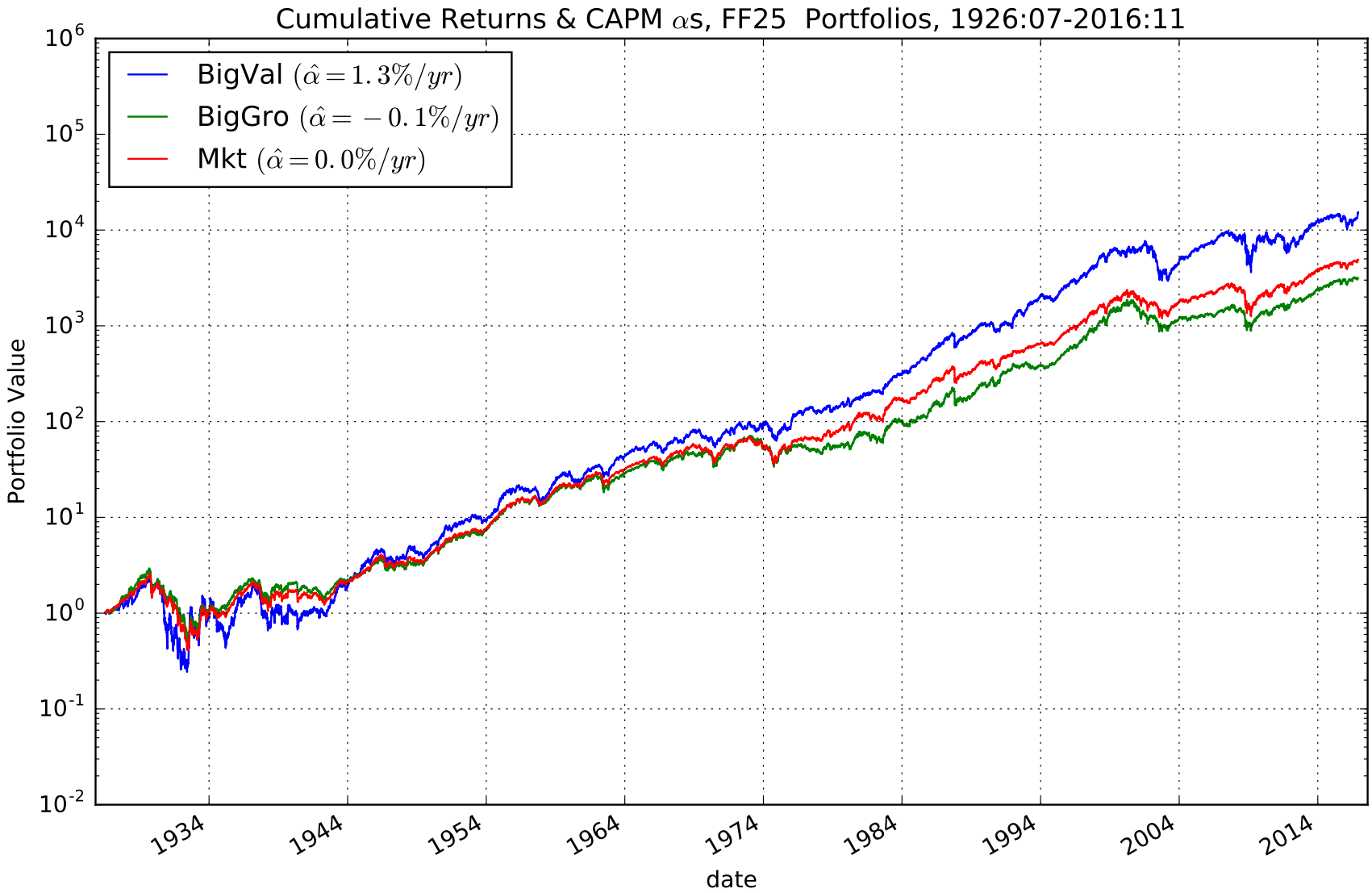
Table 2
Anomalies during periods of high and low investor sentiment: excess returns on long-short strategies.

The table reports average returns in excess of the one-month T-bill in months following high and low levels of investor sentiment, as classified based on the median level of the index of Baker and Wurgler (2006). Also reported are returns on a strategy that equally combines the strategies available within a given month. The sample period is from 1965:8 to 2008:1 for all but anomaly (1), whose data begin 1974:12, and anomalies (2) and (10), whose data begin 1972:1. All *t*-statistics are based on the heteroskedasticity-consistent standard errors of White (1980).

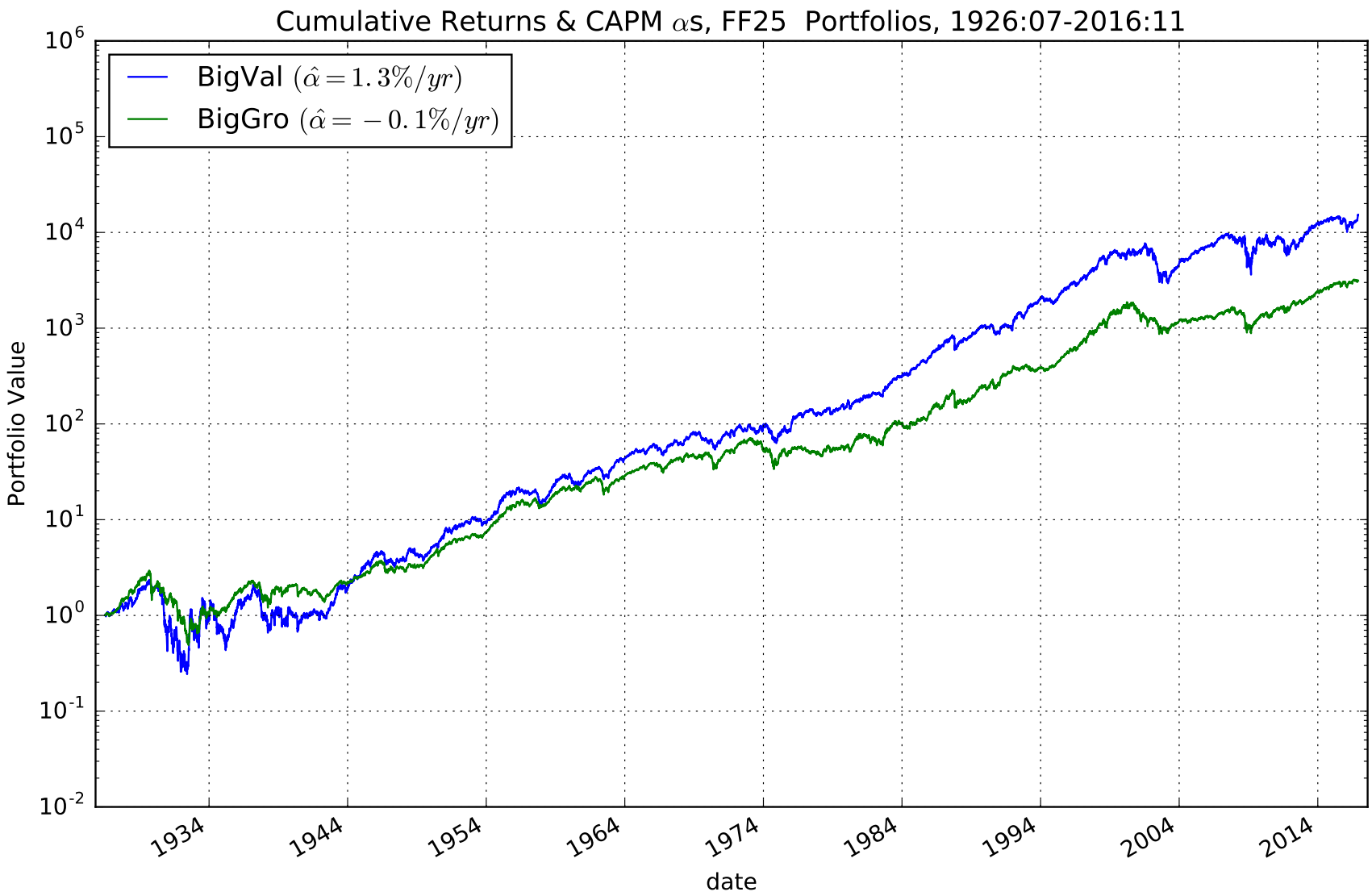
Anomaly	Long leg			Short leg			Long–short		
	High sentiment	Low sentiment	High –low	High sentiment	Low sentiment	High –low	High sentiment	Low sentiment	High –low
Failure probability	0.77	1.14	−0.38	−1.10	1.25	−2.34	1.86	−0.10	1.96
	(2.16)	(3.74)	(−0.81)	(−1.54)	(2.26)	(−2.60)	(3.25)	(−0.24)	(2.72)
Ohlson's O (distress)	0.42	0.61	−0.19	−0.98	0.61	−1.59	1.40	−0.00	1.40
	(1.14)	(2.06)	(−0.41)	(−1.69)	(1.33)	(−2.15)	(3.81)	(−0.01)	(2.85)
Net stock issues	0.64	0.75	−0.11	−0.50	0.63	−1.13	1.14	0.12	1.02
	(2.22)	(3.04)	(−0.28)	(−1.26)	(2.10)	(−2.28)	(5.71)	(0.88)	(4.20)
Composite equity issues	0.53	0.72	−0.19	−0.28	0.69	−0.97	0.81	0.02	0.79
	(1.93)	(3.08)	(−0.52)	(−0.72)	(2.13)	(−1.91)	(3.19)	(0.13)	(2.46)
Total accruals	0.37	1.07	−0.71	−0.57	0.84	−1.41	0.94	0.23	0.70
	(0.82)	(3.10)	(−1.25)	(−1.06)	(2.22)	(−2.14)	(3.11)	(1.04)	(1.88)
Net operating assets	0.50	0.92	−0.43	−0.57	0.69	−1.26	1.07	0.24	0.83
	(1.36)	(3.01)	(−0.90)	(−1.37)	(2.20)	(−2.41)	(4.66)	(1.29)	(2.84)
Momentum	0.78	1.43	−0.64	−1.24	0.34	−1.58	2.03	1.09	0.93
	(1.69)	(4.12)	(−1.11)	(−2.14)	(0.76)	(−2.16)	(4.49)	(3.12)	(1.64)
Gross profitability	0.59	0.79	−0.20	−0.06	0.64	−0.70	0.65	0.15	0.50
	(1.84)	(2.73)	(−0.47)	(−0.18)	(2.48)	(−1.62)	(2.93)	(0.64)	(1.53)
Asset growth	0.79	1.22	−0.43	−0.60	0.68	−1.27	1.39	0.54	0.85
	(2.14)	(3.26)	(−0.81)	(−1.30)	(1.92)	(−2.20)	(5.04)	(2.34)	(2.37)
Return on assets	0.61	0.66	−0.05	−1.10	0.44	−1.55	1.72	0.22	1.50
	(1.60)	(2.10)	(−0.10)	(−1.78)	(1.00)	(−2.02)	(4.01)	(0.65)	(2.74)
Investment-to-assets	0.44	1.38	−0.94	−0.47	0.78	−1.25	0.91	0.60	0.30
	(1.19)	(4.13)	(−1.90)	(−1.14)	(2.25)	(−2.32)	(4.48)	(2.93)	(1.06)
Combination	0.56	0.95	−0.39	−0.68	0.65	−1.32	1.23	0.31	0.93
	(1.72)	(3.51)	(−0.93)	(−1.54)	(1.96)	(−2.41)	(6.64)	(2.64)	(4.25)

† from Stambaugh, Yu, and Yuan (2012)

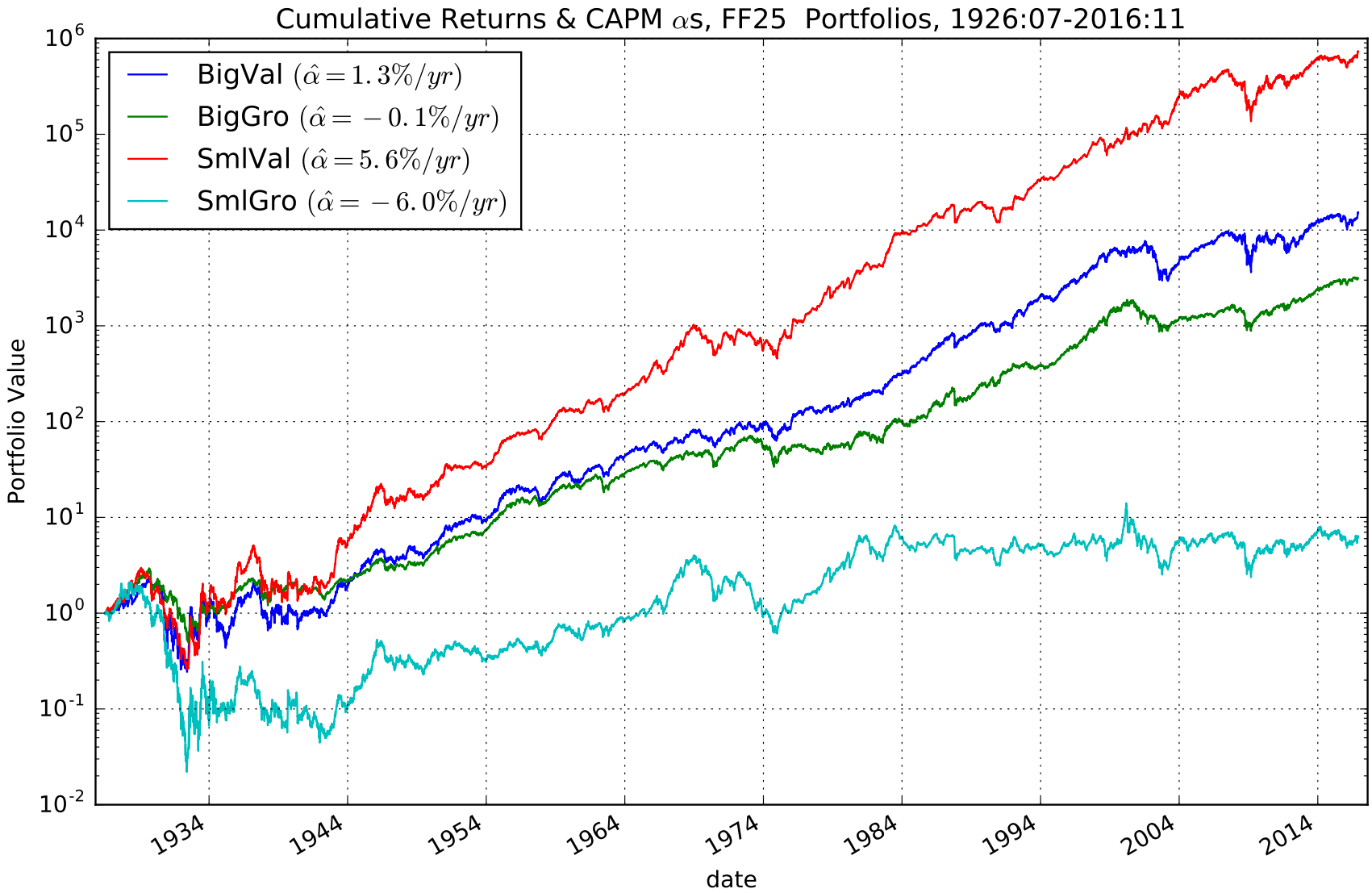
Size-Value Interactions



Size-Value Interactions



Size-Value Interactions



Estimation

- As noted earlier, the estimation here uses a LASSO approach:

$$\tilde{\beta}_t = \underset{b_{s,k}}{\text{argmin}} \sum_{i=1}^{N} \left(R_{i,t} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{s,k} p_k(\tilde{C}_{i,s,t-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(\sum_{k=1}^{L+2} b_{s,k}^2 \right)^{1/2}$$

where

$$p_1(c) = 1, \;\; p_2(c) = c, p_3(c) = c^2$$

and

$$p_k(c) = \max\{c - t_{k-3}, 0\}^2 \;\text{ for }\; k = 4, \ldots, L+2$$

Estimation

- However, since the first step LASSO approach (with BIC used to select λ_1) "selects too many characteristics."
- FNW therefore perform a second-stage estimation

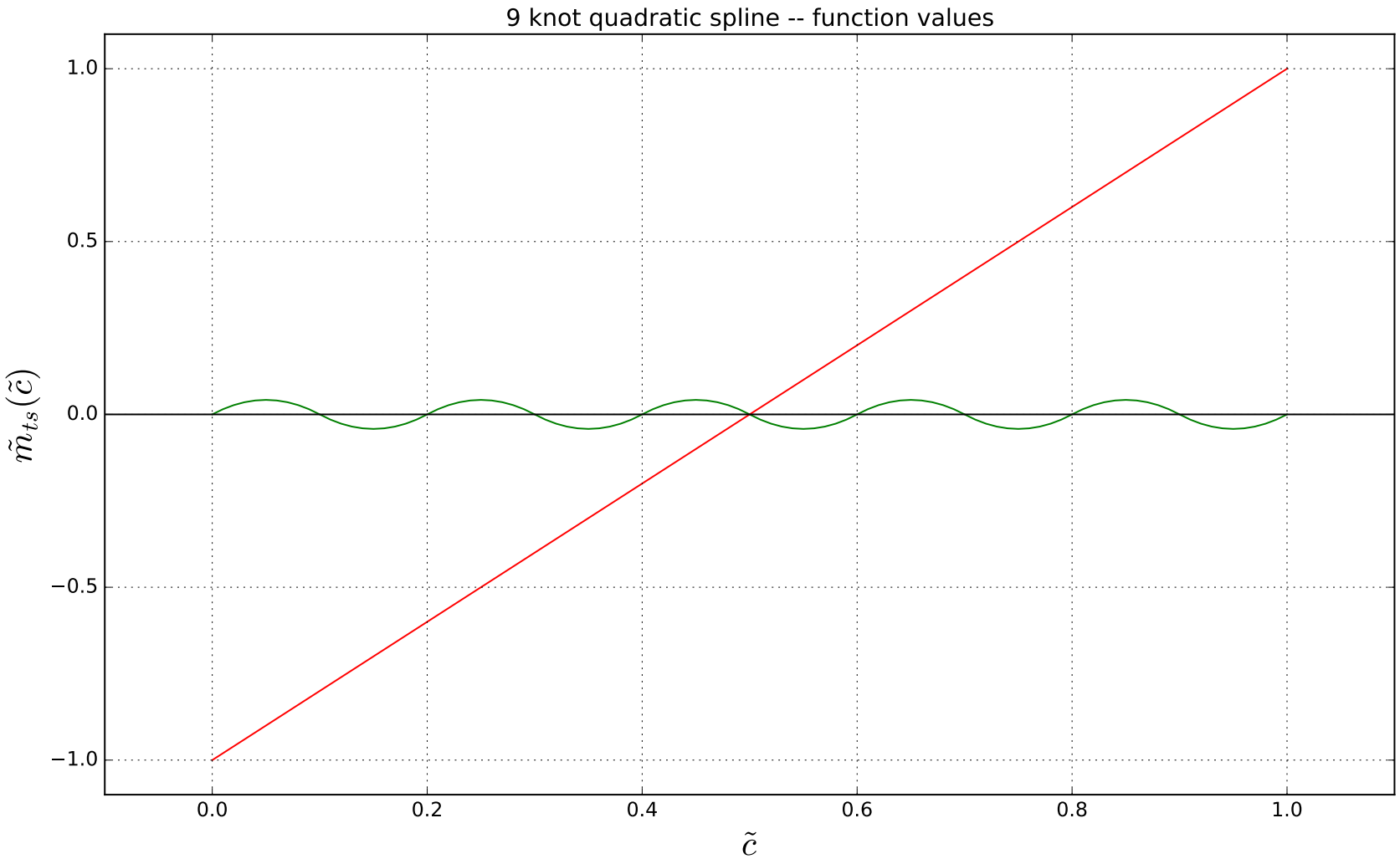
$$\breve{\beta}_t = \underset{b_{s,k}}{\text{argmin}} \sum_{i=1}^{N} \left(R_{i,t} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{s,k} p_k(\tilde{C}_{i,s,t-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(w_s \sum_{k=1}^{L+2} b_{s,k}^2 \right)^{1/2}$$

where

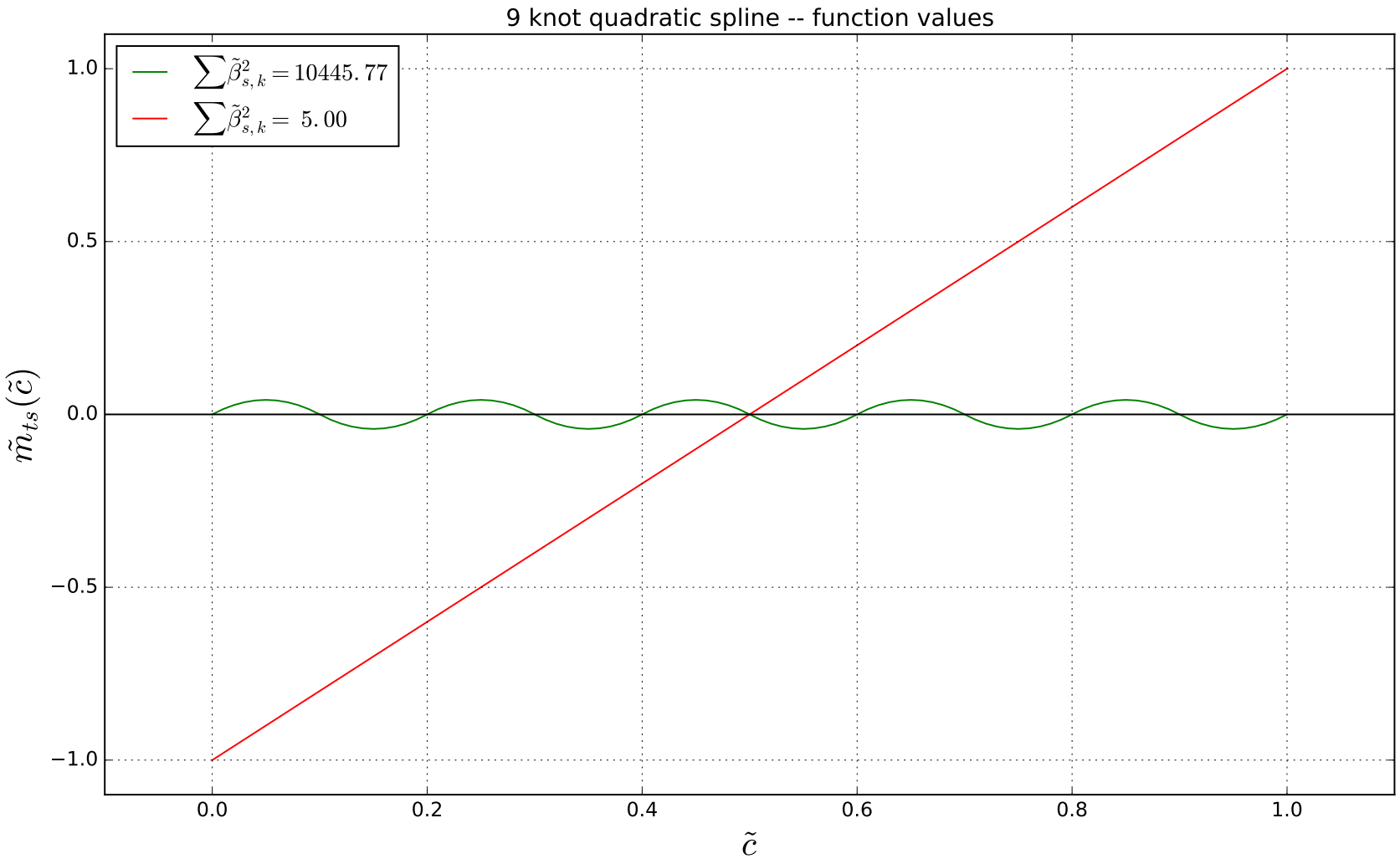
$$w_s = \begin{cases} \left(\sum_{k=1}^{L+2} \tilde{\beta}_{s,k}^2 \right)^{-1/2} & \text{if} \quad \sum_{k=1}^{L+2} \tilde{\beta}_{s,k}^2 \neq 0. \\ \infty & \text{if} \quad \sum_{k=1}^{L+2} \tilde{\beta}_{s,k}^2 = 0. \end{cases}$$

- What this second stage will do is to "throw out" characteristics that, in the first round estimation, had a low $|\tilde{\beta}_{s,t}|$.

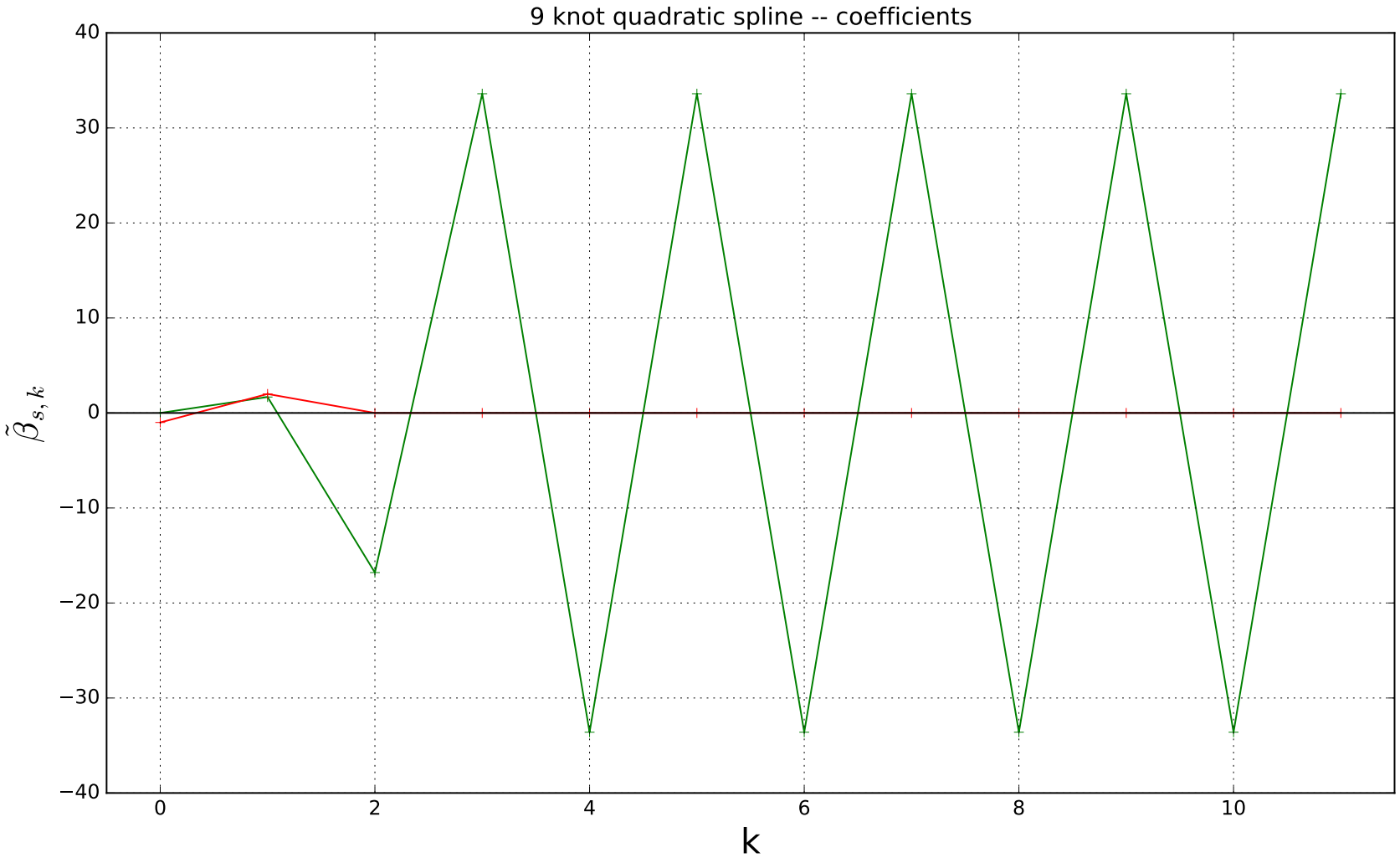
Two Functions...



Two Functions...

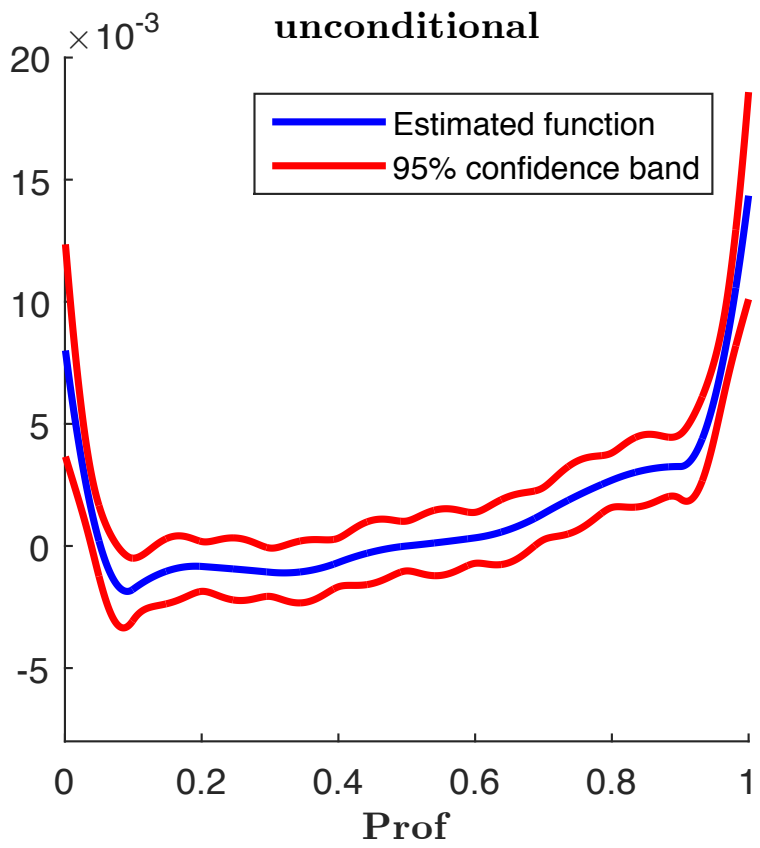


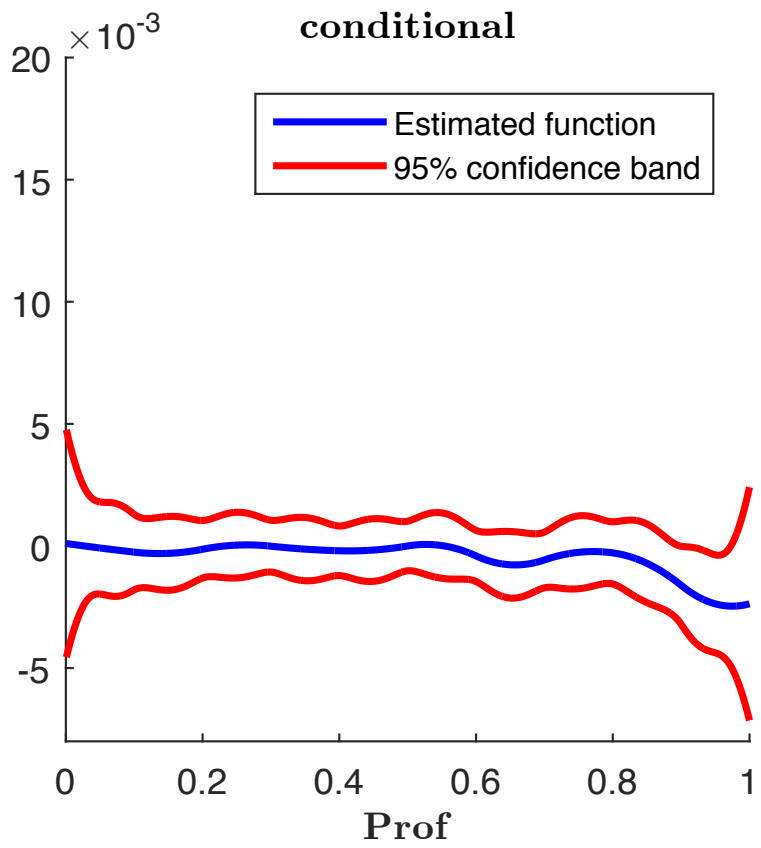
Function Coefficients



VW vs. EW (Constant Weight) Portfolios

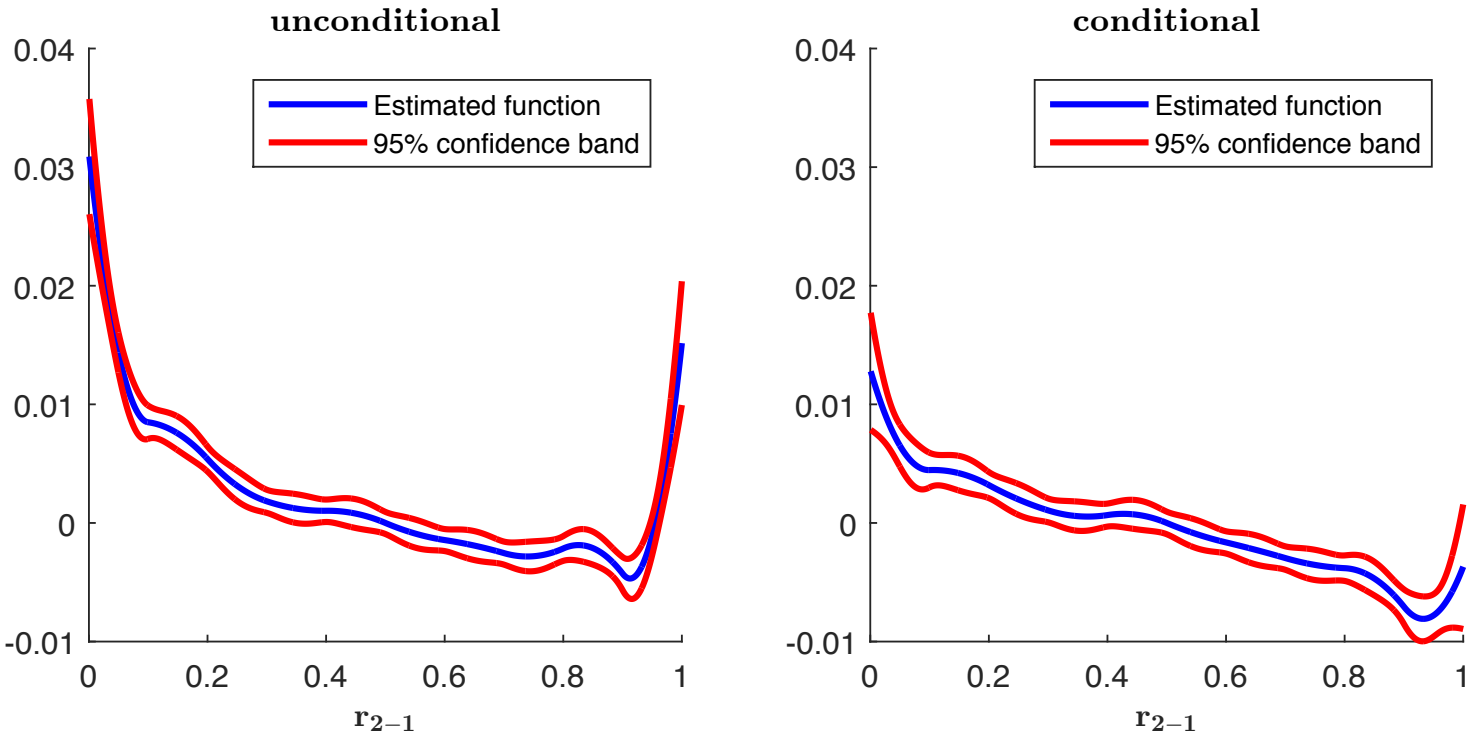
From Figure 7, lower panel:





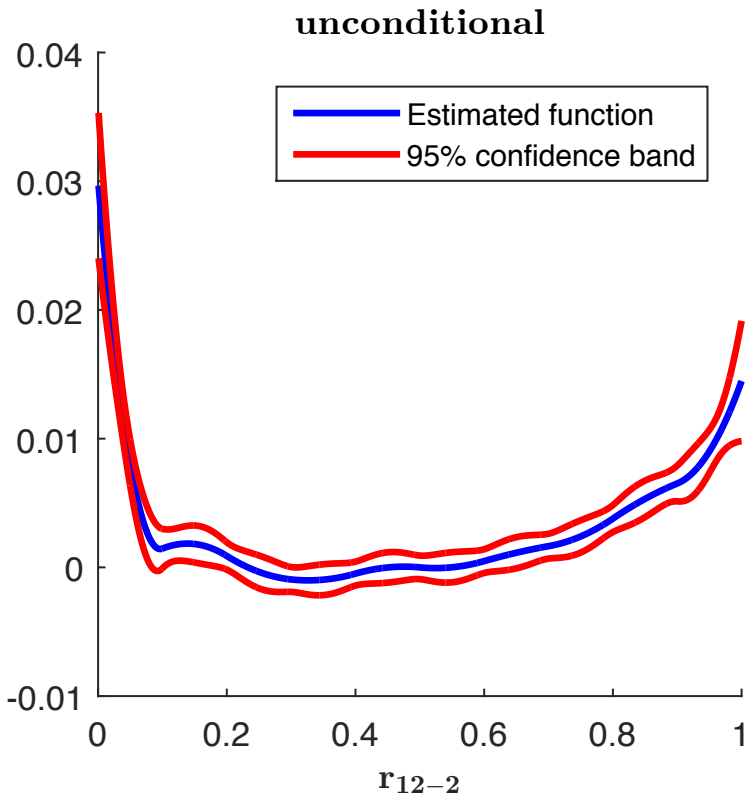
VW vs. EW (Constant Weight) Portfolios

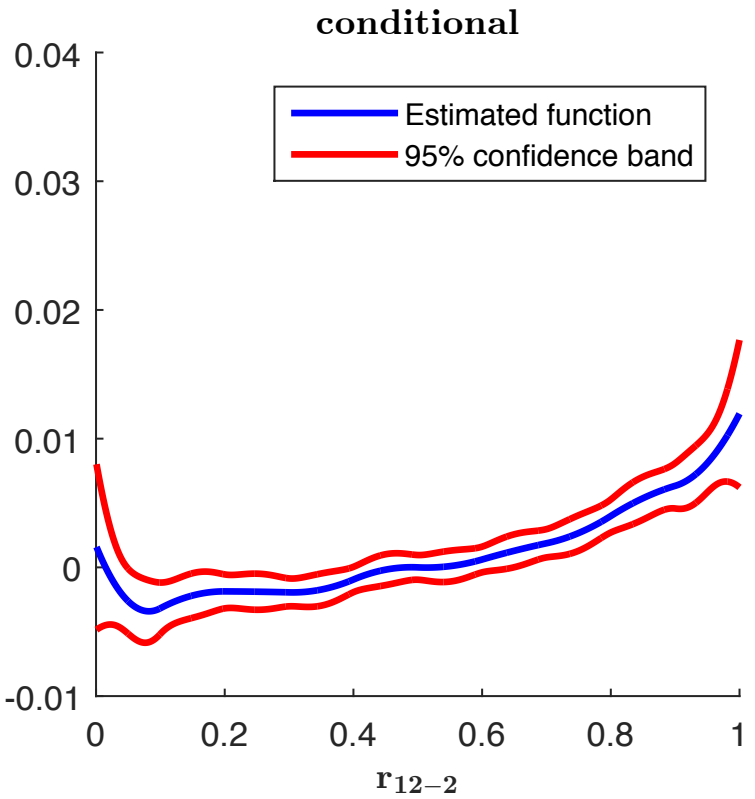
From Figure 9, upper panel:



VW vs. EW (Constant Weight) Portfolios

From Figure 9, lower panel:





Out-of-sample Results

- It is possible that what is driving the really-good out-of-sample Sharpe-ratio of the NP method (3.42-NP/2.26-Linear) is the constant-weight estimation
  - This will tend to overweight illiquid firms with substantial negative serial correlation.
- Since STR is a really persistent phenomenon, EW portfolios of illiquid firms will continue to perform well out of sample.
  - Recall that, for VW portfolios, the difference is far smaller (1.24-NP/1.01-Linear)
- It shouldn't be that hard to implement an estimation approach that fixes this.

Other Concerns

- Should the model specification be enhanced?
  - e.g., cross-sectional size-value interactions; time-series sentiment interaction.

Other Concerns

- Should the model specification be enhanced?
  - e.g., cross-sectional size-value interactions; time-series sentiment interaction.
- Risk-analysis

$$w^* \neq \mu \qquad\qquad w^* = \Sigma^{-1}\mu$$

  - *A high SR may not reflect a good* $\mathbb{E}[\mathbf{R}]$ *model.*
  - The analysis should incorporate estimates of the covariance structure if the objective is to maximize the Sharpe ratio

Other Concerns

- Should the model specification be enhanced?
  - e.g., cross-sectional size-value interactions; time-series sentiment interaction.
- Risk-analysis

$$w^* \neq \mu \qquad\qquad w^* = \Sigma^{-1}\mu$$

  - *A high SR may not reflect a good* $\mathbb{E}[\mathbf{R}]$ *model.*
  - The analysis should incorporate estimates of the covariance structure if the objective is to maximize the Sharpe ratio
- Is parsimony a reasonable objective?
  - Is the double exponential prior reasonable?
  - If the objective is to maximize the OOS Sharpe-ratio, why not try something like a Black and Litterman (1991) approach?

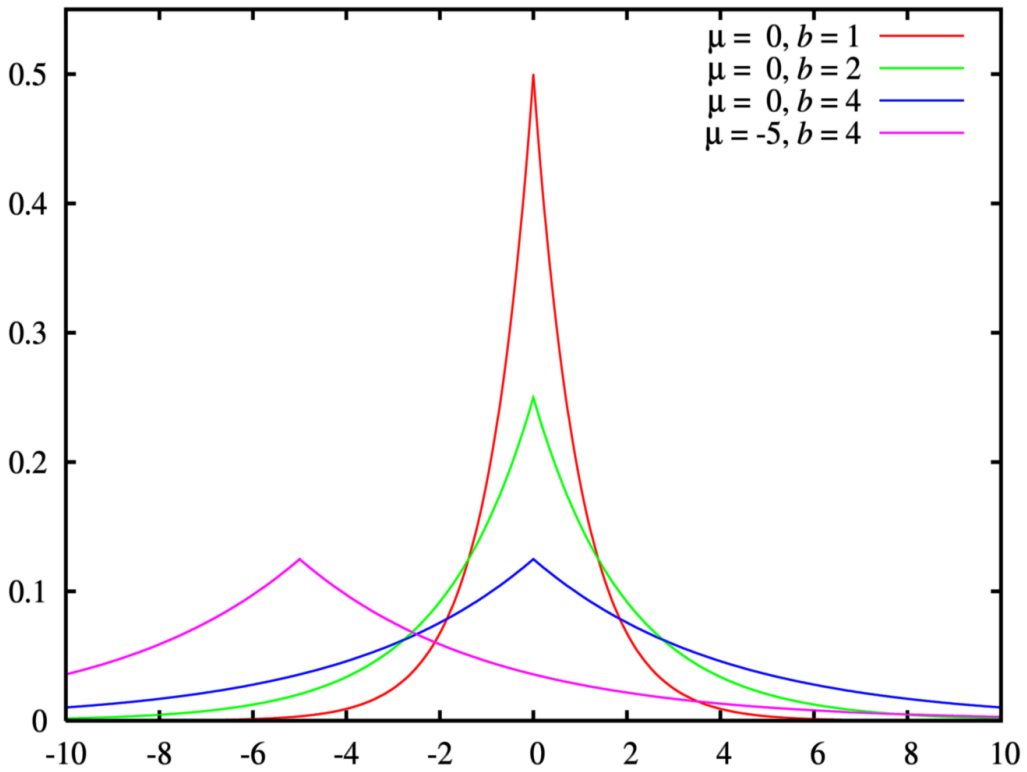
Other Concerns

- Should the model specification be enhanced?
  - e.g., cross-sectional size-value interactions; time-series sentiment interaction.
- Risk-analysis

$$w^* \neq \mu \qquad\qquad w^* = \Sigma^{-1}\mu$$

  - *A high SR may not reflect a good* $\mathbb{E}[\mathbf{R}]$ *model.*
  - The analysis should incorporate estimates of the covariance structure if the objective is to maximize the Sharpe ratio
- Is parsimony a reasonable objective?
  - Is the double exponential prior reasonable?
  - If the objective is to maximize the OOS Sharpe-ratio, why not try something like a Black and Litterman (1991) approach?
- What are we learning about different models of the price formation process from this exercise?

Double-Exponential Prior



References I

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2010, Liquidity biases in asset pricing tests, *Journal of Financial Economics* 96, 215–237.

——— , 2013, Noisy prices and inference regarding returns, *The Journal of Finance* 68, 665–714.

Black, Fischer, and Robert Litterman, 1991, Global portfolio optimization, *Financial Analysts' Journal* 48, 28–43.

Daniel, Kent D., and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

Fama, Eugene F., and Kenneth R. French, 2017, Choosing factors, Chicago Booth Paper No. 16-17.

Fama, Eugene F., and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2017, Taming the factor zoo, Chicago Booth Paper No. 17-04.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2015, ... and the cross-section of expected returns, *Review of Financial Studies* p. hhv059.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2017, Shrinking the cross section, University of Michigan Business School Working Paper.

Linnainmaa, Juhani T, and Michael R Roberts, 2016, The history of the cross section of stock returns, USC working paper.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.