Robust Regression for QuantFin and Fama-French 1992 Redux

R. Douglas Martin

Professor Emeritus
Applied Mathematics and Statistics
University of Washington
doug@amath.washington.edu

QWAFAFEW, Boston
May 21, 2019

NOTE:

This slide deck is a slightly modified version of the talk slide deck, whereby:

(1) The order of the sections has been changed by moving the “Best 1-D Robust Outlier Methods” first section to the last section and renaming it “Robust Location Estimates Applications”

(2) A few slides have been added and a few have been deleted

Main Reference is Chapter 5 of:

Robust Statistics: Theory and Methods (2019). 2nd Ed.
Maronna, Martin, Yohai & Salibian-Barrera, Wiley. (**MMYS**)

- Companion R package: RobStatTM (2019, beta)

  Maintainer: Matias Salibian-Barrera

To install from CRAN load, and view functions and data sets:

https://cran.r-project.org/package=RobStatTM

To load, and view functions and data sets:

```
> library(RobStatTM)
> ls("package:RobStatTM")
> data(package = "RobStatTM")
```

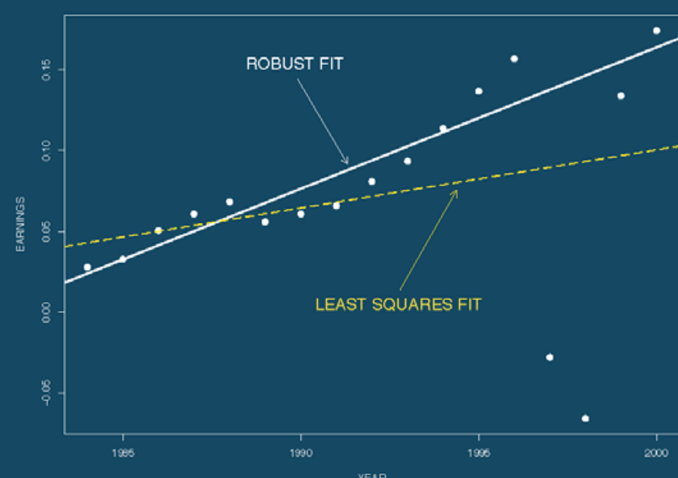
Wiley Series in Probability and Statistics

Second Edition

ROBUST STATISTICS

THEORY AND METHODS (WITH R)

Ricardo A. Maronna
R. Douglas Martin
Víctor J. Yohai
Matías Salibian-Barrera



WILEY

Outline

1. Robust regression overview
2. Fama-French 1992 results
3. FF92 Redux with robust regression
4. Two models not studied in FF92
5. Fundamental factor models use
6. Robust location estimator applications
7. Take away and open questions

Appendix

1. Robust Regression Overview

Data Oriented Viewpoint

- Not much influenced by outliers
- A good fit to the bulk of the data
- Reliable multi-D outlier detection

Provides a diagnostic check on classical estimates

- **Tukey (1979)**

  "… It is perfectly proper to use both classical and robust methods routinely, and only worry when they differ enough to matter. But when they differ, you should think hard."

Least Squares (LS)

$$r_i = \mathbf{x}_i'\boldsymbol{\theta} + \varepsilon_i, \quad i = 1, \cdots, n$$

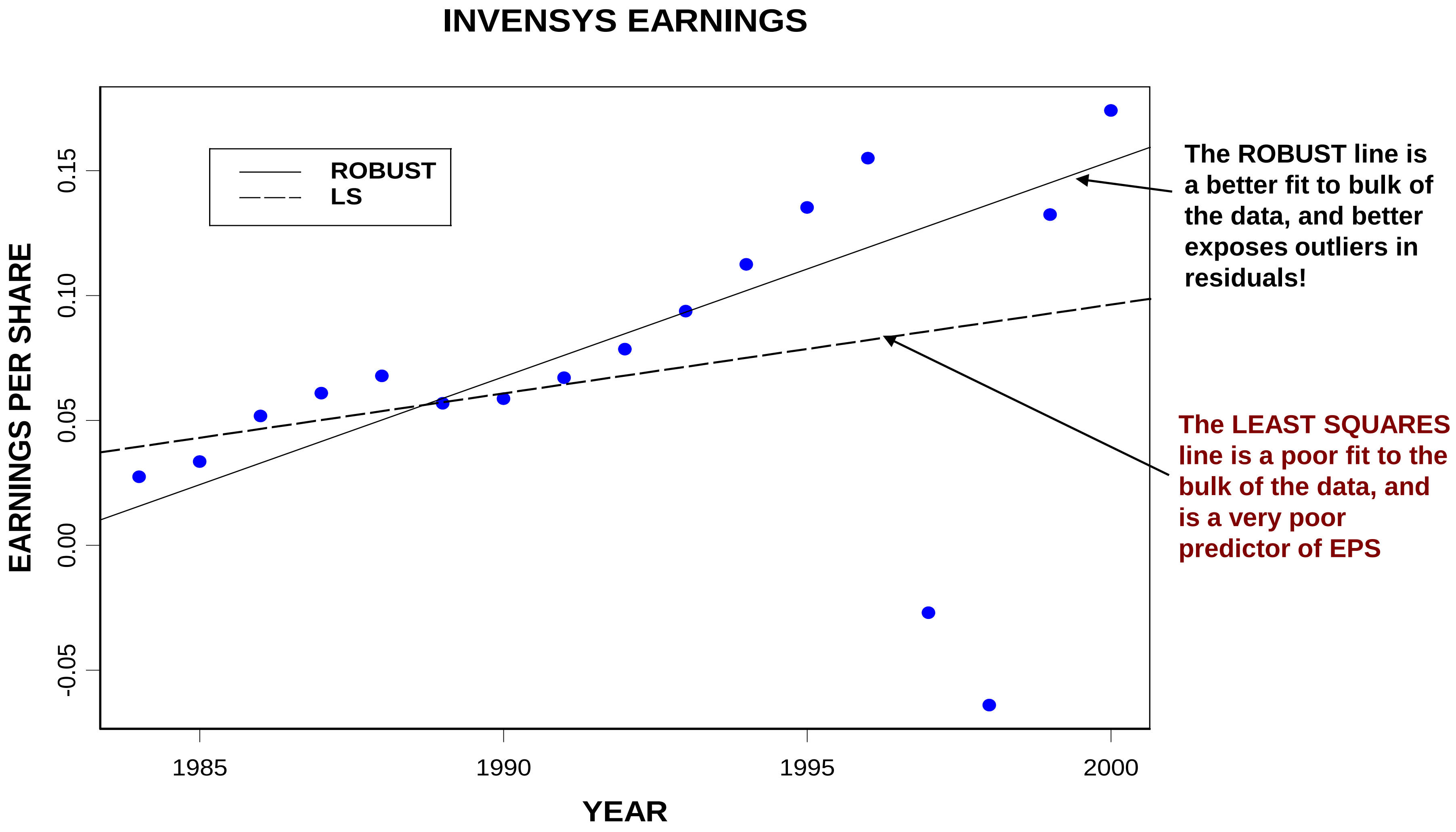
$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(r_i - \mathbf{x}_i'\boldsymbol{\theta} \right)^2$$

- Maximum-likelihood estimator (MLE) for normally distributed ε_i
- Best linear unbiased estimator (BLUE)   (so what?)

LS is Totally Lacking in Robustness Toward Outliers

- Returns often have non-normal distributions and outliers
- Linearity is not at all enough to cope with outliers
- Outliers in r_i and/or $\mathbf{x}_i$ can have arbitrarily large influence on $\hat{\boldsymbol{\theta}}$

Robust vs. Least Squares Fits for EPS



Regression M-Estimators

“M” = maximum-likelihood type: Huber (1964, 1973)

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \rho\left(\frac{r_{\mathrm{i}} - \mathbf{x}_i'\boldsymbol{\theta}}{\hat{s}_o}\right)$$

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathbf{x}_{\mathrm{i}} \cdot \psi\left(\frac{r_{\mathrm{i}} - \mathbf{x}_i'\hat{\boldsymbol{\theta}}}{\hat{s}_o}\right) = 0 , \qquad \psi = \rho\,'$$

“MM-estimators”: M-estimators with a highly robust initial estimator – crucial for non-convex ρ, Yohai (1987).

Robustness Goals for Choice of ρ and ψ

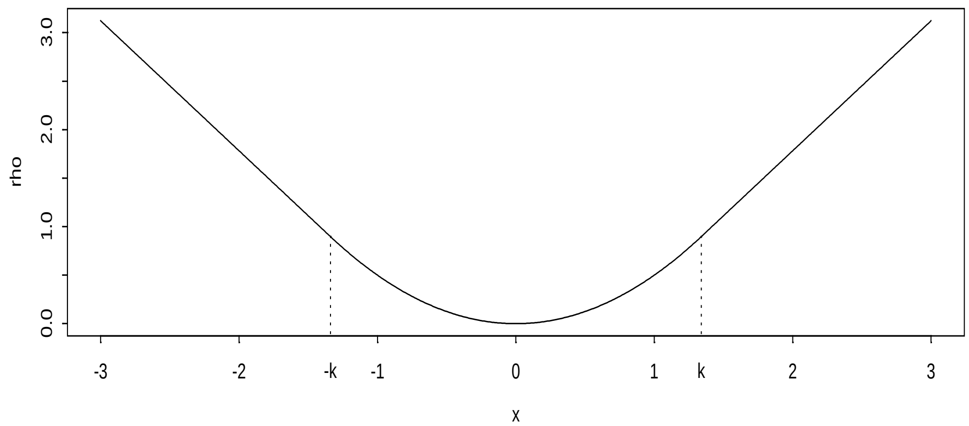
Two Key Simultaneous Goals

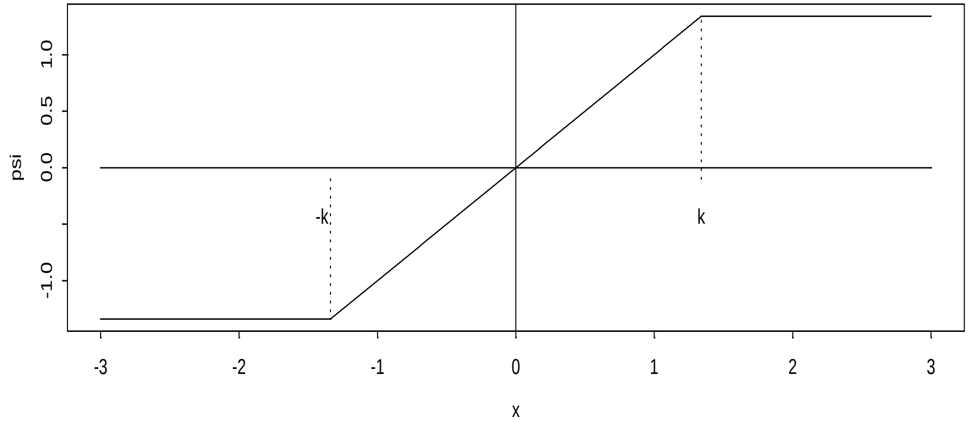
- Estimator variance is only very slightly larger than that of LS in the case of normal distributions (high normal distribution "efficiency")

- The maximum estimator bias due to outliers is minimized

See Appendix for sketch of the theory and references to detailed theory in MMYS (2019).

Huber Optimal ρ and ψ





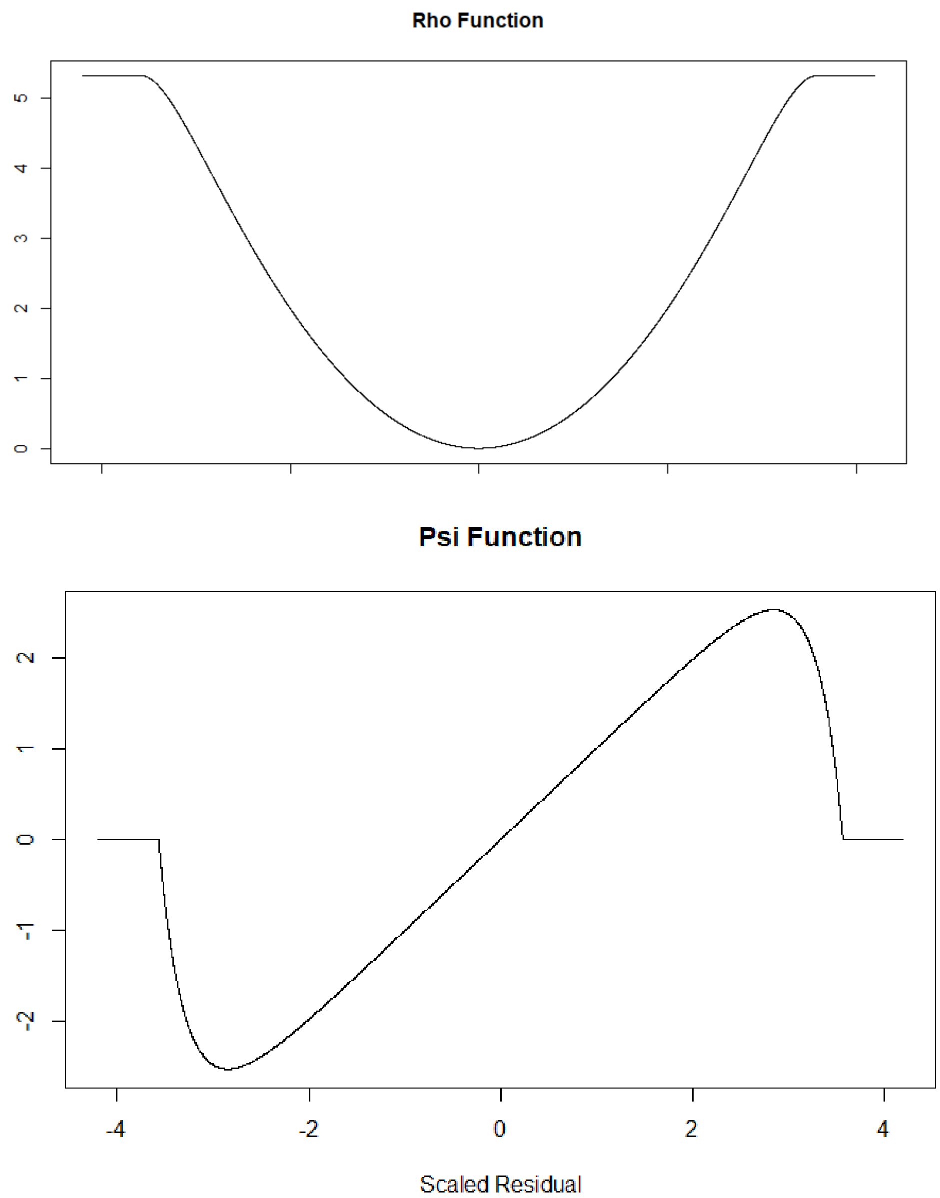
Good news: It is a convex optimization problem, and minimizes maximum variance

Bad news: It can result in arbitrarily large bias for the Tukey-Huber model (Martin, Yohai & Zamar, 1989).

None-the-less, is better than not using any robust regression method

N.B. Axioma uses this robust regression estimator, see., e.g., Axioma paper 062 (2015), and Axioma AXWWW21-1 (2015).

Yohai-Zamar-Svarc Optimal ρ and ψ



<u>A non-convex optimization</u> problem. But a very reliable MM-estimator algorithm exists for finding global minimum.

See MMYS (2019) for details, Section 5.8.1 for formula for psi function.

99% normal distribution efficiency version, with **smooth outlier rejection**, rejects outliers for which:

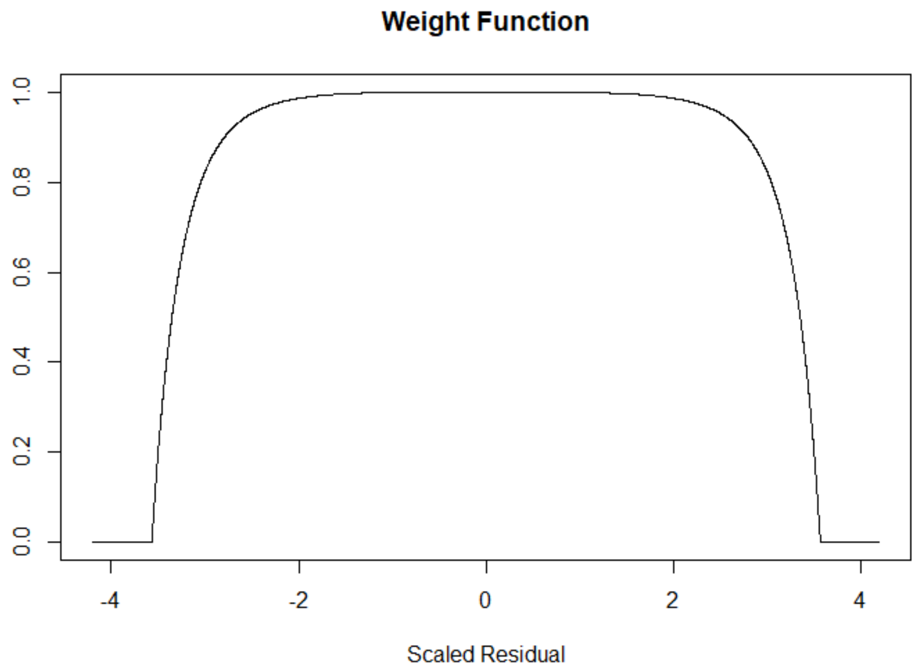
$$\frac{|r_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}|}{\hat{s}_o} > 3.568$$

Weighted Least Squares Version of Estimator

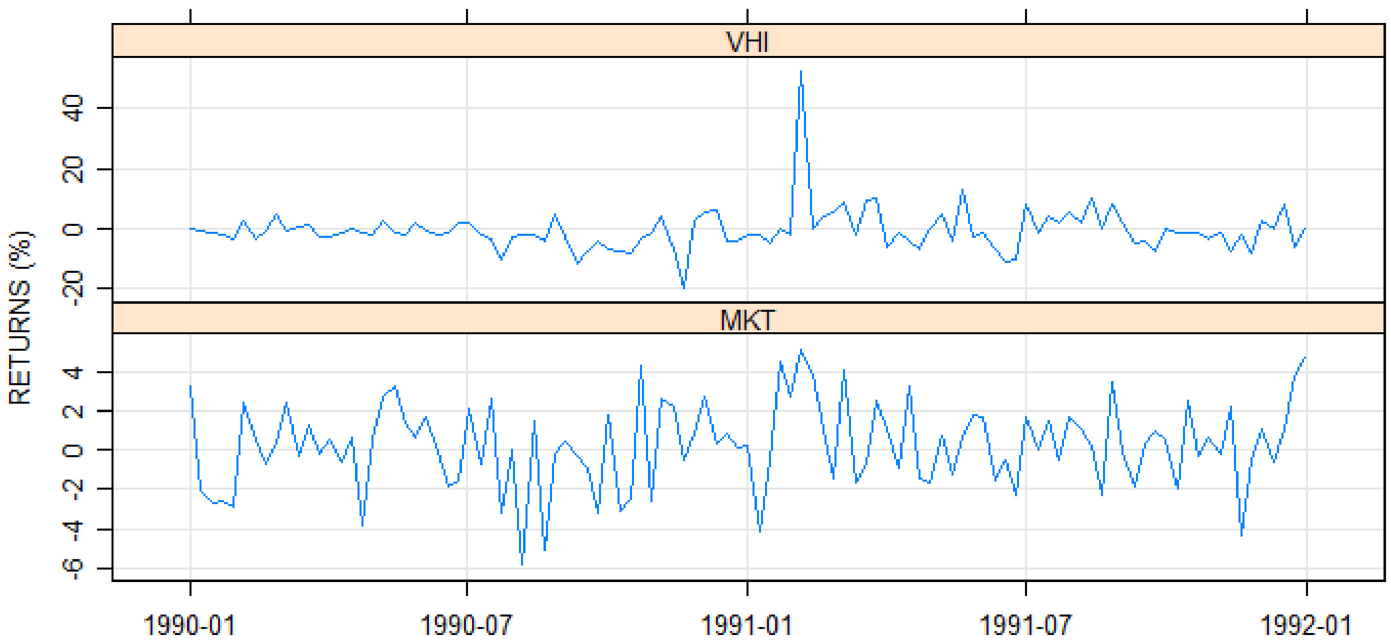
$$\sum_{i=1}^{n} \mathbf{x}_i \cdot W_{opt}\left(\frac{r_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}}{\hat{s}_o}\right)\cdot\left(r_i - \mathbf{x}_i'\hat{\boldsymbol{\theta}}\right) = 0$$

$$W_{opt}(t) = \frac{\psi_{opt}(t)}{t}$$

Uses a very robust initial estimate, and solves by iterative re-weighting.

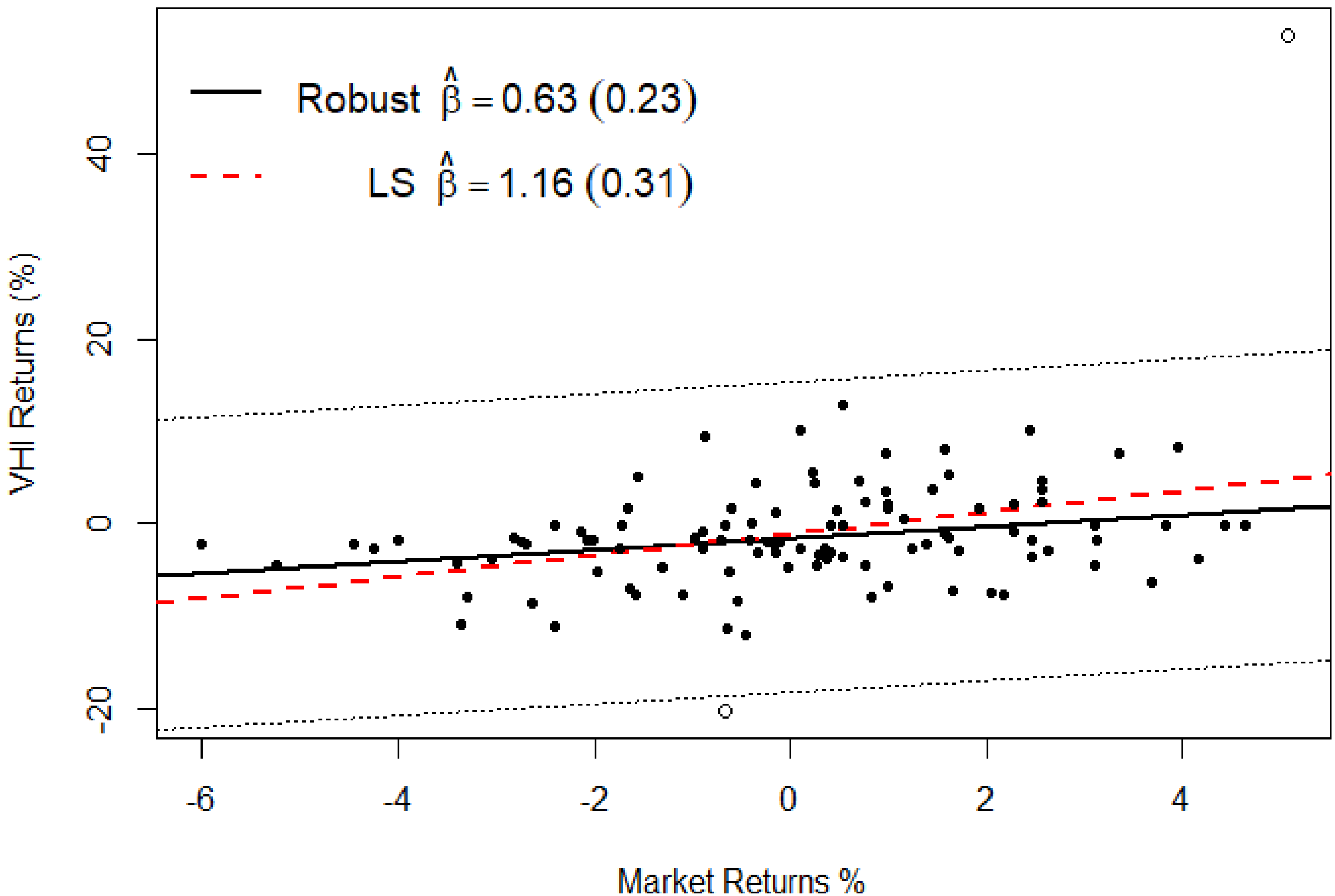


Example 1: Single-Index Model for VHI



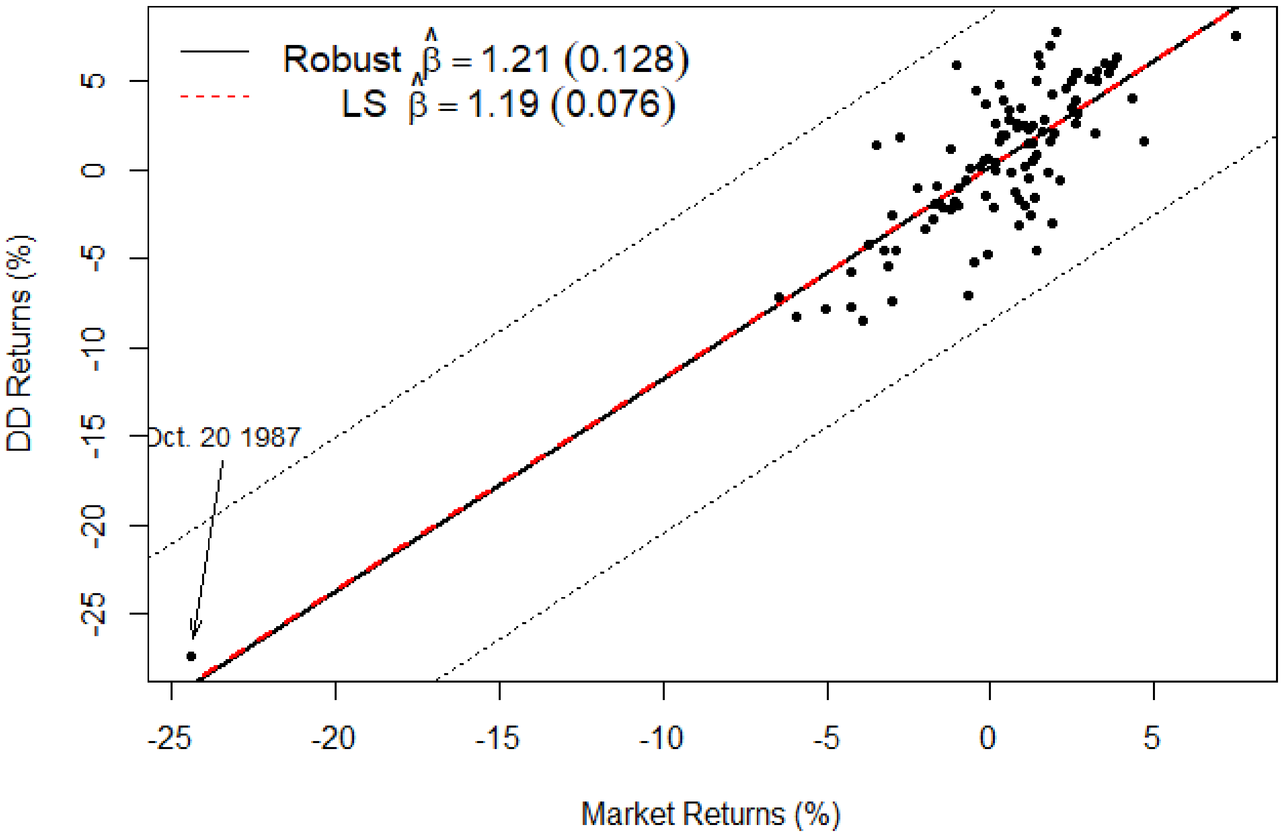
Code for the above plot, and the plots on the next two slides, is provided in the Appendix.

- LS beta is almost twice the robust beta
- Robust beta standard error is smaller than that of LS beta



Example 2: Single-Index Model for VHI

Important fact: "good" outliers are not rejected



2. Fama French 1992 Results

Eugene F. Fama and Kenneth R. French (1992). "The Cross-Section of Expected Stock Returns", *Journal of Finance*.

For an overview of empirical asset pricing, including brief discussion of research on many pricing anomalies, see:

Bali, Engle and Murray (2016). *Empirical Asset Pricing: The Cross-Section of Stock Returns*, Wiley

Cross-Section Regression Models

$$\mathbf{r}_t = \mathbf{X}_{t-1}\boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \cdots, T$$

Factor exposures

Regression slopes

Least Squares (LS) Fitted Models

$$\mathbf{r}_t = \mathbf{X}_{t-1}\hat{\boldsymbol{\theta}}_t + \hat{\boldsymbol{\varepsilon}}_t \qquad \hat{\boldsymbol{\theta}}_t = \left(\hat{\theta}_{1,t}, \hat{\theta}_{2,t}, \cdots, \hat{\theta}_{K,t}\right)'$$

t-Tests of Significance

Sample mean of time-series of slopes $\hat{\theta}_{k,t}, \quad t = 1, 2, \cdots, T$

Fama-French 1992 Goal

Determine which of the factors below explain the cross-section of expected returns (which factors “price risk”)

β	CAPM beta (special portfolios to reduce EV)
ln(ME)	Size (ME is market equity in $M)
ln(BE/ME)	Book-to-Market (often just B/M)
E(+)/P	Positive Earnings to Price
E/P Dummy	Negative Earnings to Price Dummy
ln(A/ME), ln(A/BE)	Leverage factors (A = book assets)

Fama-French (1992) Table III - 1963-1990

Ignore

strange

β	ln(ME)	ln(BE/ME)	ln(A/ME)	ln(A/BE)	E/P Dummy	E(+)/P
0.15 (0.46)						
	−0.15 (−2.58)					
−0.37 (−1.21)	−0.17 (−3.41)					
		0.50 (5.71)				
			0.50 (5.69)	−0.57 (−5.34)		
					0.57 (2.28)	4.72 (4.57)
	−0.11 (−1.99)	0.35 (4.44)				
	−0.11 (−2.06)		0.35 (4.32)	−0.50 (−4.56)		
	−0.16 (−3.06)				0.06 (0.38)	2.99 (3.04)
	−0.13 (−2.47)	0.33 (4.46)			−0.14 (−0.90)	0.87 (1.23)
	−0.13 (−2.47)		0.32 (4.28)	−0.46 (−4.45)	−0.08 (−0.56)	1.15 (1.57)

beta **does not** price risk !

returns are **negatively** related to size

returns are **positively** related to BE/ME

3. FF92 Redux with Robust Regression

For the vast majority of the stocks (~ 97-99%) we found:

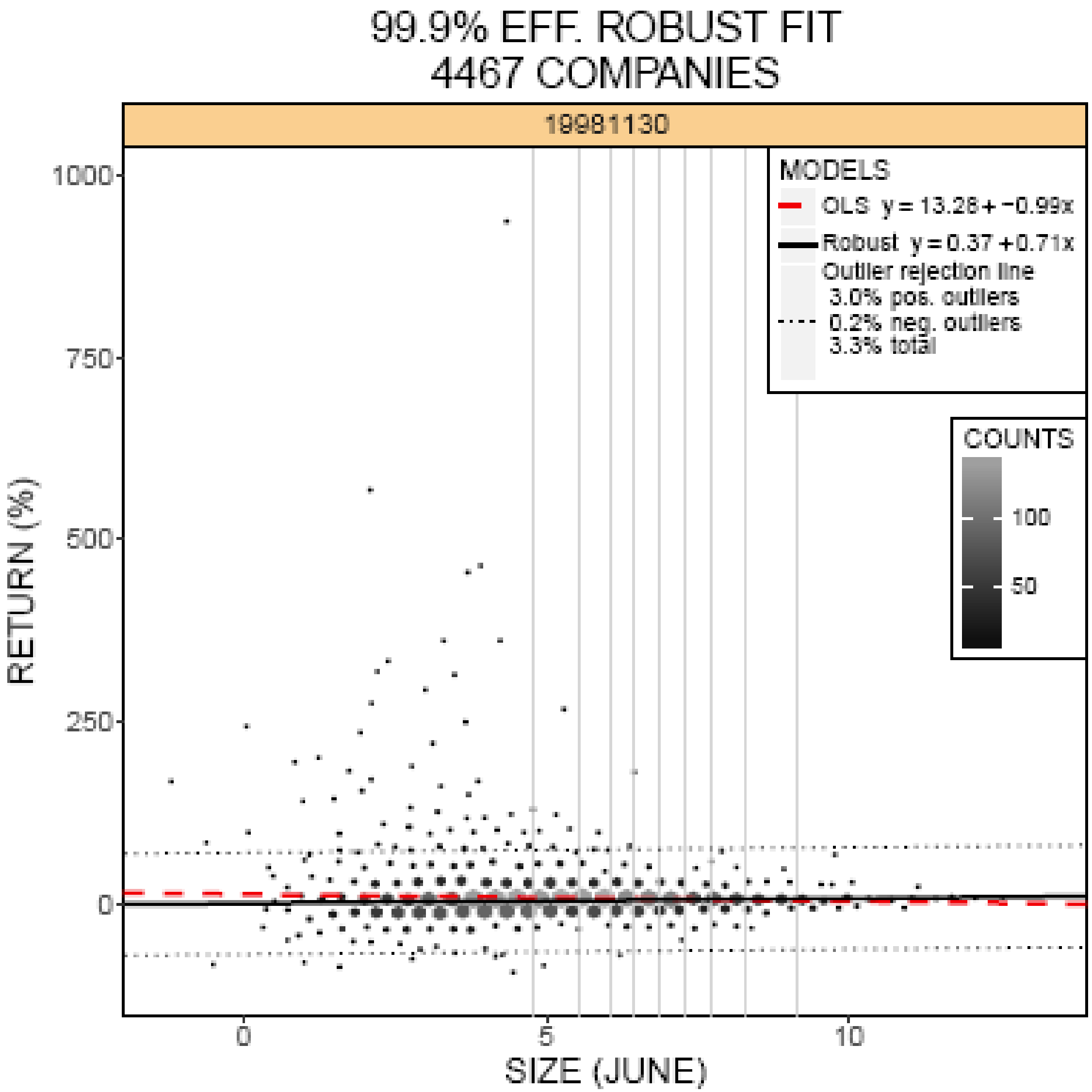
- **Different conclusions with robust regression than FF92:**
  - Equity returns are positively related to firm size
  - Beta relationship is significant and negative

- **New results for two models not in FF92:**
  - E/P prices risk
  - Beta and size interaction term

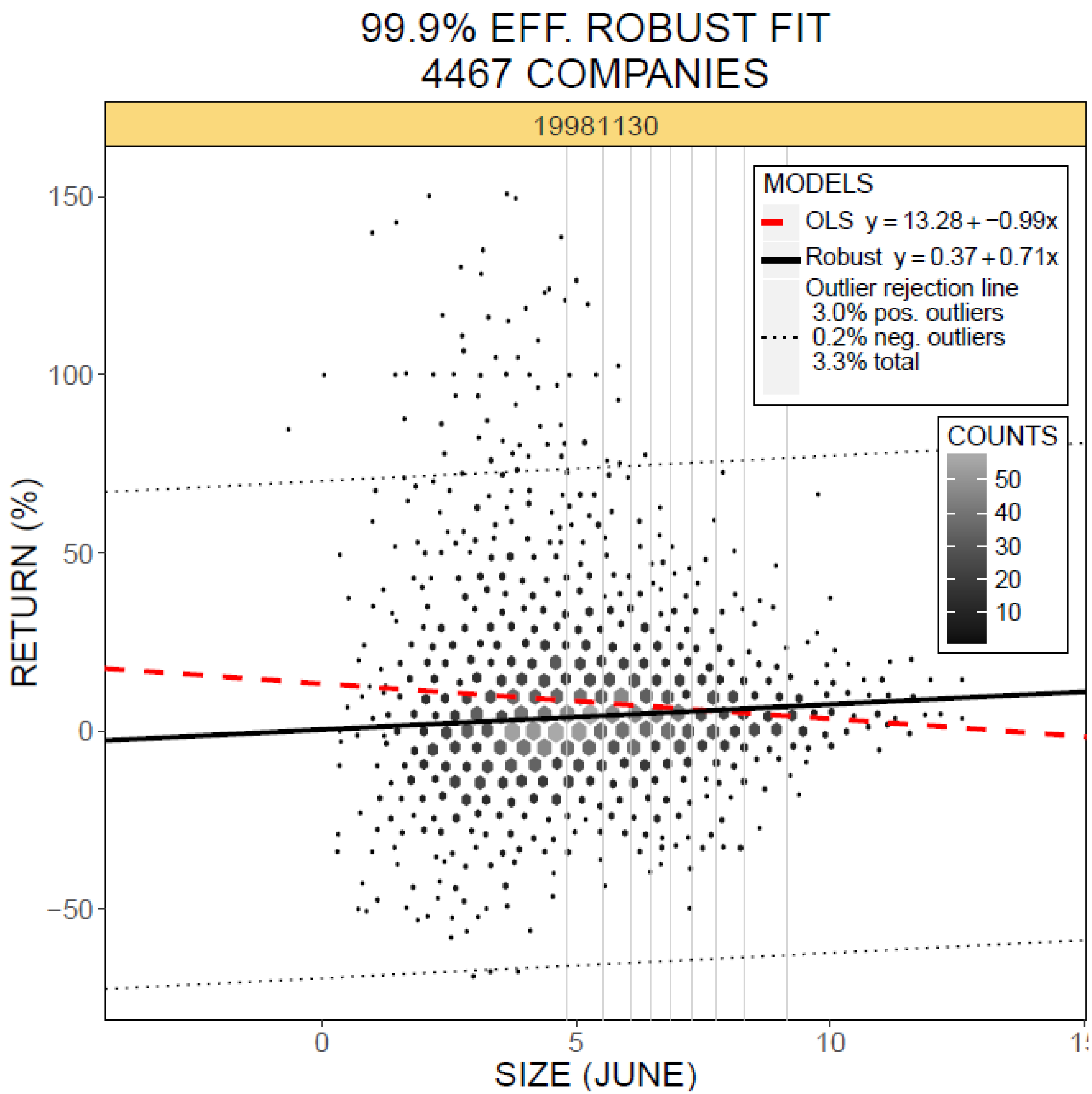
*Joint work with Christopher G. Green. See Green and Martin (2017), SSRN Abstract ID 2963855.

Nov. 1998 Returns vs Size LS & Robust Fits

Full vertical range view



Same data after .1% vertical trimming for a better view



Returns vs Size

Factor	Method	1963–1990		1963–2015		1980–2015	
Size	LS (FF92)	−0.15	(−2.58)				
	LS (GM)	−0.13	(−2.33)	−0.14	(−3.45)	−0.10	(−2.25)
	LTS 5% (CCW04)	0.22	(5.79)				
	LTS 1% (KR97)	0.14	(2.63)				
	Robust (GM)	0.21	(4.01)	0.28	(8.44)	0.39	(11.48)

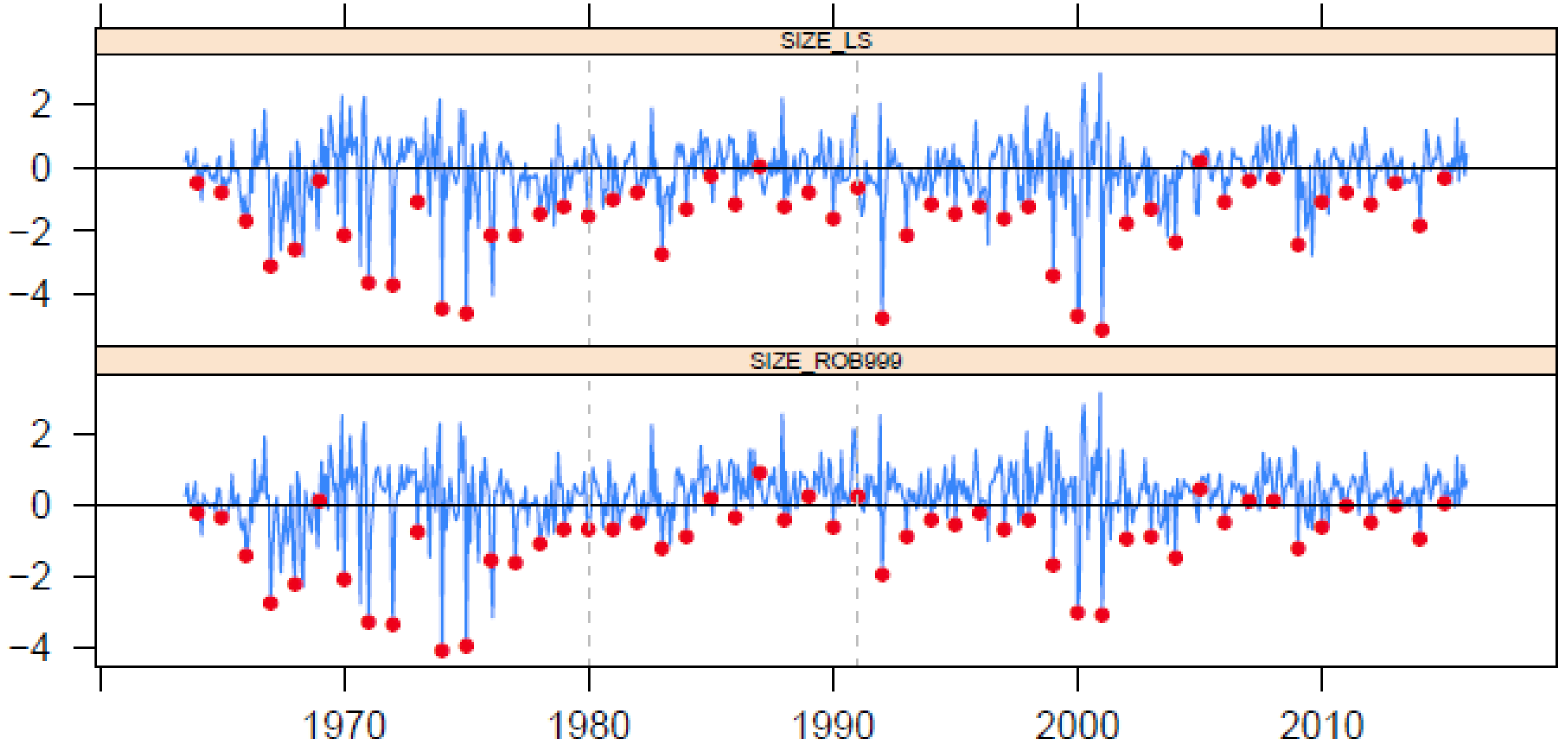
huge t-stats

Mean % Outliers Rejected = 1.54% (= median in this case)

KR97 = Knez & Ready (1997)
CCW04 = Chou, Chou & Wang (2004)

LTS = least trimmed squares

Monthly Slopes of Returns Regressed on Size

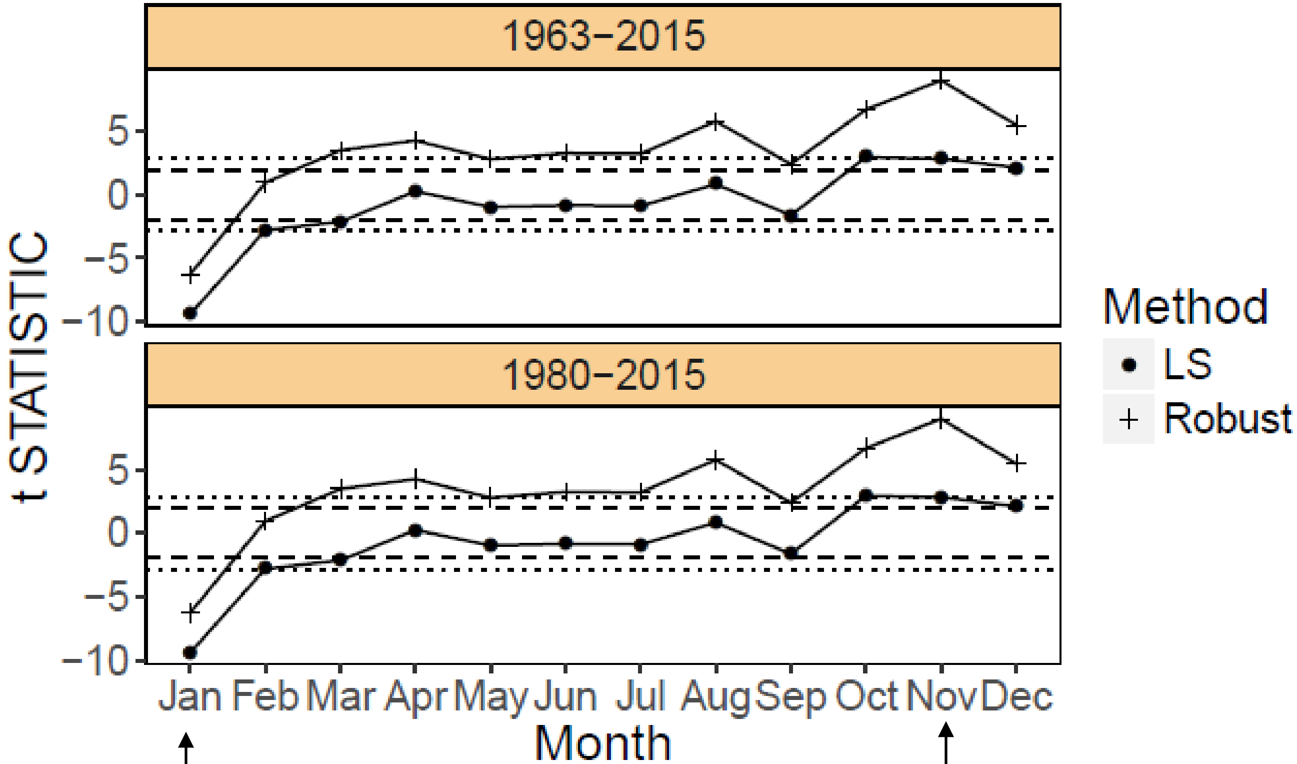


red dots illustrate a well-known January size effect

N.B. Existence of outliers and serial correlation, thus one should use a robust location estimator with HAC: Croux et al. (2003).

Monthly Analysis of Returns vs Size

rejection regions with and without multiple comparisons adjustment



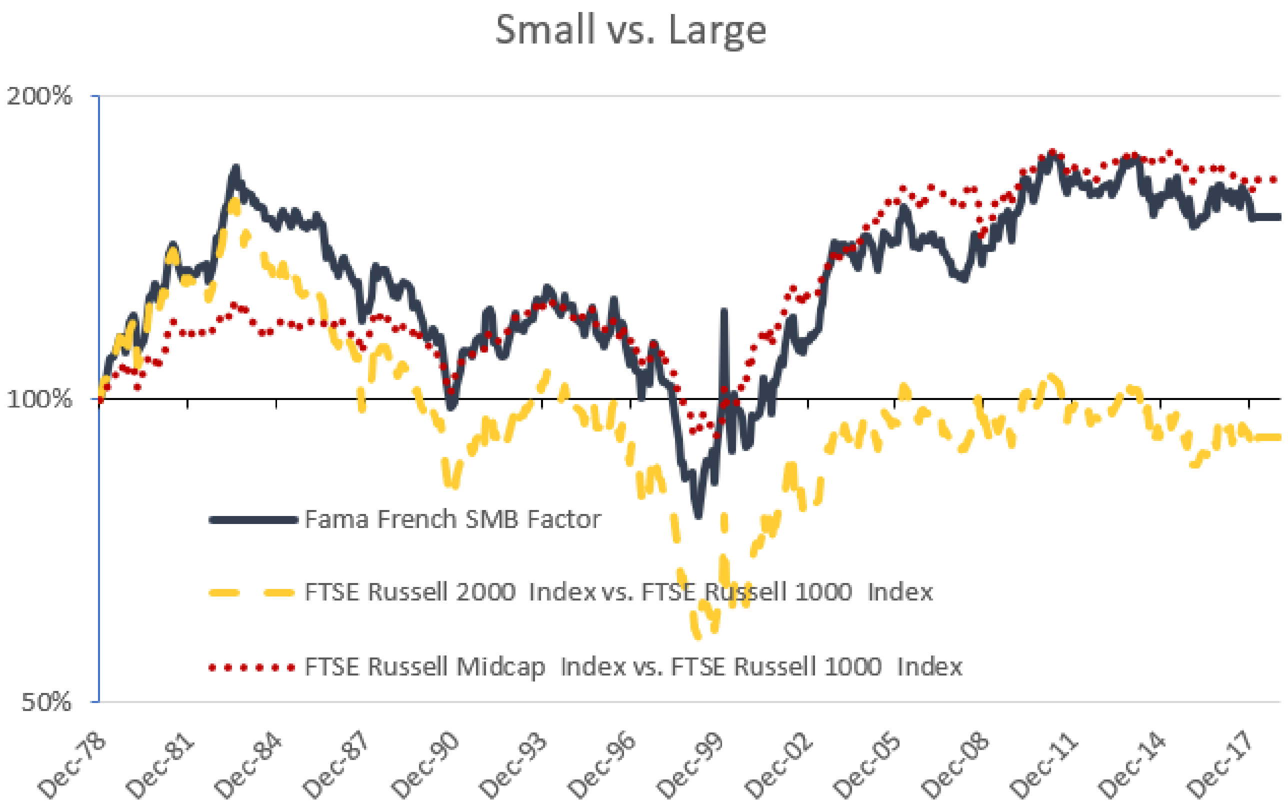
well-known
January effect

A not-so-well-known
Q4 effect ??

Advice for Evaluating Factor Premia

Tom Philips:

“Attempt to replicate the returns of a factor using publicly available indices, preferably ones that discard the bottom 5%-10% of the market’s total capital. Such a replication allows the investigator to determine if a strategy is tradeable, and also real-time permits performance monitoring.”



Returns vs Beta

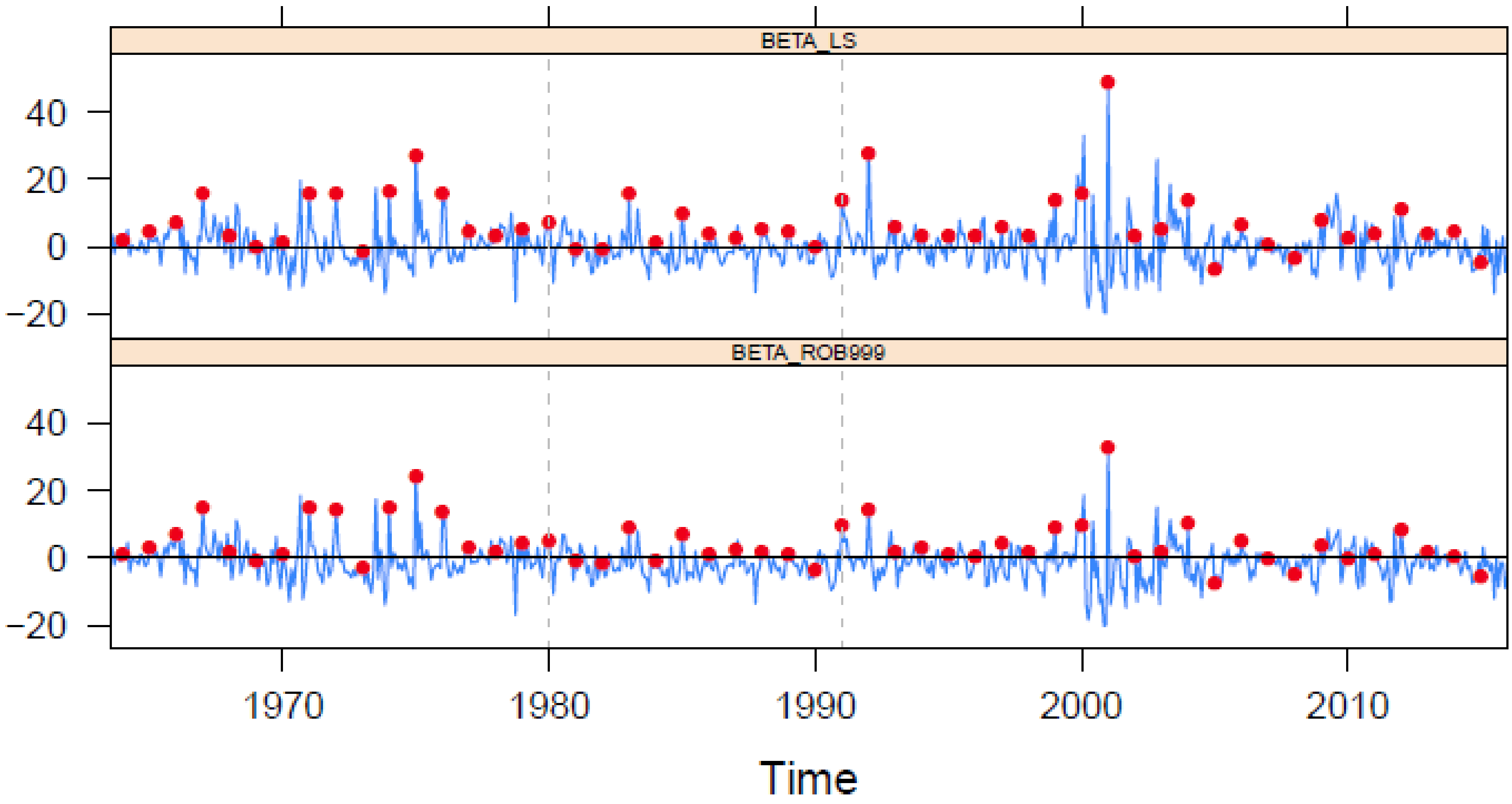
Factor	Method	Beta 1963–1990		1963–2015		1980–2015	
Beta	LS (FF92)	0.15	(0.46)				
	LS (GM)	0.05	(0.14)	0.16	(0.64)	−0.08	(−0.25)
	LTS 5% (CCW04)	−0.40	(−1.73)				
	Robust (GM)	−1.08	(−3.47)	−1.25	(−5.72)	−1.70	(−6.66)

highly significant t-stats

Slopes of Returns Regressed on Beta

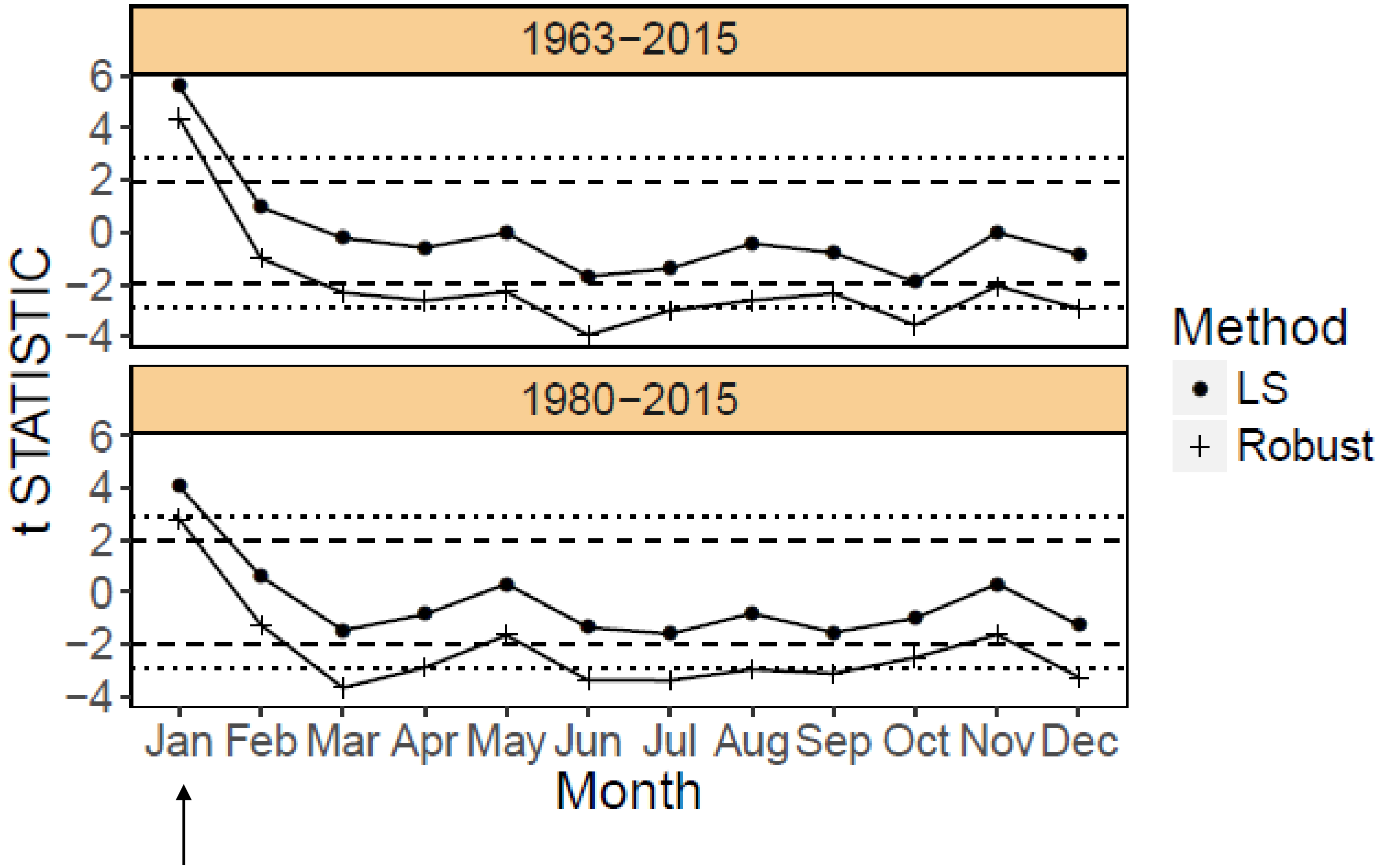
Monthly time series

red dots = Januaries



Monthly Analysis of Returns vs Beta

rejection regions with and without multiple comparisons adjustment



Beta January effect

4. Two Models Not Studied in FF92

Returns vs Earnings-to-Price

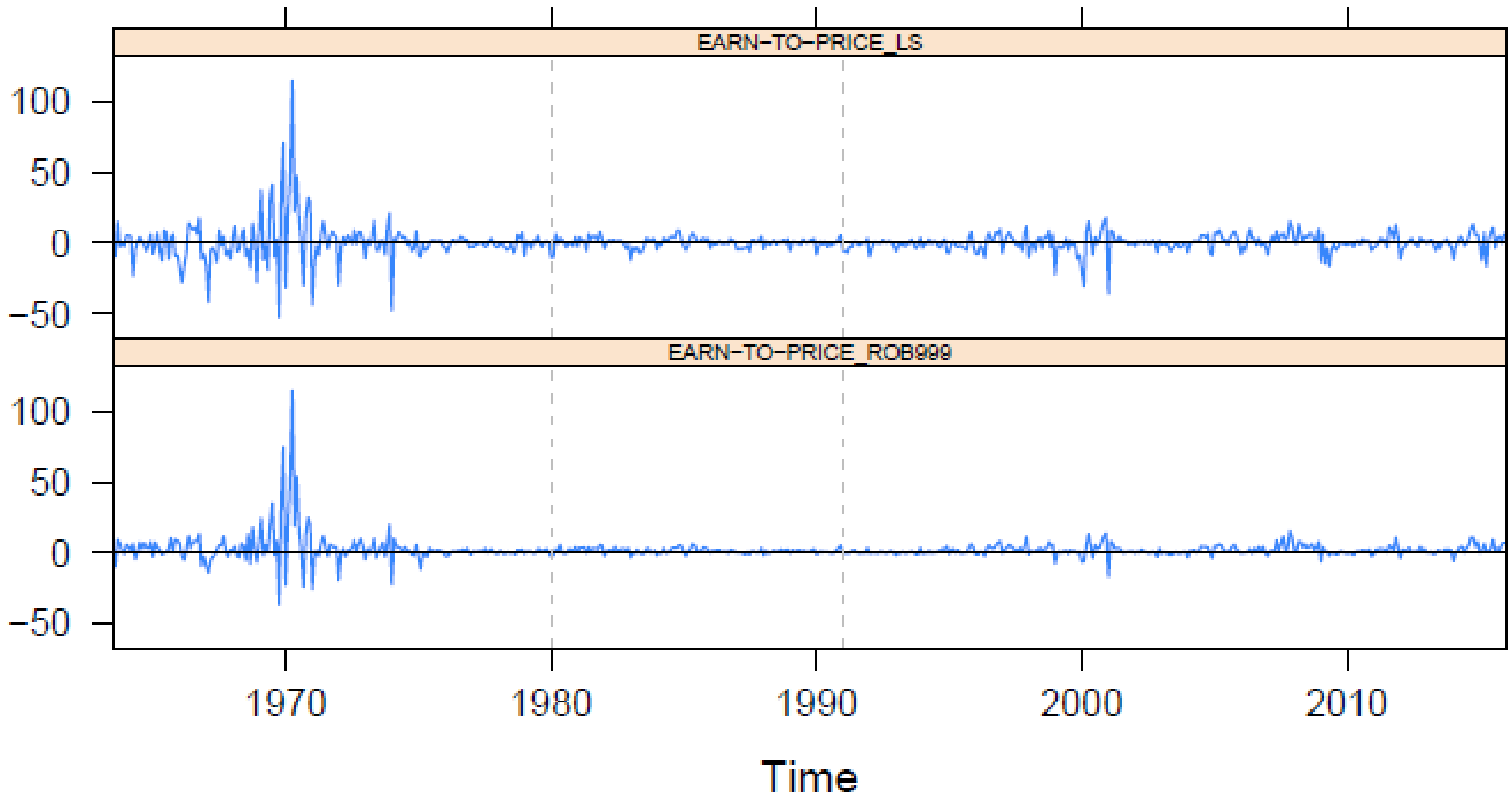
Factor	Regression	Location Estimate	t-Statistic	Unaltered Earnings-to-Price 1963–1990		1963–2015		1980–2015	
E/P	LS	mean	uncorrected	0.81	(1.14)	0.62	(1.52)	0.41	(1.57)
		robust	Croux et al.	0.32	(0.90)	0.56	(2.41)	0.70	(3.08)
	Robust	mean	uncorrected	2.92	(5.01)	2.47	(7.74)	1.86	(13.18)
		robust	Croux et al.	1.88	(9.73)	1.82	(13.74)	1.71	(15.46)

“uncorrected” = not corrected with Newey-West (should be done for classic t-test)

“Croux et al.” = Croux et al. (2003) standard error serial autocorrelation correction (AC)

N.B. Typically using a robust mean estimator of time series of slopes, and corresponding robust t-test will improve the power of the test, even without AC

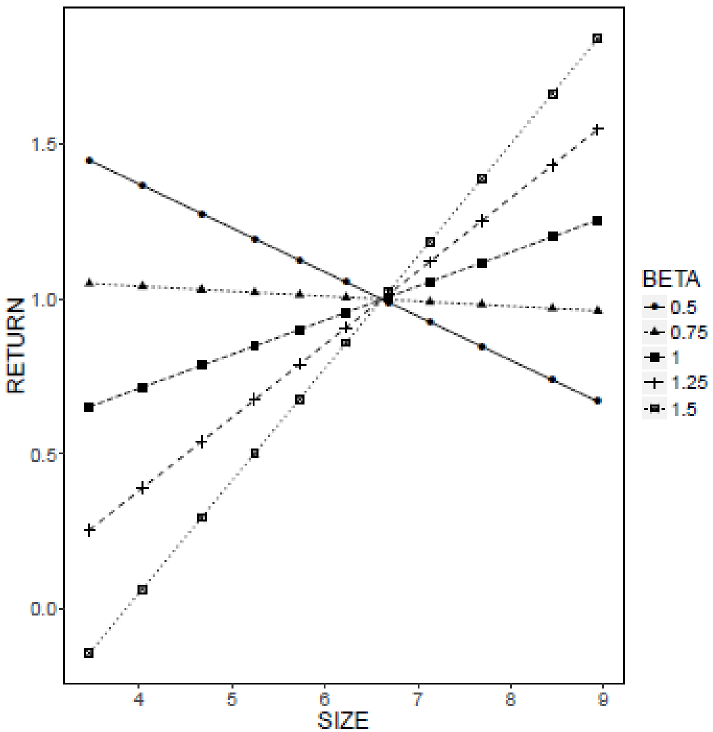
Slopes of Returns Regressed on E/P

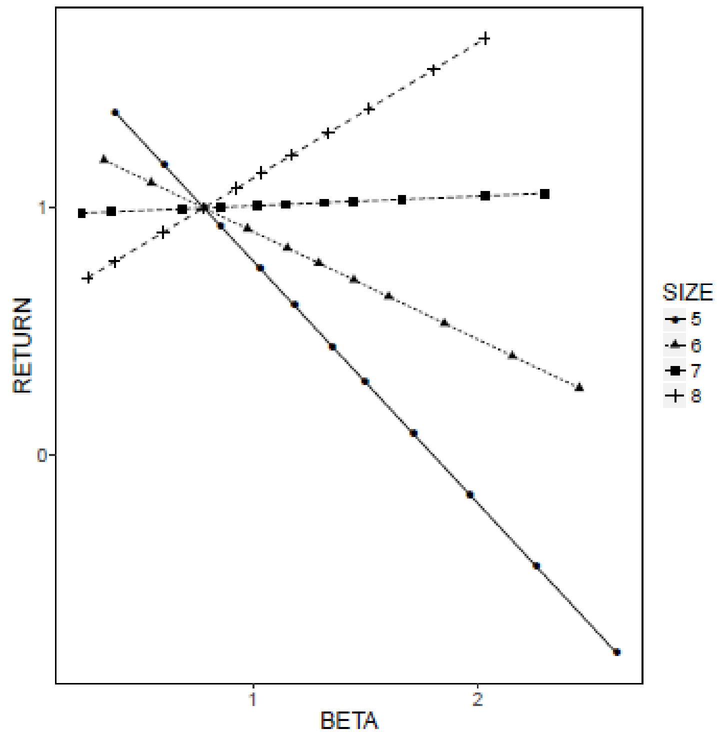


Size-Beta Interaction Model

With LS the only significant coefficient is interaction for 2 time periods, but Robust Regression coefficients are all highly significant for all 3 time periods.

$$returns = 3.63 - 3.54\times\beta - .4\times\text{SIZE} + .51\times\beta\times\text{SIZE} + noise$$





SIZE = ln(ME): (5, 6, 7, 8) = ($148M, $403M, $1.1B, $3.0B)

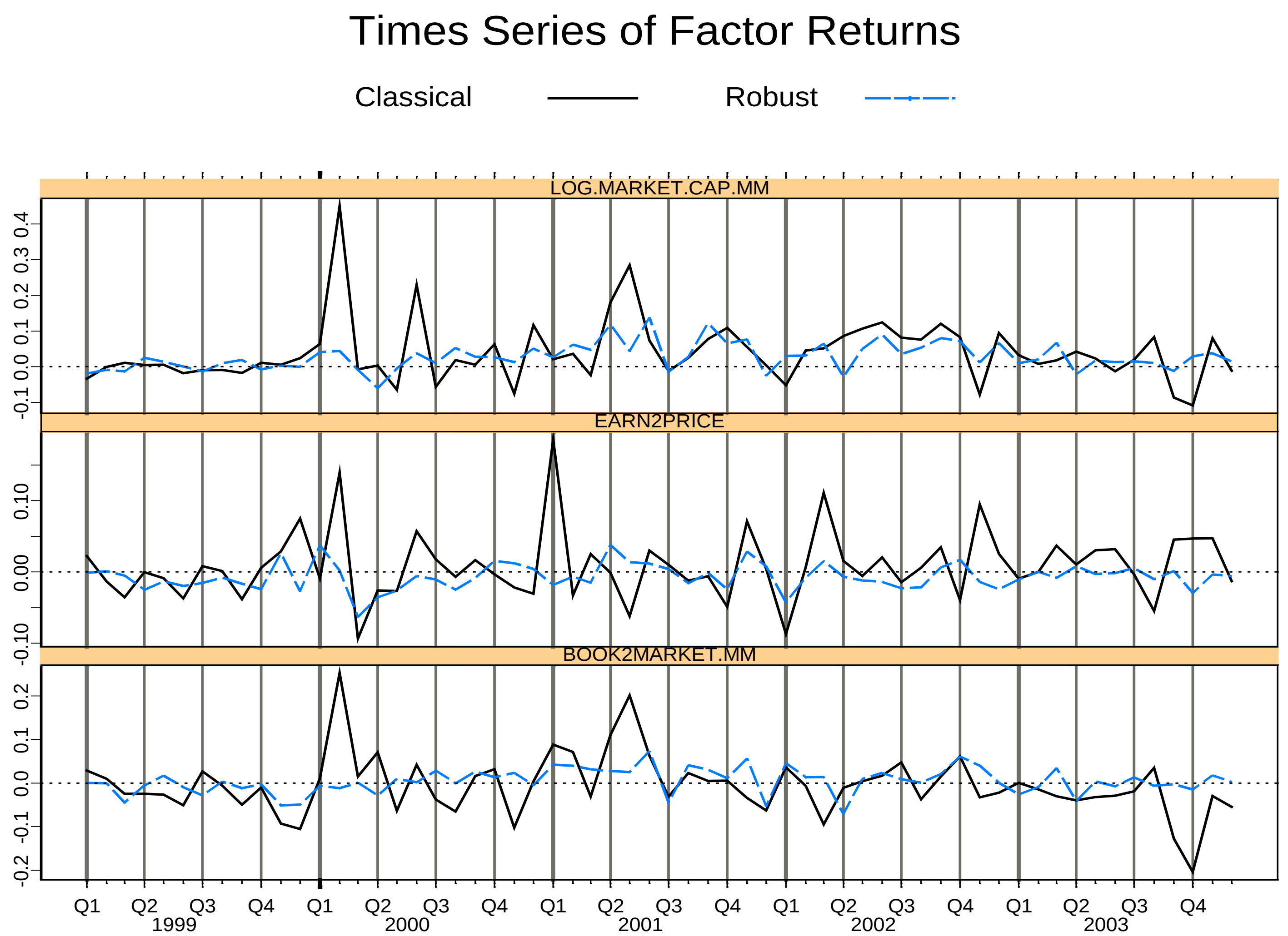
5. Fundamental Factor Models Use

- Axioma has responded to the need for robust regression in fundamental factor models by using Huber M-estimator
- Outliers abound in returns and in factor exposures, more so in the latter than one may think
- Price paid for using LS is more volatile factor returns and cross-section correlation in residuals. The former can result in over-stating the factor contribution to risks.

The following two slides illustrate the last point.

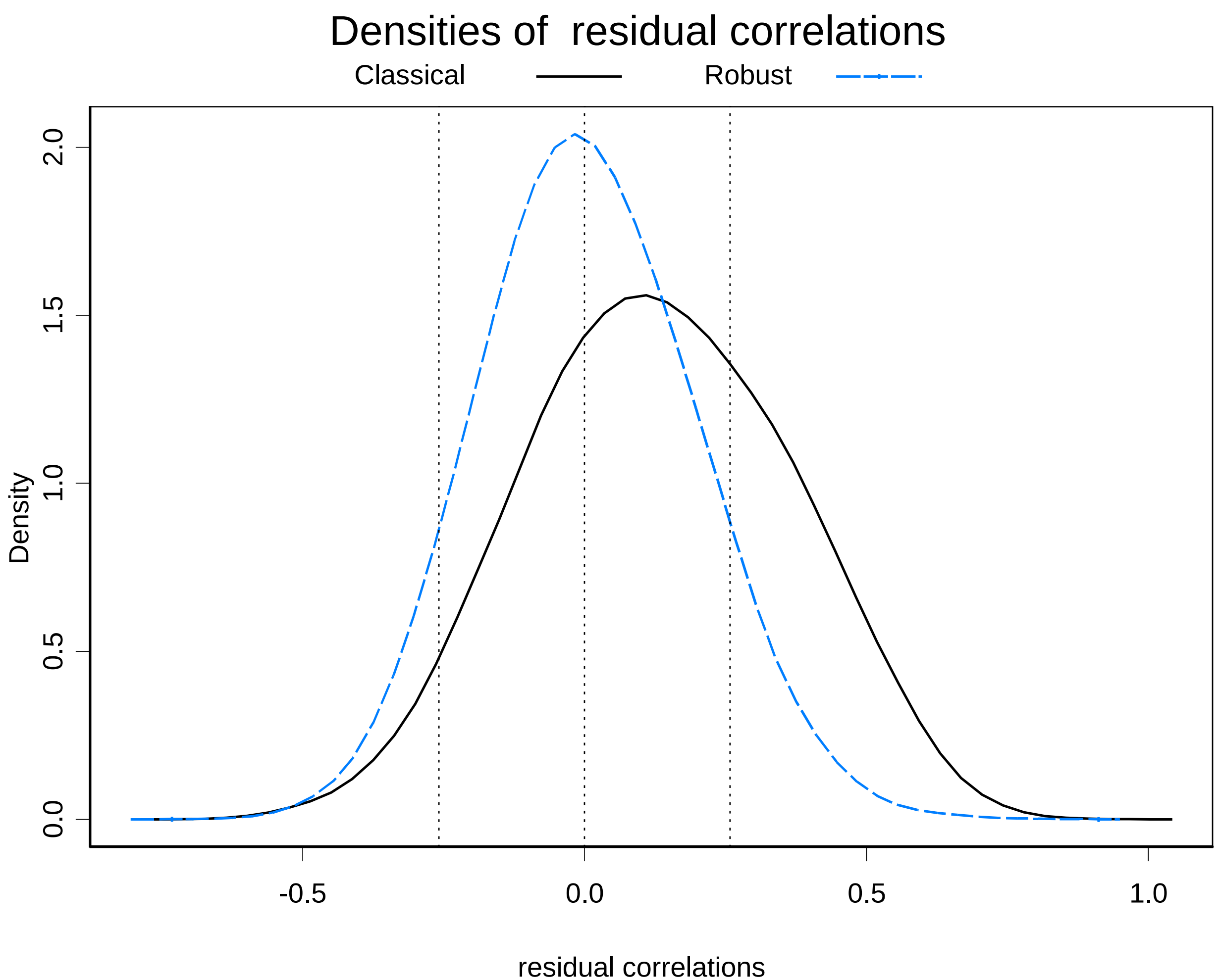
Robust versus Classical Factor Returns

Three factors: size, E/P, B/M, monthly returns



Residuals Cross-Section Correlations

LS results in positive average of residuals cross-correlations, which does not happen with robust regression



6. Robust Location Estimator Apps

- Special case of robust regression with intercept only
- Robust deciles analysis of expected returns to factors
- Outlier cleaning for risk & performance estimation

Location M-Estimator

The model: $$r_t = \mu + s \cdot \varepsilon_t , \qquad t = 1, 2, \cdots, n$$

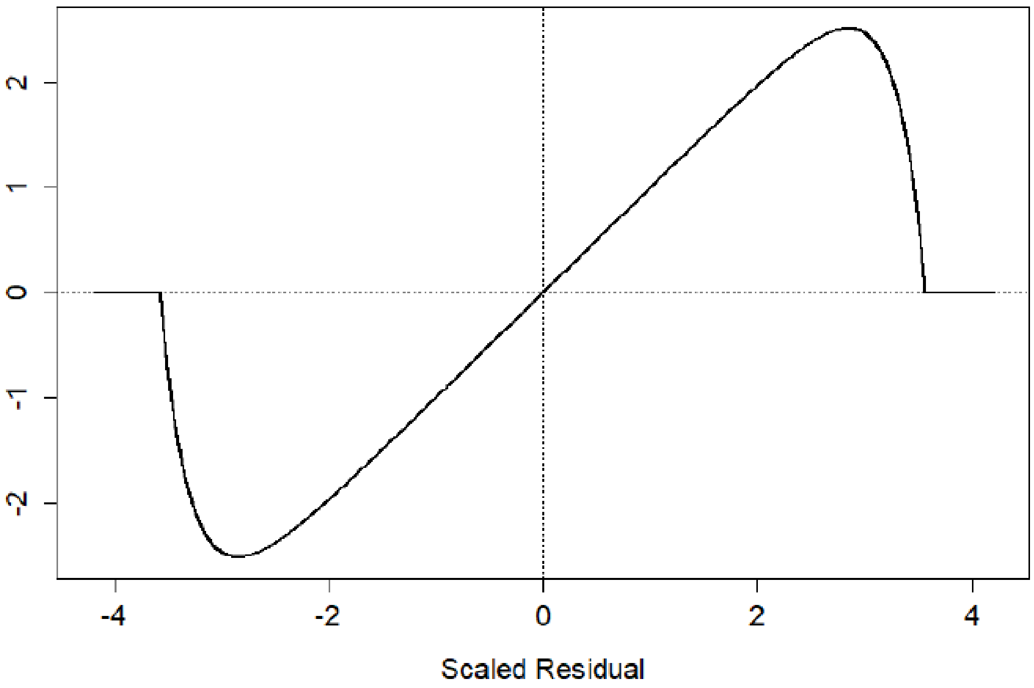
$$\hat{\mu} = \operatorname{argmin}_{\mu} \sum_{t=1}^{n} \rho\left(\frac{r_t - \mu}{\hat{s}} \right)$$

$$\sum_{t=1}^{\mathrm{n}} \psi\left(\frac{r_t - \hat{\mu}}{\hat{s}} \right) = 0 , \qquad \psi = \rho\,'$$

Very easy to solve! See Section 2.8.1 of MMYS.

The R function in RobStatTM: **`locScaleM()`**

ψ with 99% Normal Distribution Efficiency



Smooth **outlier rejection:**

Rejects data for which:

$$\frac{|r_i - \hat{\mu}|}{\hat{s}} > 3.568$$

Formula for psi function in MMYS Section 5.8.1

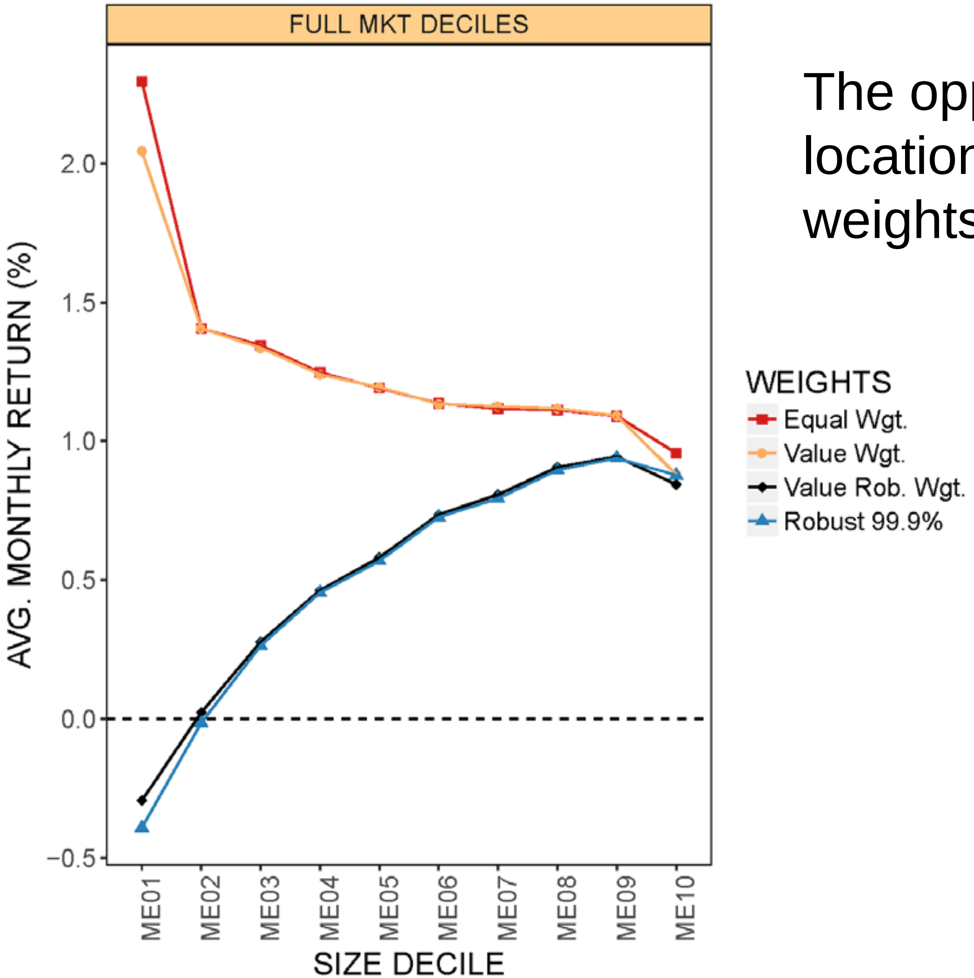
Virtues of this Location Estimator

- Fraction of outliers trimmed is data adaptive
- Can reject outliers asymmetrically, e.g., if more positive outliers than negative outliers (and conversely)

Robust Expected Returns Factor Deciles Analysis

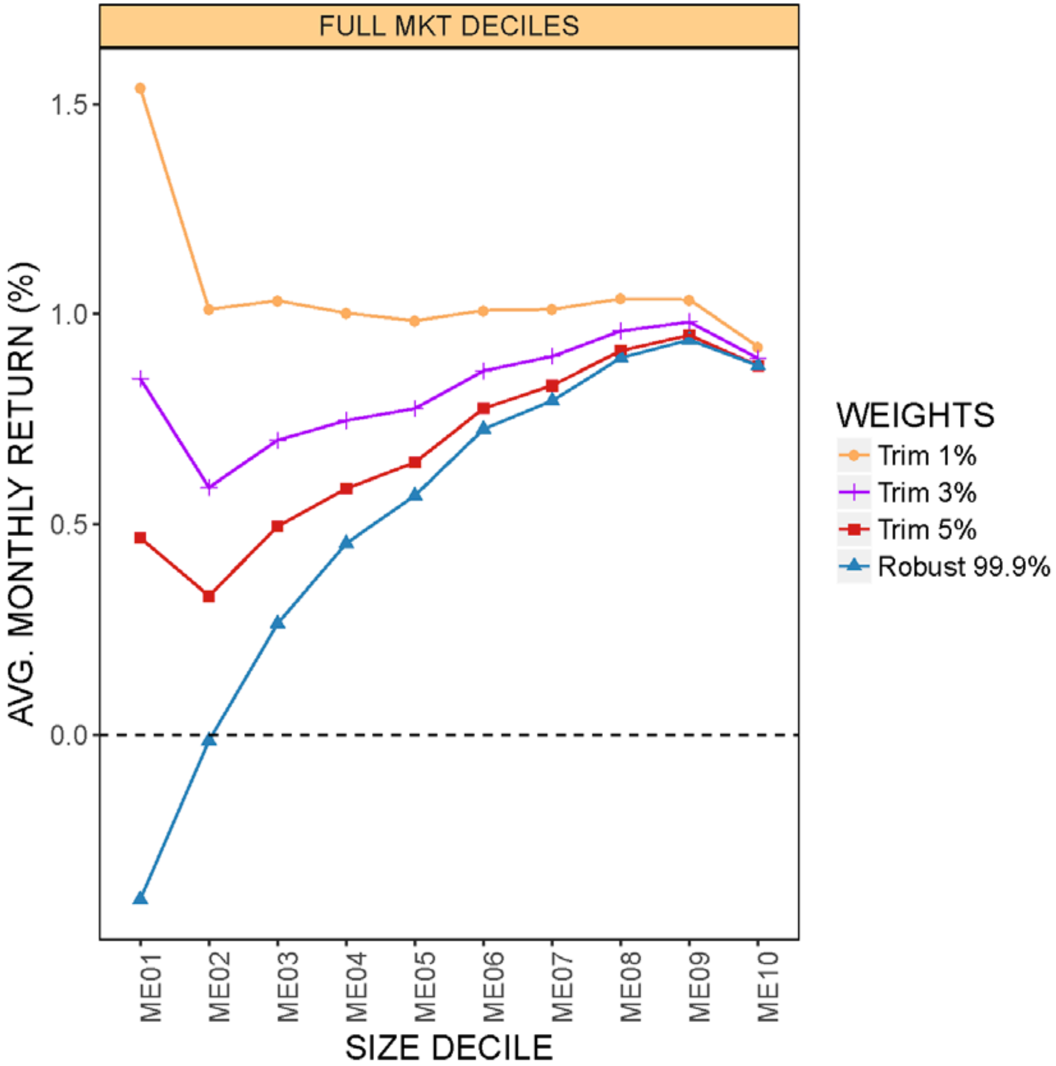
Reverses the common wisdom that “Returns decrease with firm size”

The opposite is true with robust location M-estimator portfolio weights that rejects outliers



from time series of decile portfolios

Trimming will Not Suffice !

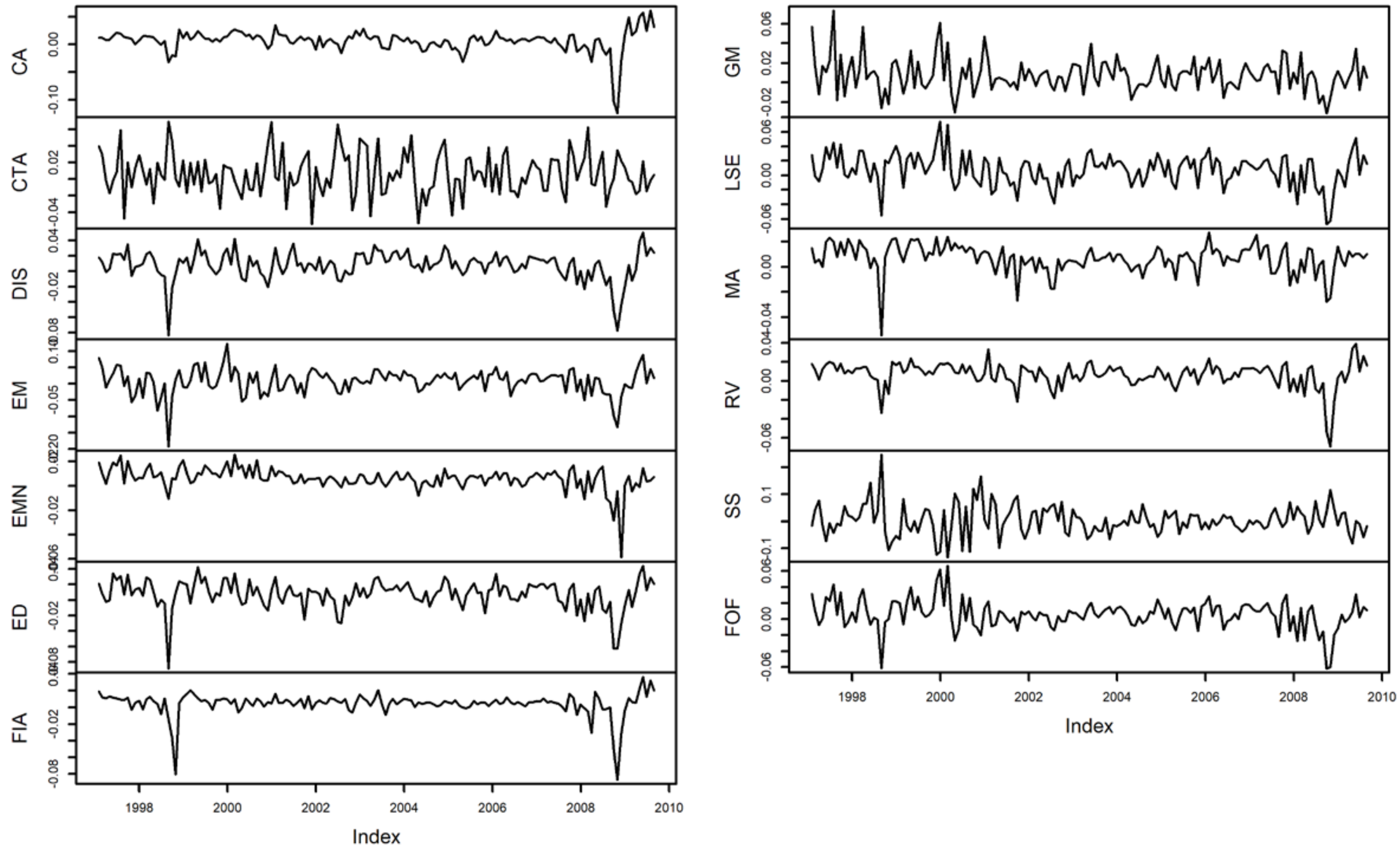


Limitations of Trimming and Winsorizing

- No data driven way of choosing trimming fraction
- Rigidly symmetric outlier treatment

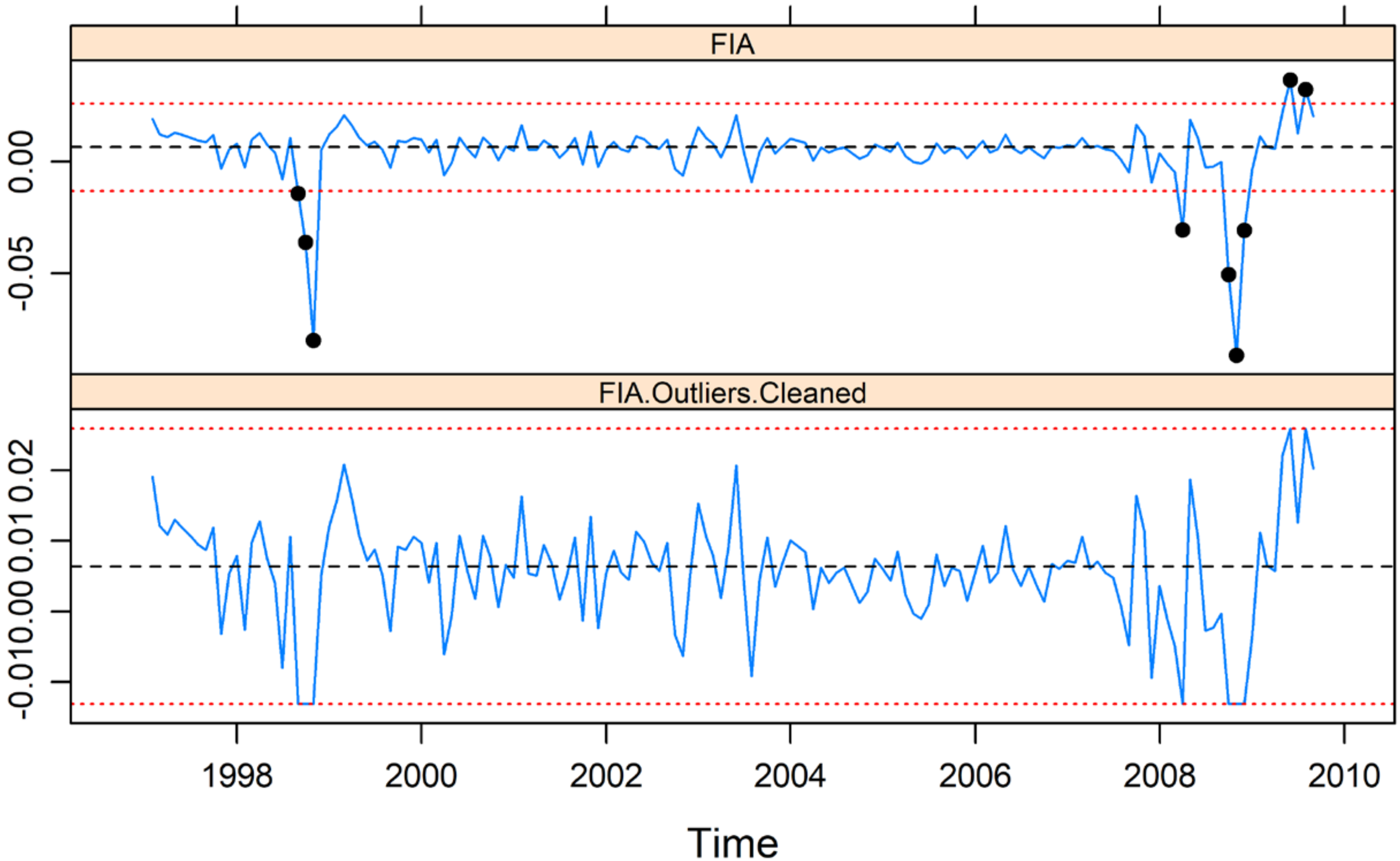
Risk & Performance Estimator Outlier Cleaning

You need to compute risk and performance estimators for the following hedge funds returns with outliers.



Application to the FIA Hedge Fund Returns

Automatic outlier detection and shrinkage



Outlier Impact on ES and SR Estimators

	ES	ES CL	seCorIF	seCorIF CL	SR	SR CL	seCorIFAdapt	seCorIFAdapt CL
FIA	-0.042	-0.013	0.025	0.002	0.299	0.761	0.185	0.148
CTAG	-0.045	-0.045	0.004	0.004	0.258	0.258	0.08	0.08

seCorIF and seCorIFAdapt are new estimator standard error computational method that is accurate when returns are serially correlated as well as uncorrelated.

Chen and Martin (2019). “Standard Errors of Risk and Performance Estimators with Serially Correlated Returns”, https://ssrn.com/abstract=3085672.

7. Take-Aways and Open Questions

Take-Aways

- Empirical asset pricing studies can benefit considerably by using robust regression (and other robust methods) as a complement to LS
- Fundamental factor model construction for portfolio optimization and risk management could similarly benefit.

Open Questions

- Connection with low-vol anomaly (Blitz & van Vliet, 20017, Baker et al., 2011)
- Outliers and Asness et al. (2015) "Quality-Junk" factor?
- Life-time and other properties of positive outliers?
- What is the full story about the negative beta relationship?
- Financial implications of the size-beta interaction model?

Appendix

Robust Regression Theory

MMYS (2019) Chap. 5

Tukey-Huber model for regression $r_i = \mathbf{x}_i'\boldsymbol{\theta} + \varepsilon_i$

$$(r_i, \mathbf{x}_i) \sim F = (1-\gamma)\cdot N(r, \mathbf{x}) + \gamma \cdot H(r, \mathbf{x})$$

multivariate normal distribution

any joint distribution

Robustness goals for estimator $\hat{\theta}_{ROB}$

When $\gamma = 0$: High efficiency $\longrightarrow$ $\mathrm{EFF}\left(\hat{\theta}_{ROB}\right) = \dfrac{\mathrm{var}\left(\hat{\theta}_{LS}\right)}{\mathrm{var}\left(\hat{\theta}_{ROB}\right)} = 99\%$

When $0 < \gamma < 1/2$: Minimize maximum bias of $\hat{\theta}_{ROB}$ over all $H(r, \mathbf{x})$

Main Large Sample Theory Results

- First result (Huber, 1964, 1973)
  - M-estimators that minimize maximum variance subject to symmetric distributions constraint
  - Lacks bias robustness (can have arbitrarily large bias)

- Important result

  Yohai & Zamar, 1997; Svarc, Yohai & Zamar (2002); MMYS Chap. 5.8.1
  - MM-estimator with high normal distribution efficiency and min-max bias over Tukey-Huber model.

- Even better result (Maronna & Yohai, 2015 ; Ch 5.9.3 MMYS)
  - Fully efficient DCML estimator

Example 1 R Code

```
library(devtools) # Needed to install PCRM
install_github("kecoli/PCRM") # Install PCRM
library(PCRM)  # Load PCRM
(names(retVHI))
ret12 = retVHI[,1:2]
tsPlot(ret12,cex = .8)

library(RobStatTM)  # Must first install from CRAN
x=(retVHI[,2]-retVHI [,3])*100
y=(retVHI[,1]-retVHI [,3])*100
fit.ls = lm(y~x)
ctrl = lmrobdet.control(efficiency = 0.99,family =
"optimal")
fit.rob = lmrobdetMM(y~x,control = ctrl)
coef(fit.ls)
coef(fit.rob)
```

```
plotLSandRobustVHI = function(x)
{
  ret = x
  x=(ret[,2]-ret[,3])*100
  y=(ret[,1]-ret[,3])*100
  fit.ls = lm(y~x)
  fit.rob = lmrobdetMM(y~x, control=
                       lmrobdet.control(efficiency=0.99,family="optimal"))
  plot(x,y, pch=20, xlab="Market Returns %",ylab="VHI Returns (%)",
       type="n",main="")
  abline(fit.rob, col="black", lty=1, lwd=2)
  abline(fit.ls, col="red", lty=2, lwd=2)
  abline(fit.rob$coef[1]+3*1.29*fit.rob$scale,fit.rob$coef[2],lty=3,col="black")
  abline(fit.rob$coef[1]-3*1.29*fit.rob$scale,fit.rob$coef[2],lty=3,col="black")
  ids=which(fit.rob$rweights==0)
  points(x[-ids], y[-ids], pch=20)
  points(x[ids], y[ids], pch=1)
  legend("topleft",
         legend=c(expression("Robust " ~ hat(beta)==0.63~(0.23)),
                  expression("       LS " ~ hat(beta)==1.16~(0.31))),
        lty=1:2, col=c("black", "red"), bty="n", lwd=c(2,2), cex=1.2)
}
plotLSandRobustVHI(retVHI)
```

Example 2 R Code

```
plotLSandRobustDD = function(x)
{
  ret = x
  x=(ret[,2]-ret[,3])*100
  y=(ret[,1]-ret[,3])*100
  fit.ls = lm(y~x)
  fit.rob = lmrobdetMM(y~x, control=
                       lmrobdet.control(efficiency=0.99,family="optimal"))
  plot(x,y, pch=20, xlab="Market Returns (%)", ylab="DD Returns (%)", type="n")
  abline(fit.rob, col="black", lty=1, lwd=2)
  abline(fit.ls, col="red", lty=2, lwd=2)
  abline(fit.rob$coef[1]+3*1.29*fit.rob$scale,fit.rob$coef[2],lty=3,col="black")
  abline(fit.rob$coef[1]-3*1.29*fit.rob$scale,fit.rob$coef[2],lty=3,col="black")
  points(x, y, pch=20)
  legend("topleft",
         legend=c(expression("Robust " ~ hat(beta)==1.21 ~ (0.128)),
                  expression("     LS " ~ hat(beta)==1.19 ~ (0.076))),
         lty=1:2, col=c("black", "red"), bty="n", cex=1.2 )
  id = which(retDD <=-0.24)
  arrows(x[id]+1, y[id]+11, x[id]+0.1, y[id]+1, angle=15, length=0.1)
  text(x[id]+1, y[id]+12.5, labels="Oct. 20 1987", cex=0.9)
}
plotLSandRobustDD(retDD))
```