# ROBUST STATISTICS FOR QUANTITATIVE FINANCE

## Tutorial R-Finance 2018

R. Douglas Martin*
Director of Computational Finance Program
University of Washington

doug@.amath.washington.edu

Chicago, June 1

Joint work with "*Robust Statistics: Theory and Methods* (*2018*)" co-authors Ricardo Maronna, Victor Yohai, Matias Salibian-Barrera, and joint work with Ruben Zamar and Kjell Konis

# John Tukey On Good Statistical Practice*

"What are today's obligations of good statistical practice? I suggest they should include these two:

1. where one or more analyses robust of efficiency are available, one (or more) of them should be among those applied (in parallel) to the data, and
2. its (their) results should be compared with those of the other analyses used"

* Useable Resistant/Robust Techniques of Analysis (1979)

# 1. Robust Statistics: Theory & Methods

**The Book**

R. A. Maronna, R. D. Martin, V. J. Yohai and Matias Salibian-Barrera (2018), 2nd Edition, Wiley (MMYS)

**The Companion R Package RobStatTM**

Developed and maintained by Matias Salibian-Barrera

To install RobStatTM, first install devtools, then use:

devtools::install_github("msalibian/RobStatTM")

# Book Chapters

1. Introduction
2. Location and Scale
3. Measuring Robustness
4. Linear Regression 1
5. Linear Regression 2 *****
6. Multivariate Analysis *****
7. Generalized Linear Models *****
8. Time Series
9. Numerical Algorithms
10. Asymptotic Theory of M-Estimators
11. Description of Data Sets

***** Chapters with most new material

# RobStatTM Main Modeling Functions

**Robust Regression**

| | |
|---|---|
| lmrobM | M-estimator regression with fixed design |
| lmrobdetMM | MM estimator regression |
| lmrobdetDCML | DCML estimator regression |
| lmrobdet.control | tuning parameters for lmrobdetMM/DCML |
| rob.linear.test | likelihood ratio test of a linear hypothesis |
| lmrobdetMM.RFPE | robust FPE model selection criterion |
| step.lmrobdetMM | stepwise regression with RFPE |

**Multivariate**

| | |
|---|---|
| MultiRobu | multivariate location and scatter/covmat |
| GSE | covmat for independent outliers in variables |
| KurtSD | new multivariate location & scatter/covmat |
| SMPCA | PCA based on minimizing residual M-scale |

## Logistic Regression

| | |
|---|---|
| ▪ WBYlogreg | redescending weighted M-estimator |
| ▪ BYlogreg | M-estimator |
| ▪ WMLlogreg | weighted maximum-likelihood estimator |

## Time Series

| | |
|---|---|
| ▪ arima.rob | filtered tau-estimators |

## Utility Functions

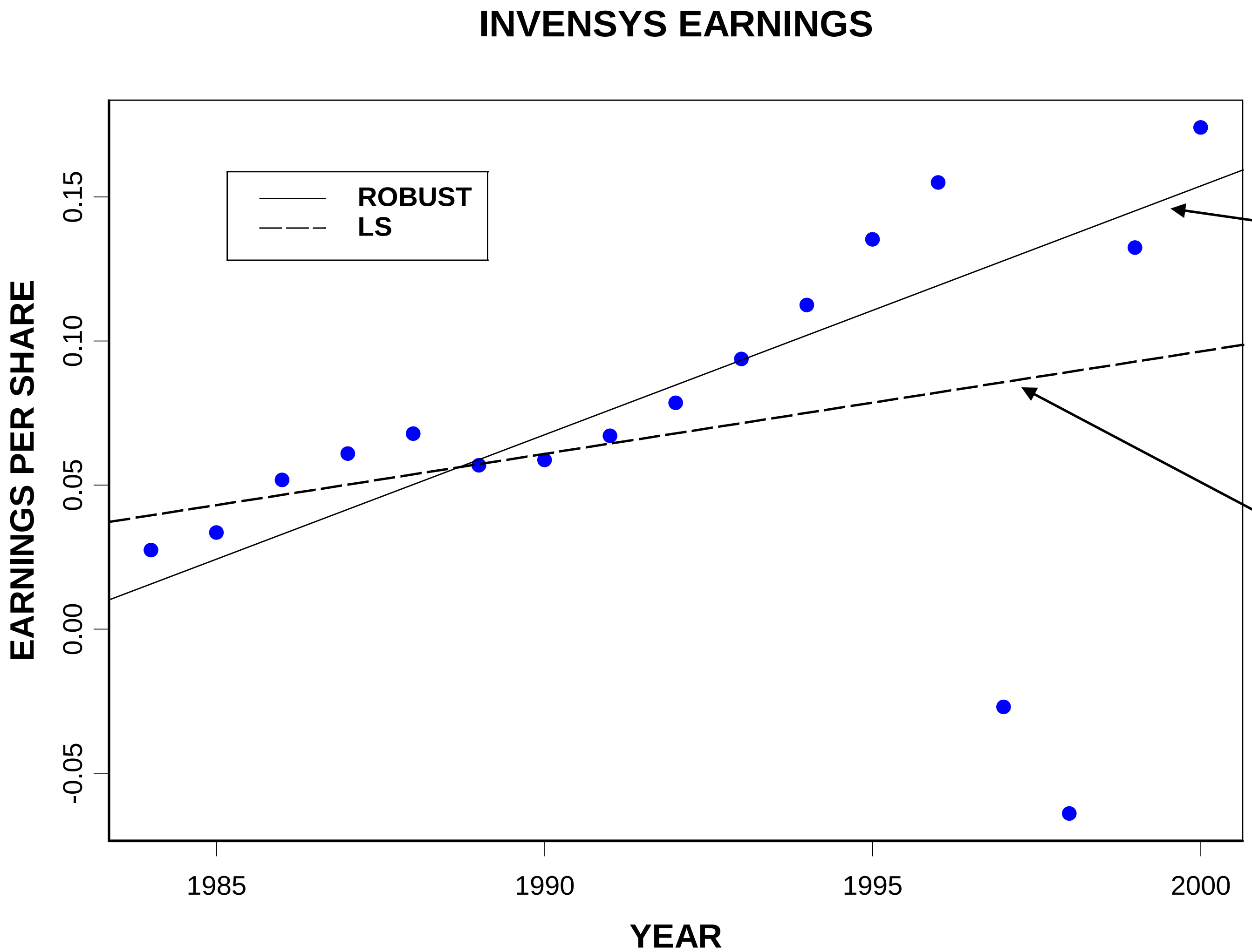
- Many . . .

# 2. Key Robustness Concepts

## Data-Oriented Robustness

- Not much influenced by outliers
- Good model fit to bulk of the data
- Reliable returns outlier detection

# Robust vs. LS Prediction of EPS




**The ROBUST line fits bulk of data well by down-weighting the 2-D outliers, and provides a good predictor of EPS**

**The CLASSIC LEAST SQUARES line is a poor fit to the bulk of the data, and is a very poor predictor of EPS**

Example from user in DuPont Corporate Finance

# Robustness Models and Methods

- Tukey-Huber outliers model

*Estimator Evaluation Methods*

- Variance and efficiency
- Bias and breakdown point
- Mean-squared error

*Leading Estimators*

- M-estimators and MM-estimators
- S-estimators

# Tukey-Huber Outlier Generating Model

Fixed arametric distribution, often univariate/multivariate normal

Unknown asymmetric, or non-elliptical distributions

$$\boxed{F(\mathbf{r}) = (1-\gamma)\cdot F_{\boldsymbol{\theta}}(\mathbf{r}) + \gamma \cdot H(\mathbf{r})}$$

Unknown, often .01 to .05 or .10 but want to do well for $0 \le \gamma < .5$

# Estimator Variance and Efficiency

Finite-sample variance $V(\hat{\theta}_n, F) \quad \rightarrow 0 \text{ as } n \rightarrow \infty$

Asymptotic variance $V(\theta, F)$

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, V(\theta, F)\right)$$

Variance efficiency:

$$Var(\hat{\theta}_{ROBUST}, F) = \frac{\text{var}(\hat{\theta}_{MLE}, F)}{\text{var}(\hat{\theta}_{ROBUST}, F)}$$

# Estimator Bias

<u>Finite Sample Bias</u>

$$\mathrm{B}\left(\hat{\theta}_n\right) = \mathrm{E}\left(\hat{\theta}_n\right) - \theta$$

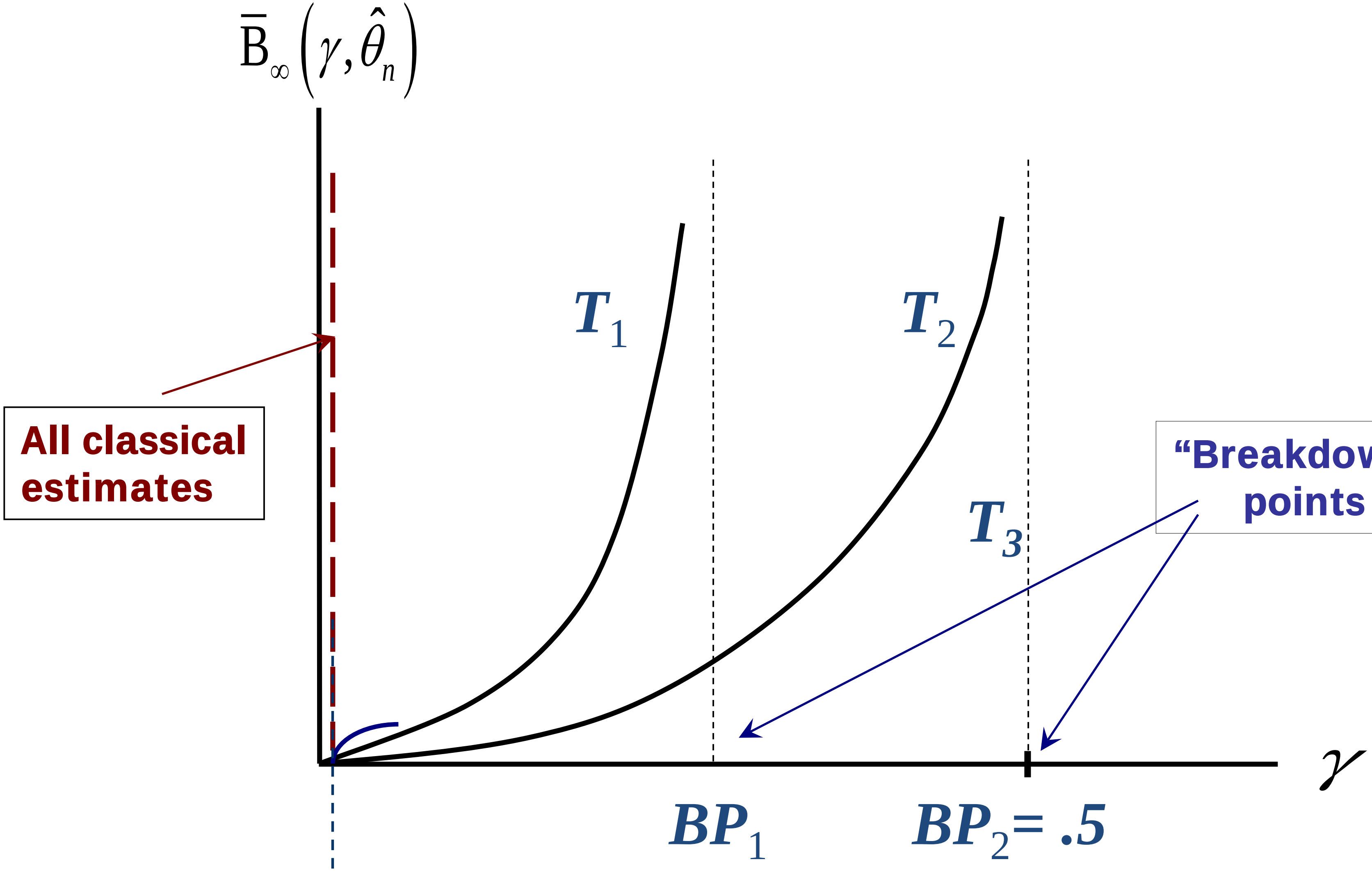
<u>Asymptotic Bias</u> $\hat{\theta}_n \to \theta_\infty$

$$\mathrm{B}_\infty\left(\hat{\theta}_n\right) = \theta_\infty - \theta$$

Both of the above depend upon $H$ and $\gamma$ under the Tukey-Huber model.

# Asymptotic Maximum Bias Curves

Want to find estimator with "smallest" curve



# Asymptotic Regression Theory Results

**Asymptotic Theory Based Estimators**

- Min-max variance location estimation (Huber, 1964)
- Min-max bias estimation of scale (Martin & Zamar, 1989, 1993)
- Min-max bias regression (Martin, Yohai & Zamar, 1989, 1993)
- Efficiency-constrained min-max bias regression

  (Yohai & Zamar, 1997, Svarc, Yohai & Zamar, 2002)
- Efficient min-max bias regression

  (Gervini & Yohai, 2002, Maronna & Yohai, 2014)

# Finite-Sample Mean-Squared Error (mse)

$$\boxed{\mathrm{mse}_F\left(\hat{\theta}_n, \gamma\right) = \mathrm{var}_F\left(\hat{\theta}_n, \gamma\right) + \mathrm{bias}_F^2\left(\hat{\theta}_n, \gamma\right)}$$

The above are used to rank robust estimators, **with extensive Monte Carlo simulations**, and then make recommendations of which estimator to use.

# 3. Location M-Estimates

Maximum-Likelihood Type Estimators (Huber, 1964)

$$r_i = \mu + \sigma \cdot \varepsilon_i, \quad i = 1, 2, \cdots, n \qquad \varepsilon_i \ i.i.d. \sim f_0$$

assumed known for the moment

$$\begin{aligned}\hat{\mu}_{MLE} &= \arg\max_{\mu} L(\mu \mid r_1, r_2, \cdots, r_n) \\ &= \arg\max_{\mu} \prod_{i=1}^{n} \frac{1}{\sigma} f_0\left(\frac{r_i - \mu}{\sigma}\right) \\ &= \arg\min_{\mu} \sum_{i=1}^{n} \rho\left(\frac{r_i - \mu}{\sigma}\right)\end{aligned}$$

where $\rho(x) = -\log(f_0)$

# MLE Estimating Equations

With $\psi(x) = \rho'(x)$ the corresponding "estimating" equation is:

$$\sum_{i=1}^{n} \psi\left(\frac{r_i - \hat{\mu}}{\sigma}\right) = 0$$

To implement, one needs to use a scale estimate $\hat{\sigma}$, as described in MMYS:

$$\sum_{i=1}^{n} \psi\left(\frac{r_i - \hat{\mu}}{\hat{\sigma}}\right) = 0$$

Use $\hat{\sigma} = \mathrm{MADM}$

(median absolute deviaton about median, normalized to be unbiased for $\sigma$ under normality)

# Special cases of MLE's

Normal Distributions (Least Squares, Sample Mean)

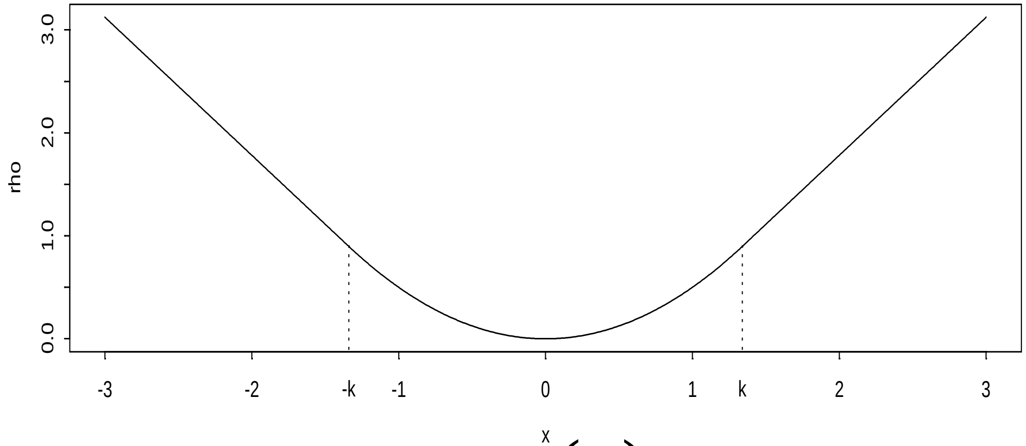
$$\rho(x) = c + \frac{x^2}{2}, \qquad \psi(x) = x$$

Exponential Distributions (L1/LAD, Sample Median)

$$\rho(x) = c + |x|, \qquad \psi(x) = \mathrm{SGN}(x)$$

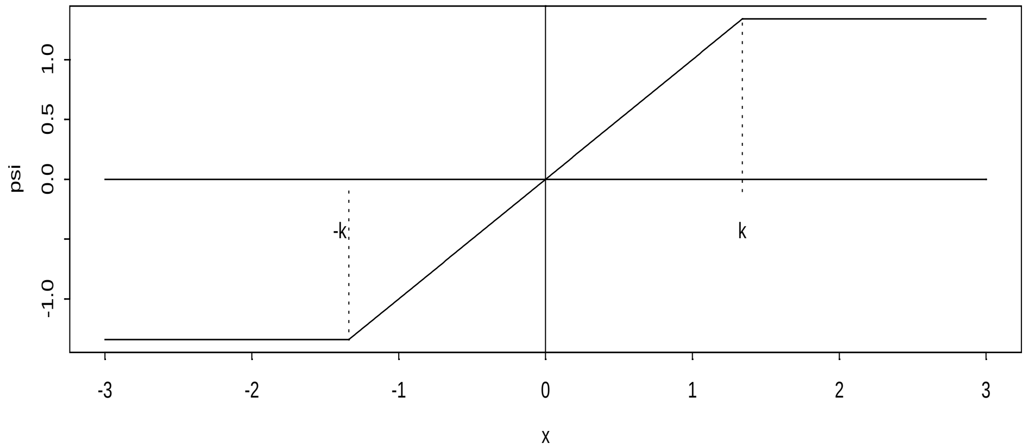
# Huber Min-Max Variance MLE Solution (1964)

$$\rho(x)$$




Unbounded, like LS in the middle and like LAD outside. Is an MLE for a distribution that is normal in the middle and exponential in the tails

$$\psi(x)$$

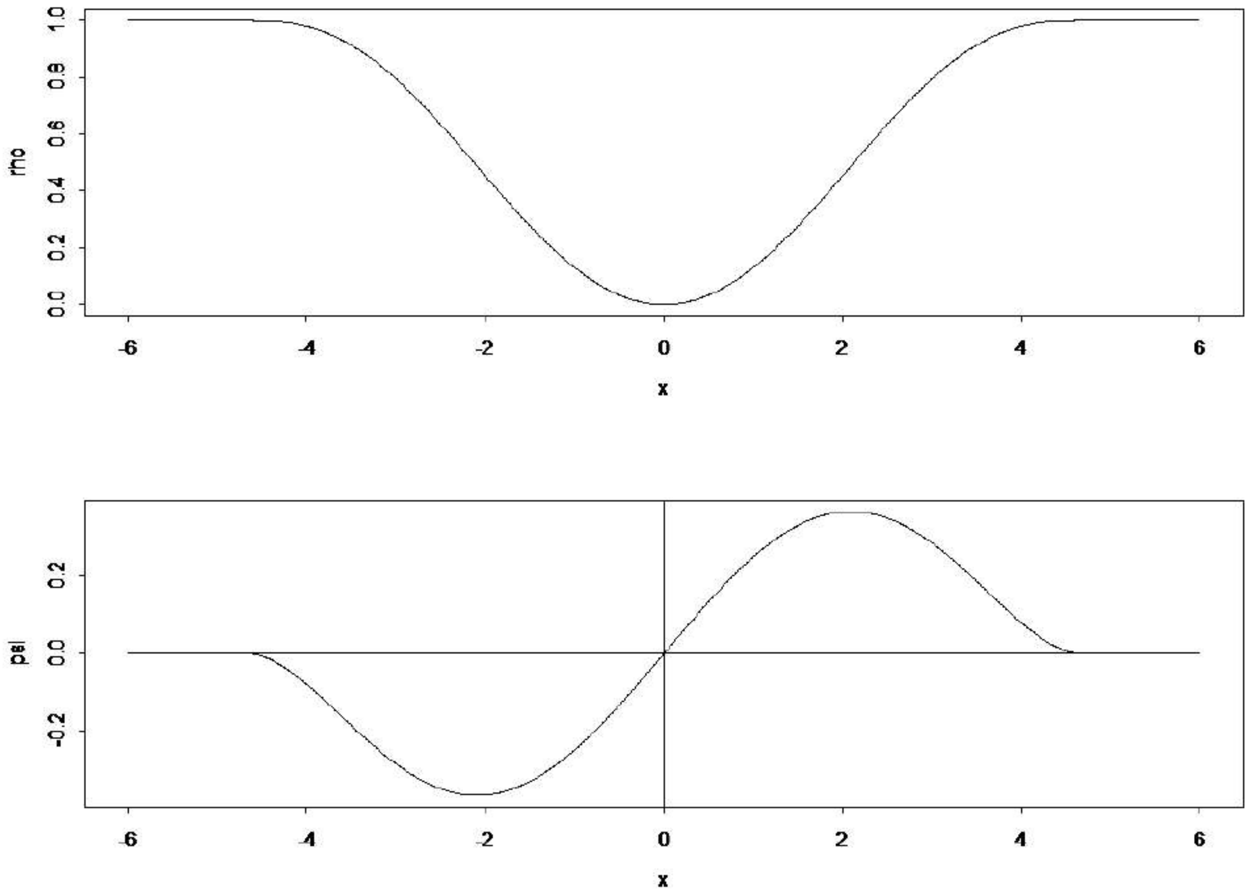



Better robust M-estimators use bounded loss functions and re-descend to zero.

# Tukey Bisquare Rho and Psi Functions

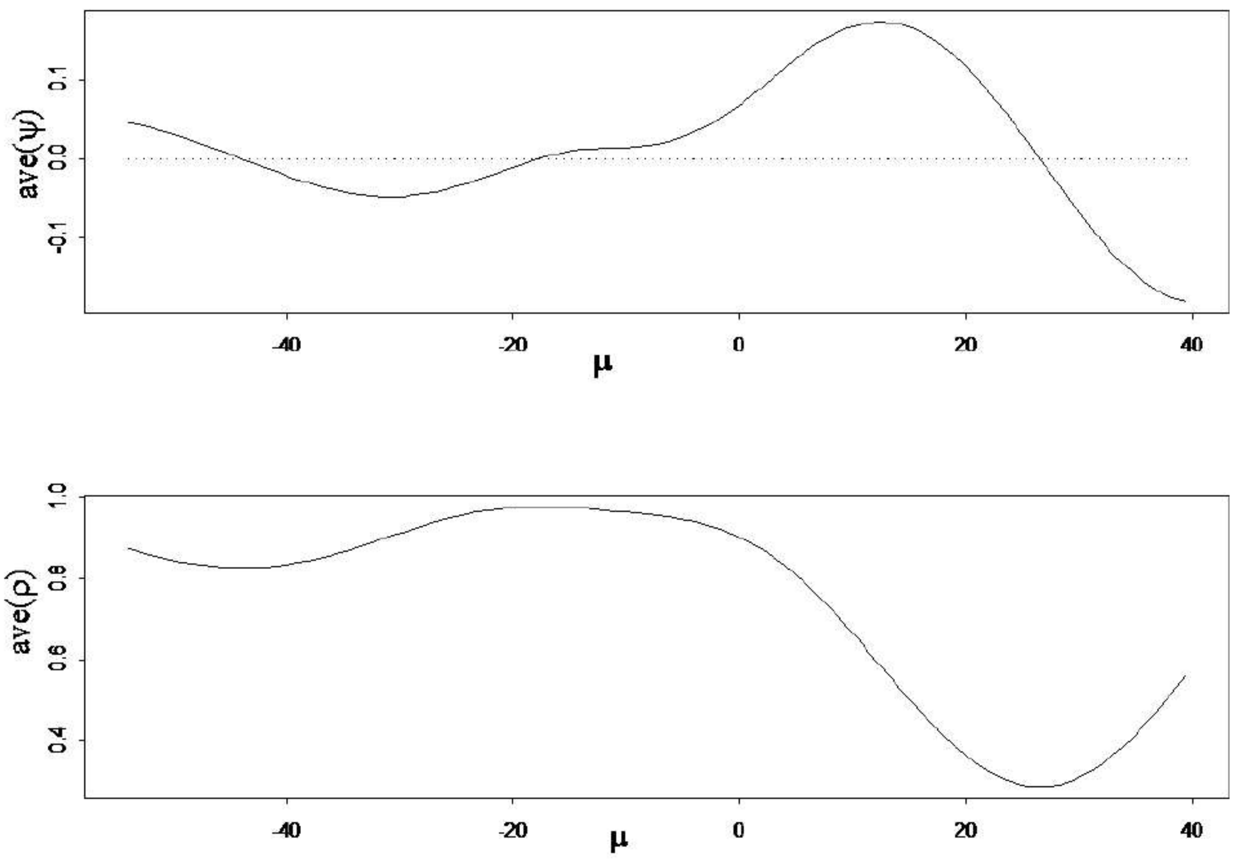
**Not an MLE** (no pdf yields this rho and psi)

See MMYS for formulas



# M-Estimator Solves a Nonlinear Equation

- Essentially unique root for convex rho and monotone psi
- Multiple roots with redescending psi, e.g.,Tukey bisquare




MMYS Section 2.8, Figure 2.6

# Need a Good Numerical Algorithm

- Weighted least squares (WLS) version of M-estimator

$$\sum_{i=1}^{n} \psi\left(\frac{r_i - \hat{\mu}}{\text{MADM}}\right) \;\rightarrow\; \sum_{i=1}^{n} W\left(\frac{r_i - \hat{\mu}}{\text{MADM}}\right)\left(r_i - \hat{\mu}\right) = 0$$

Data-dependent weight defined by: $W\left(t\right) = \dfrac{\psi(t)}{t}$

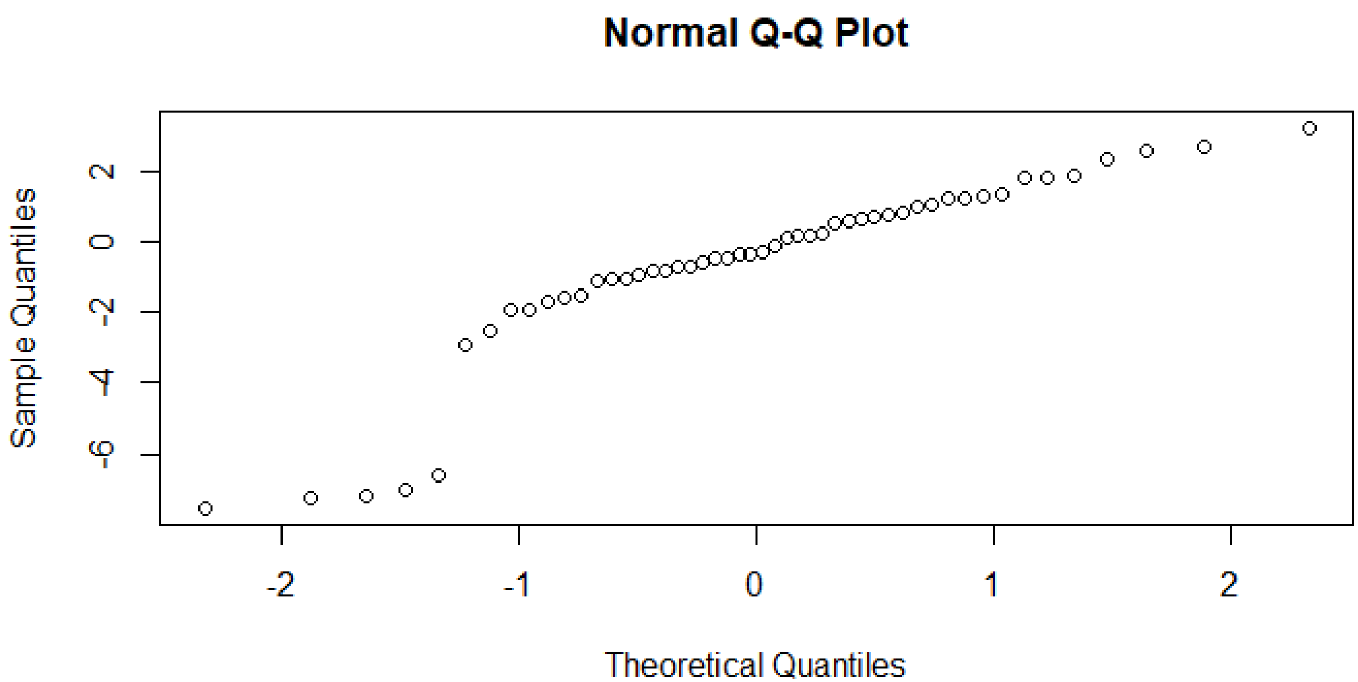
- RobStatTM uses a very reliable iterated weighted least squares (iterated weighted average) algorithm with

$$\hat{\mu}_0 = median(r_1, r_2, \cdots, r_n)$$

For convergence proof see Section 9.1 of MMYS.

# MLocDis.R() Example with Default Bisquare

```
library(RobStatTM)
set.seed(123)
r1 = rnorm(45, sd = 1.5)
r2 = rnorm(5, mean = -7, sd = .5)
r <- c(r1,r2); qqnorm(r)
```



```
> mean(r)
[1] -0.6224736
> sd(r)/sqrt(50)
[1] 0.3641196
> sd(r)
[1] 2.574714

> # robust loc, stderror, scale
> MLocDis(r)
$mu
[1] 0.08544817
$std.mu
[1] 0.2220985
$disper
[1] 1.675064
```

# Why the Location M-estimator Details?

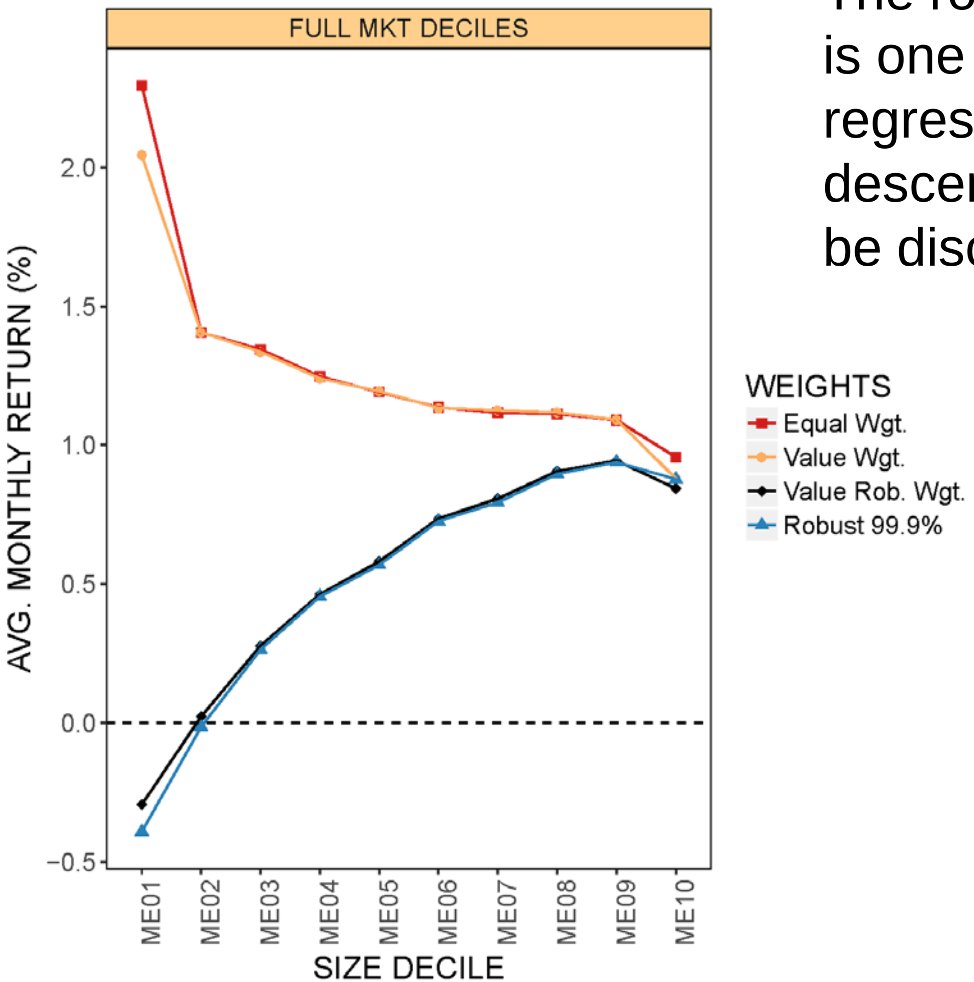
After all, there are very simple $\alpha$ trimmed mean estimators based on ordered returns $r_{(1)} \le r_{(2)} \le \cdots \le r_{(n)}$, for example for $\alpha = .05$ and $n = 60$:

$$\hat{\mu}_{trim,\,.05} \quad \frac{1}{54}\sum_{i=4}^{57} r_{(i)}$$

**A1: The most important reason is that many of the best robust estimators for regression and multivariate analysis are based on variants of M-estimator methods**.

**A2: For univariate location estimation M-estimators are better than trimmed (or Winsorized) means, as the next important example shows.**

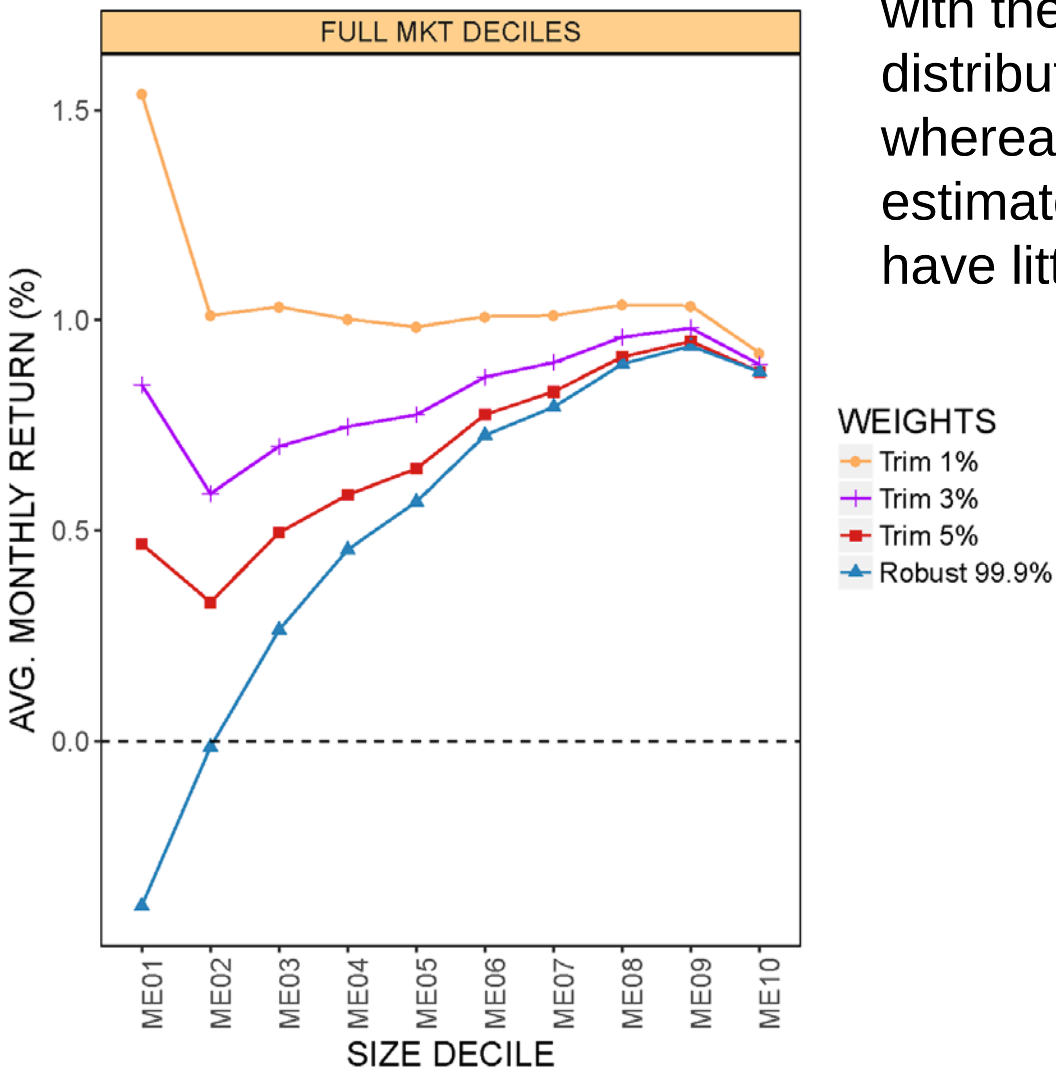
# The Common Wisdom that Returns Decrease with Firm Size is Due to Outliers in "Small" Firms




from time series of decile portfolios

The robust location estimator used is one used for optimal bias robust regression, that has a rapidly descending psi function, and will be discussed next.

# Trimming will Not Suffice




The problem with (symmetric) trimming is that it cannot cope with the positively skewed distributions of small-size firms, whereas robust location estimators with redescending psi have little problem with that.

# 4. Bias Robust High Efficiency Regression

- Two main classes of estimators:
  - MM estimators **lmrobdetMM()**
  - DCML estimators **lmrobdetDCML**()

- Try to minimize maximum asymptotic bias under Tukey-Huber model but maintain very high normal distribution efficiency

- Evaluate finite-sample MSE via simulation for wide range of sample sizes and numbers of factors

# Regression MM-Estimator Optimization

$$y_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + \varepsilon_i, \qquad i = 1, 2, \cdots, n$$

General Regression Optimization Problem

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

The Estimating Equation

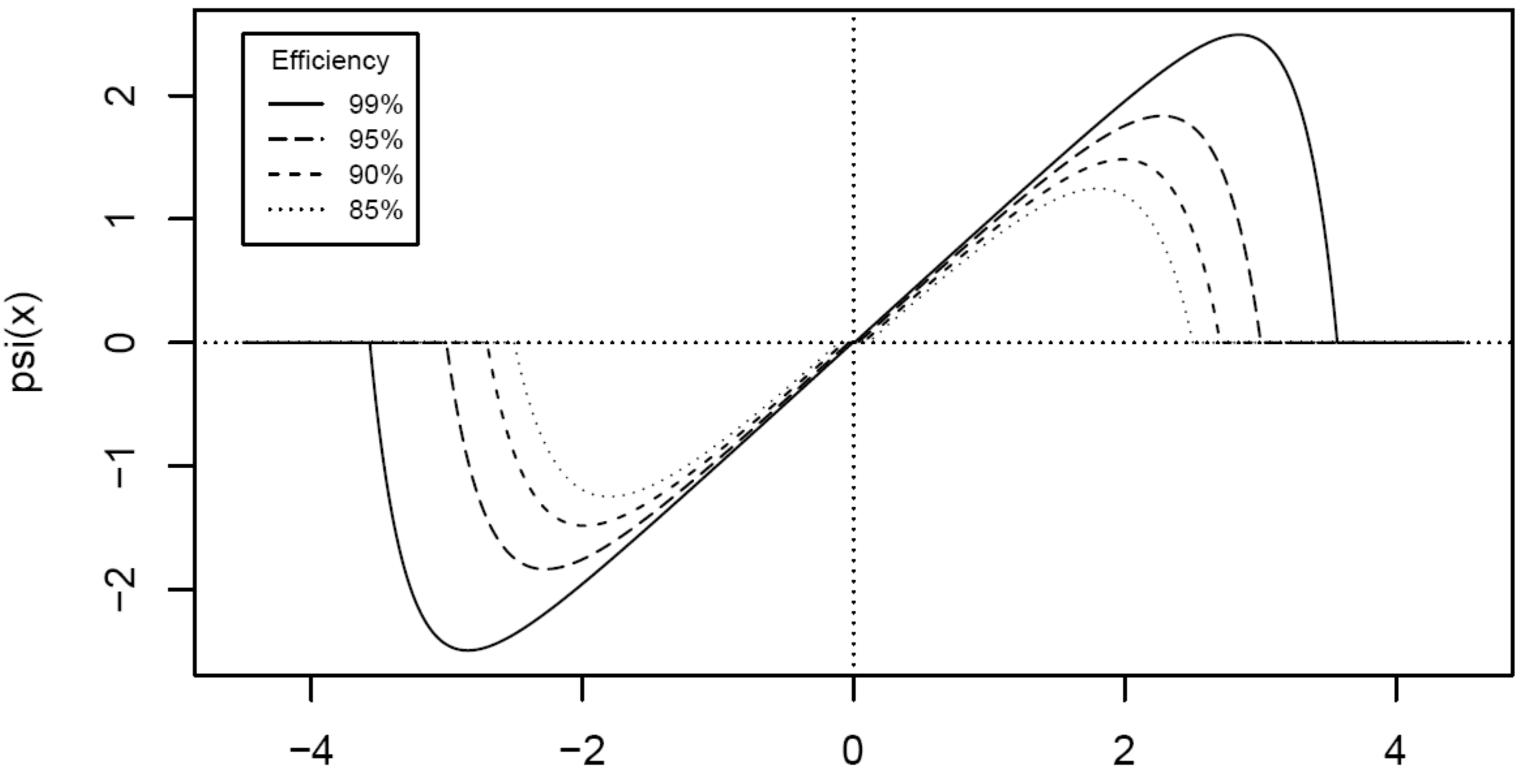
$$\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathbf{x}_i \cdot \psi\left(\frac{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}}{\hat{s}_o}\right) = \mathbf{0}$$

# Efficient Min-Max Bias Robust Regression

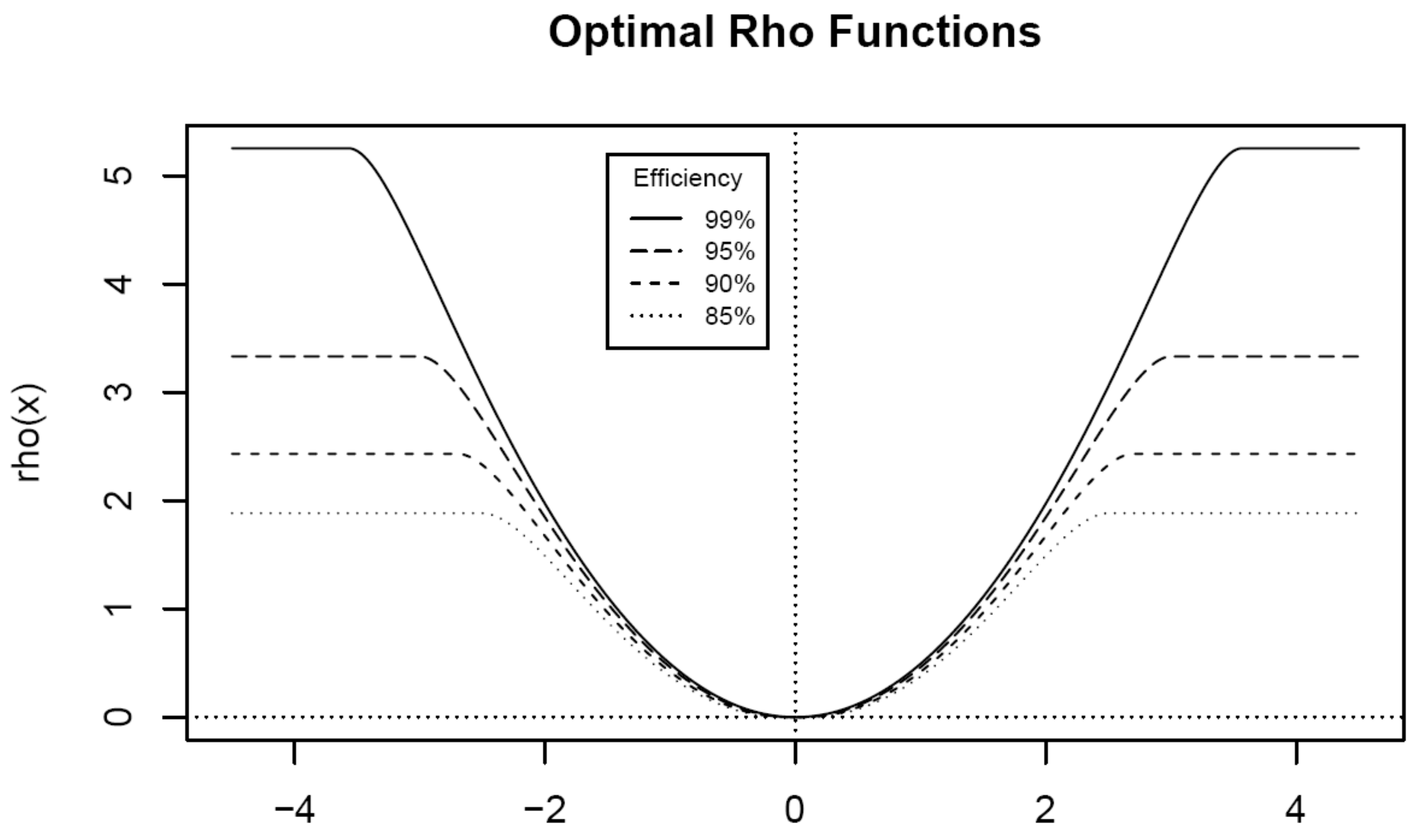
Yohai and Zamar (1997) and Svarc, Yohai & Zamar (2002)

Solution was found in terms of the $\psi$ function (not the rho!)

$$\psi_c(u) = \mathrm{sgn}(u)\left(-\frac{\varphi'(|u|)+c}{\varphi(|u|)}\right)^+$$



The analytic form of the optimal rho function, shown below was not known until this year, and is given in the working paper by Konis, Martin, Yohai and Zamar (2018)




The **robust** and **robustbase** packages usd a polynomial approximation for $\rho$ and integrated it to get $\psi$.

# Robust Regression with lmrobdetMM()

- "MM" – use a highly robust initial estimator (consisting of a (fast?) first-step initial estimator, followed by second-step S-estimator with lower efficiency but small maximum bias), and then use a fast IRWLS to find the nearest local minimum to the M-estimator equation.

- "det": Special fast deterministic first-step estimator due to Pena and Yohai (1999). Much better than previous resampling method. See MMYS Section 5.7.4

- **Very important**: A 99% normal distribution efficiency optimal optimal bias robust regression has very small maximum bias over Tukey-Huber model distributions.

# M-Estimator for Scale and S-Estimators

- M-scale estimator

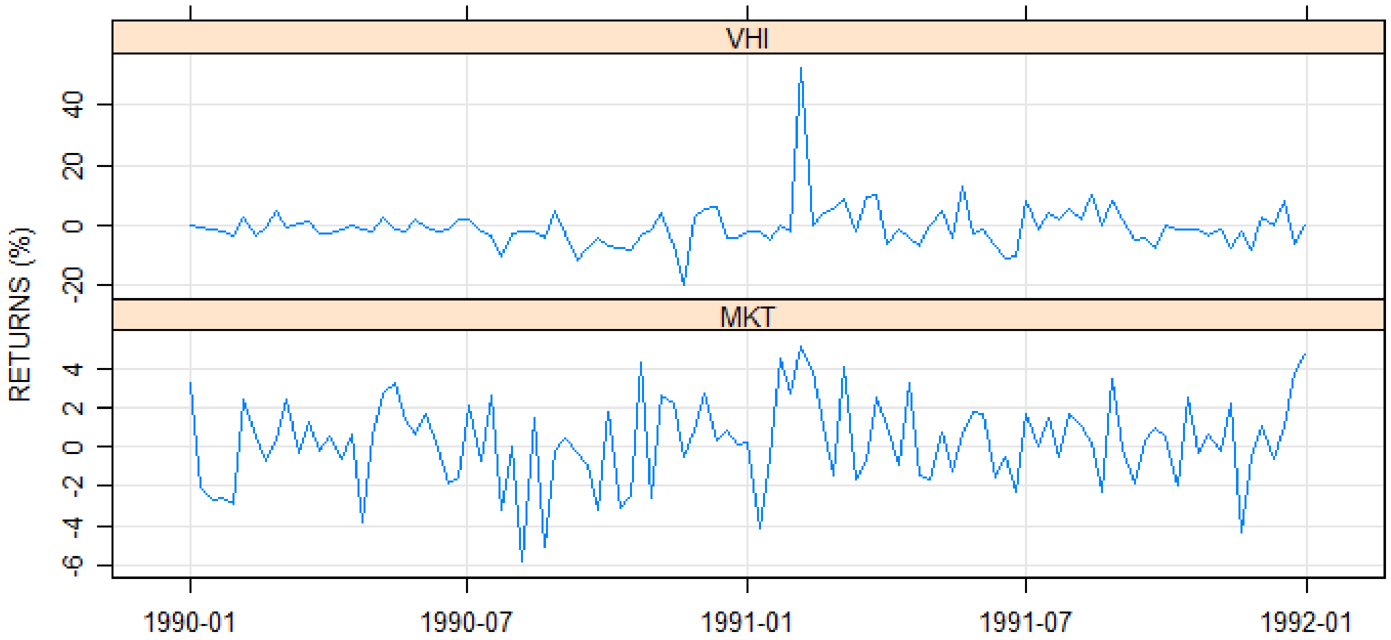
$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i}{\sigma}\right)=b$$

- Regression S-estimators:

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i-\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)=.5 \quad \rightarrow \quad \hat{\sigma}(\boldsymbol{\beta})$$

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\hat{\sigma}(\boldsymbol{\beta})$$

# Simple CAPM Example




```
> library(mpo)
> (names(retVHI))
[1] "VHI" "MKT" "RF"
> ret12 = retVHI[,1:2]
> tsPlot(ret12,cex = .8)
```

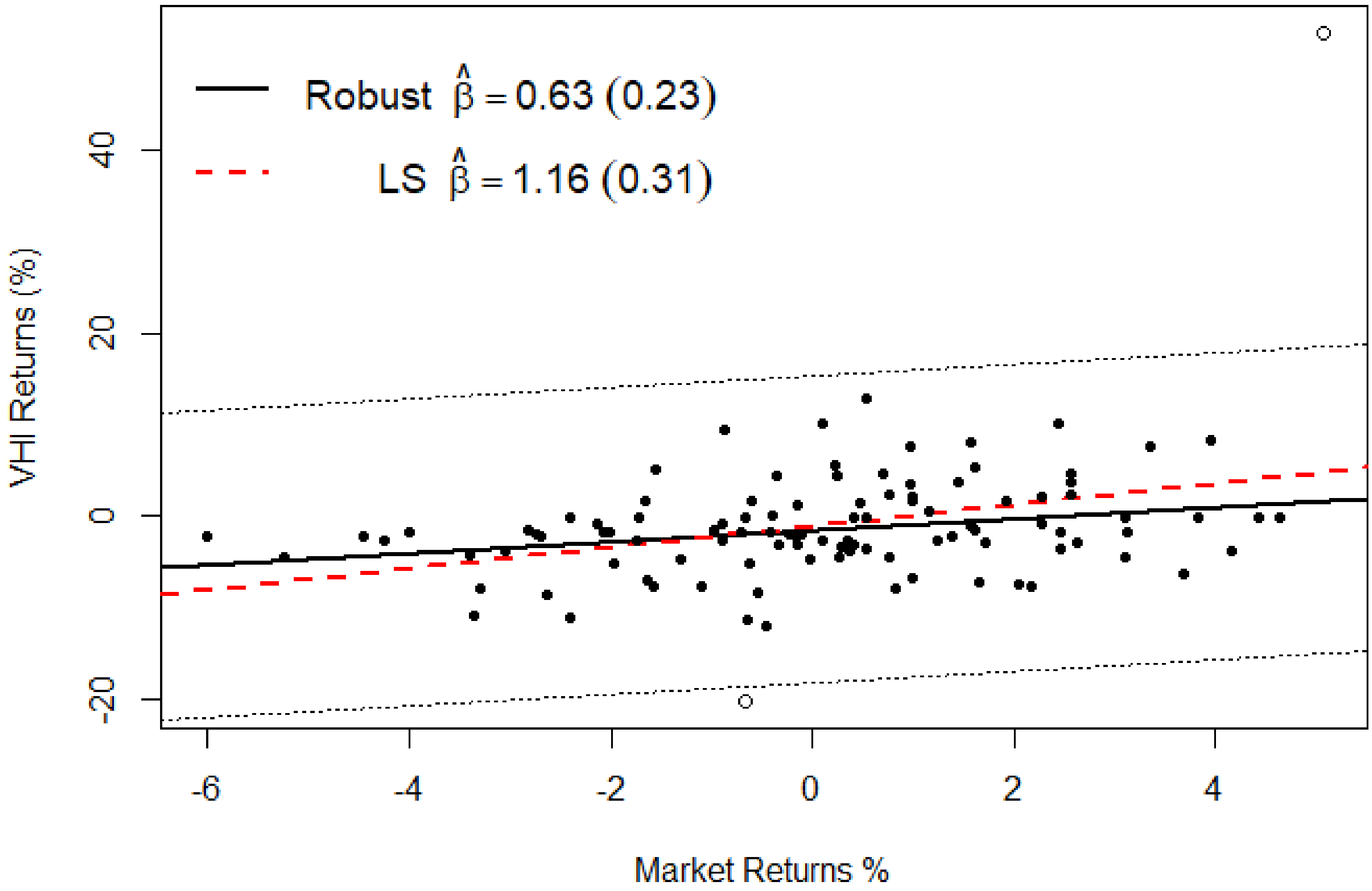
# LS Fit versus Robust Fit with lmrobdetMM()

```
> library(RobStatTM)
> x=(retVHI[,2]-retVHI [,3])*100
> y=(retVHI[,1]-retVHI [,3])*100
> fit.ls = lm(y~x)
```

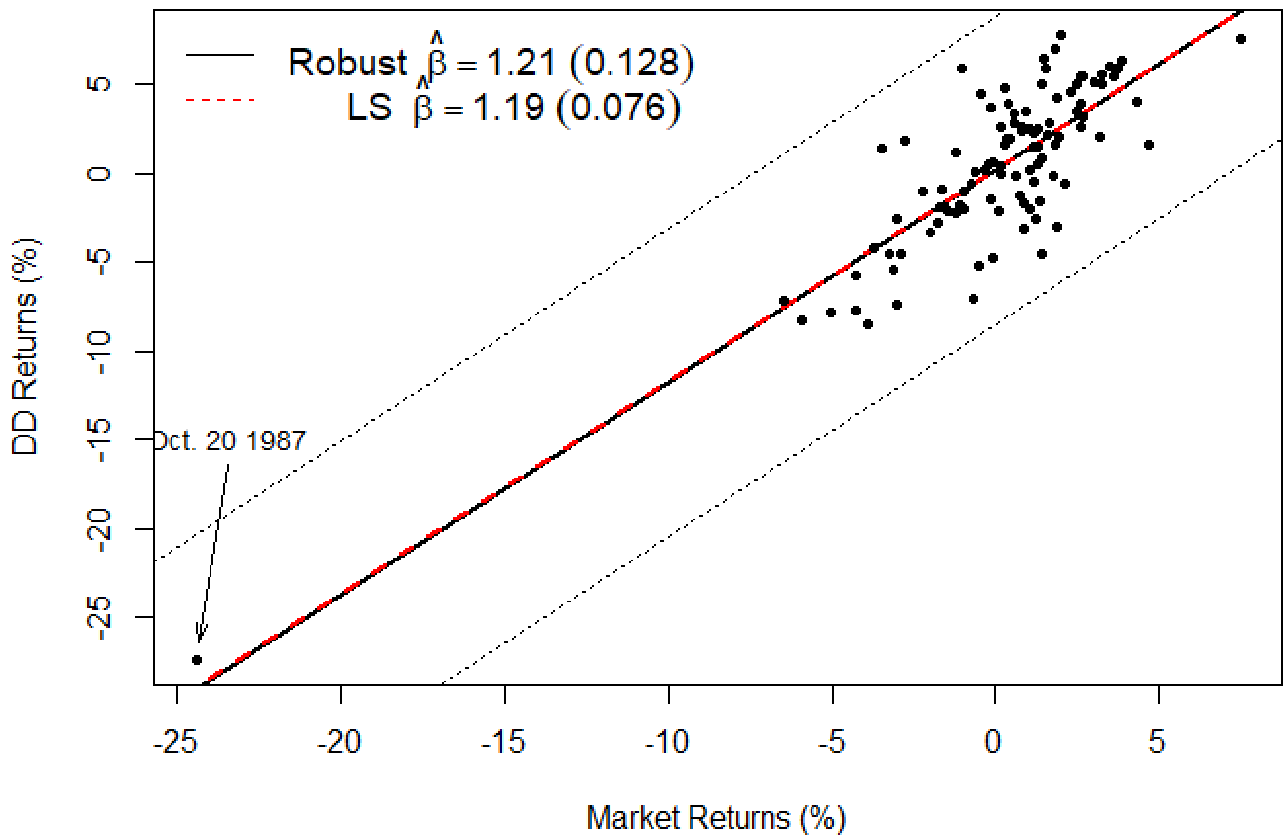
```
> ctrl=lmrobdet.control(efficiency = 0.99,
                              family = "optimal“)
> fit.rob = lmrobdetMM(y~x,control = ctrl)
```

```
> coef(fit.ls)
(Intercept)           x
  -1.098286    1.157390

> coef(fit.rob)
(Intercept)           x
 -1.4548205   0.6282898
```


Robust $\hat{\beta} = 0.63\ (0.23)$
LS $\hat{\beta} = 1.16\ (0.31)$
VHI Returns (%)
40
20
0
-20
-6
-4
-2
0
2
4
Market Returns %

# An Important Example




**N.B.** mrobdetMM() does not reject “good” outliers that are consistent with the fit to the bulk of the returns

# Robust Regression with lmrobdetDCML()

- Distance constrained maximum-likelihood (DCML)

$\hat{\boldsymbol{\beta}}_0$: a highly robust estimator, typically an MM - estimator

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} L(\boldsymbol{\beta} \mid r_1, \mathrm{r}_2, \cdots, \mathrm{r}_n) \quad \text{subject to} \quad d(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}) \le \delta$$

- For linear regression with normal errors, DCML uses the Kullback-Leibler (KL) distance estimate:

$$\Delta_{\hat{\mathrm{V}}_{\mathrm{x}}} = \hat{d}_{KL,\mathrm{V}_{\mathrm{X}}}(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}) = \frac{1}{\hat{\sigma}_0^2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_0\right)' \hat{\mathbf{V}}_{\mathrm{X}} \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_0\right)$$

robust estimate

- Thus the DCML optimization is:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) \quad \text{subject to} \quad d_{KL,\hat{\mathrm{V}}_x}(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}) \le \delta$$

Maronna and Yohai (2014) show that this leads to:

$$\hat{\boldsymbol{\beta}} = t \cdot \hat{\boldsymbol{\beta}}_{LS} + (1-t) \cdot \hat{\boldsymbol{\beta}}_0$$

where

$$t = \min\left(1, \frac{\delta}{\Delta_{\hat{\mathrm{V}}_\mathrm{x}}}\right) \qquad \delta = \delta_{p,n} = 0.3\frac{p}{n}$$

The DCML estimator was shown by Maronna & Yohai (2014) to be asymptotically fully efficient, and best finite-sample efficiency and MSE among competing estimators!

# VHI data lmrobdetDCML() fit vs. lmrobdetMM()

Modifying the code on slide 33 by making the obvious replacement of some lines with the following gives the results shown:

```
> x=(retVHI[,2]-retVHI [,3])*100
> y=(retVHI[,1]-retVHI [,3])*100
> ctrl=lmrobdet.control(efficiency=0.99,family="optimal")
> fit.DCML = lmrobdetDCML(y~x,control = ctrl)
> fit.MM = lmrobdetMM(y~x,control = ctrl)

> round(summary(fit.DCML)[9]$coefficients[2,1:2],3)
Estimate Std. Error
     0.769      0.211

> round(summary(fit.MM)[9]$coefficients[2,1:2],3)
Estimate Std. Error
     0.628      0.229
```

- Finite-sample MSE efficiency comparisons

| $p$ | $n$ | MM-Bi85 | G-Y | MM-Op99 | DCML-Bi85 | DCML-Op99 |
|---|---|---|---|---|---|---|
| 10 | 50 | 0.80 | 0.85 | 0.95 | 0.99 | 1.00 |
| | 100 | 0.82 | 0.89 | 0.97 | 0.99 | 1.00 |
| | 200 | 0.84 | 0.93 | 0.99 | 1.00 | 1.00 |
| 20 | 100 | 0.79 | 0.86 | 0.96 | 0.99 | 1.00 |
| | 200 | 0.82 | 0.92 | 0.97 | 0.99 | 1.00 |
| | 400 | 0.84 | 0.95 | 0.99 | 1.00 | 1.00 |
| 50 | 250 | 0.78 | 0.88 | 0.95 | 1.00 | 1.00 |
| | 500 | 0.82 | 0.94 | 0.98 | 1.00 | 1.00 |
| | 1000 | 0.84 | 0.97 | 0.99 | 1.00 | 1.00 |

- Finite-sample maximum MSE comparisons

| $p$ | $n$ | MM-Bi85 | G-Y | MM-Op99 | DCML-Bi85 | DCML-Op99 |
|---|---|---|---|---|---|---|
| 10 | 50 | 1.10 | 1.03 | 0.89 | 0.92 | 0.86 |
| | 100 | 0.55 | 0.48 | 0.46 | 0.45 | 0.44 |
| | 200 | 0.44 | 0.37 | 0.37 | 0.36 | 0.36 |
| 20 | 100 | 1.28 | 1.16 | 0.98 | 1.02 | 0.94 |
| | 200 | 0.58 | 0.49 | 0.47 | 0.47 | 0.45 |
| | 400 | 0.46 | 0.38 | 0.39 | 0.39 | 0.38 |
| 50 | 250 | 1.38 | 1.18 | 1.03 | 1.04 | 0.96 |
| | 500 | 0.62 | 0.51 | 0.49 | 0.49 | 0.47 |
| | 1000 | 0.52 | 0.42 | 0.41 | 0.43 | 0.40 |

# Improving Empirical Asset Pricing Research*

Green and Martin (2017). "Fama-French Redux with Robust Regression", SSRN Abstract ID 2963855

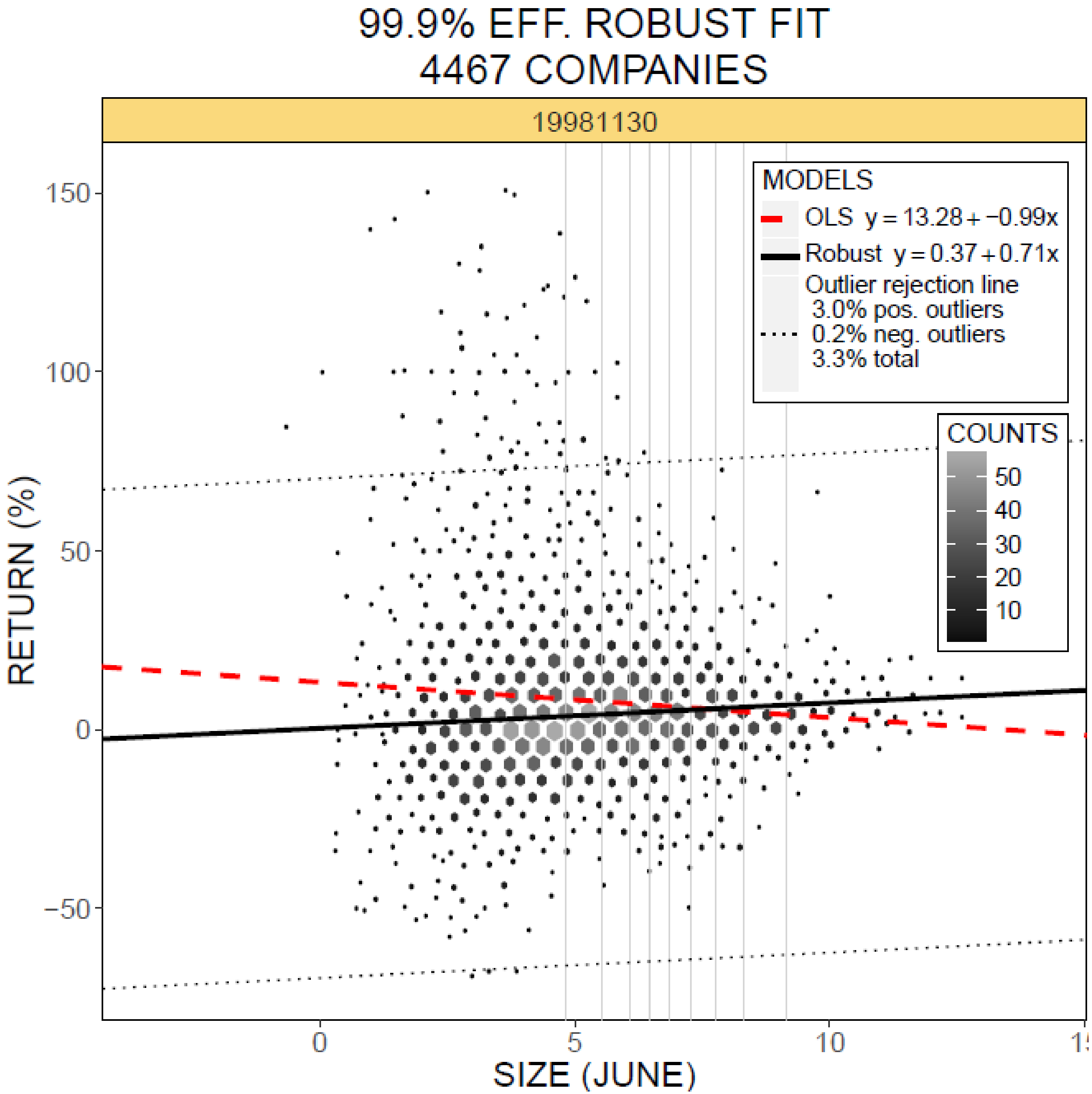
- Bias optimal robust regression with 99.9% normal distribution efficiency (using polynomial approximation for rho and psi)

- Reversed several conclusions of Fama-French 1992 (FF92), including the following:

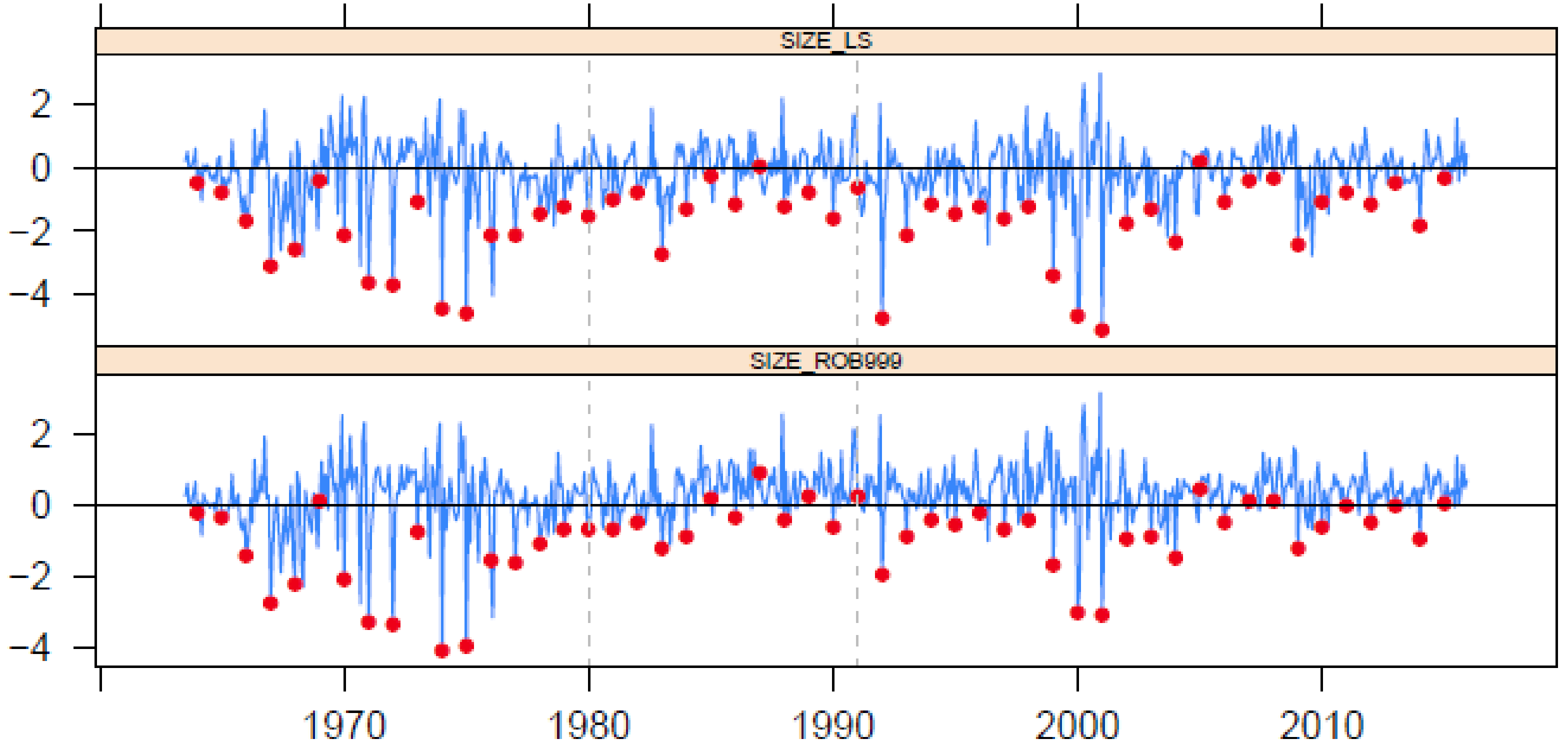
  1. Equity returns are positively related to firm size

  2. Beta relationship is significant and positive

***Robust regression is also useful for fundamental factor models**


99.9% EFF. ROBUST FIT
4467 COMPANIES
19981130
MODELS
OLS y = 13.28 + −0.99x
Robust y = 0.37 + 0.71x
Outlier rejection line
3.0% pos. outliers
0.2% neg. outliers
3.3% total
COUNTS
50
40
30
20
10
150
100
50
0
−50
RETURN (%)
0
5
10
SIZE (JUNE)

# Monthly Slopes of Returns Regressed on Size




red dots = Januaries, a well-known January effect

N.B. Presence of outliers and serial correlation, thus use a robust location estimator with HAC: Croux et al. (2003), see slide 18.

# Returns vs Size

| Factor | Method | Size 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) |

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

LTS = least trimmed squares

# Stepwise Variable Selection with RFPE

- Backward stepwise method
- Fit full model and get robust scale estimate and weights
- Use weights from full model to fit weighted least squares for each sub-model
- Use resulting sub-model beta's as initial estimate, along with robust scale from full model, to get M-estimate for sub-model
- Compute RFPE at each step, and eliminate a variable only if RFPE goes down
- The RobStatTM function is **`lmrobdetMM.RFPE()`**.

# 5. Robust Covariance Matrix Estimators

- Two main classes of estimators
  - <u>MM estimators</u> (recommended for $p < 10$)
  - <u>Rocke S estimators</u> (recommended for $p \geq 10$)

Computed with **`MultiRobu(x,type = “auto”)`**

Also: **`type = “MM”`** or **`type = “Rocke”`**

**N.B.** The well-known popular minimum covariance determinant (MCD) estimator is not as good in terms of finite-sample MSE efficiency

# M-Estimators

$$d_i^2 = \left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}\right)' \hat{\boldsymbol{\Sigma}}^{-1} \left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}\right)$$

$$\sum_{i=1}^{n} W_1(d_i)(\mathbf{x}_i - \widehat{\mu}) = \mathbf{0}$$

$$\frac{1}{n}\sum_{i=1}^{n} W_2(d_i)(\mathbf{x}_i - \widehat{\mu})(\mathbf{x}_i - \widehat{\mu})' = \widehat{\boldsymbol{\Sigma}}$$

**MM-Estimators:** M-estimator with special initial estimator

# S-Estimators

Based on use of a robust scale estimator.

$$\sum_{i=1}^{n} W\left(\frac{d_i}{\widehat{\sigma}}\right)(\mathbf{x}_i - \widehat{\mu}) = \mathbf{0}, \tag{6.31}$$

$$\frac{1}{n}\sum_{i=1}^{n} W\left(\frac{d_i}{\widehat{\sigma}}\right)(\mathbf{x}_i - \widehat{\mu})(\mathbf{x}_i - \widehat{\mu})' = c\widehat{\mathbf{\Sigma}}, \tag{6.32}$$

where

$$W = \rho' \text{ and } \widehat{\sigma} = \widehat{\sigma}(d_1, ..., d_n), \tag{6.33}$$

and $c$ is a scalar such that $\left|\widehat{\mathbf{\Sigma}}\right| = 1$. Note however that if $\rho$ is bounded (as is the usual case), $dW(d)$ cannot be monotone (Problem 6.5); actually for the estimators usually employed $W(d)$ vanishes for large $d$. Hence the estimator is not a monotone M-estimator, and therefore the estimating equations yield only *local* minima of $\widehat{\sigma}$.

# Robust Portfolio Correlations Estimators

- Outliers distort pairwise returns correlations more frequently than one might think

- Need a routine way to compute and compare classical and robust correlation estimates – use **`MultiRobu()`** for this purpose

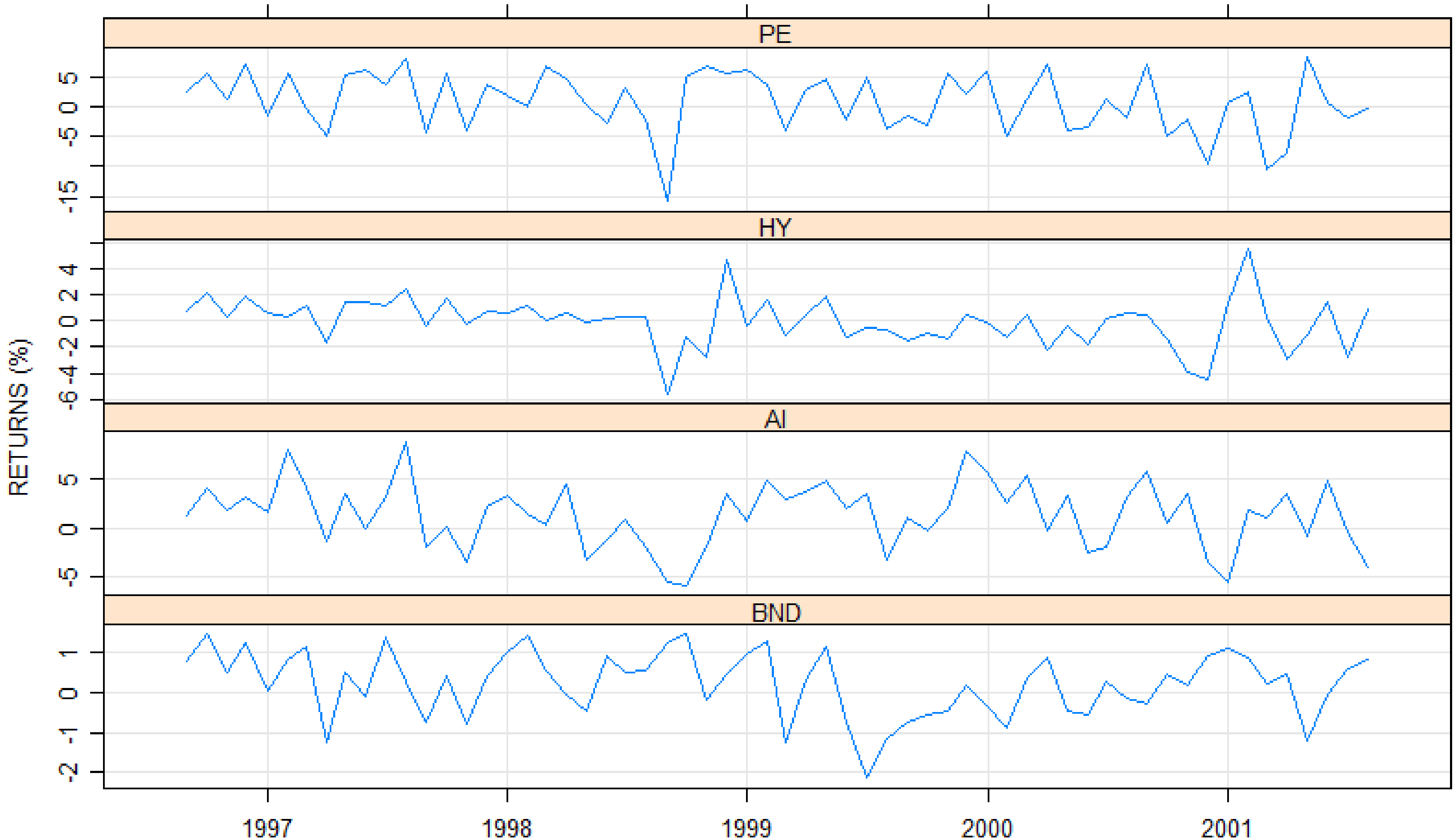
- Get robust correlations from robust covariance matrix estimate in the usual way for classical covariances*

* Can also get pairwise robust covariances and correlations directly (discuss later)

# Hedge Funds Robust Correlations

Also demonstrates use of **`fit.models()`**

```
library(xts)
library(robust)
library(mpo)
library(RobStatTM)
library(fit.models)
fmclass.add.class("covfm", "MultiRobu")
hfunds4 <- read.csv("hfunds4.ts.csv",header = TRUE)
hfunds4 <- xts(hfunds4[,-1],
                  order.by=as.Date(hfunds4[,1]))
# hfunds4 has 5 funds,and one is deleted below
# Select time interval and four of the five funds
hfunds4 <- hfunds4[-(1:81),2:5]
names(hfunds4) <- c("PE","HY","AI","BND")
tsPlot(hfunds4, cex = .8)
```

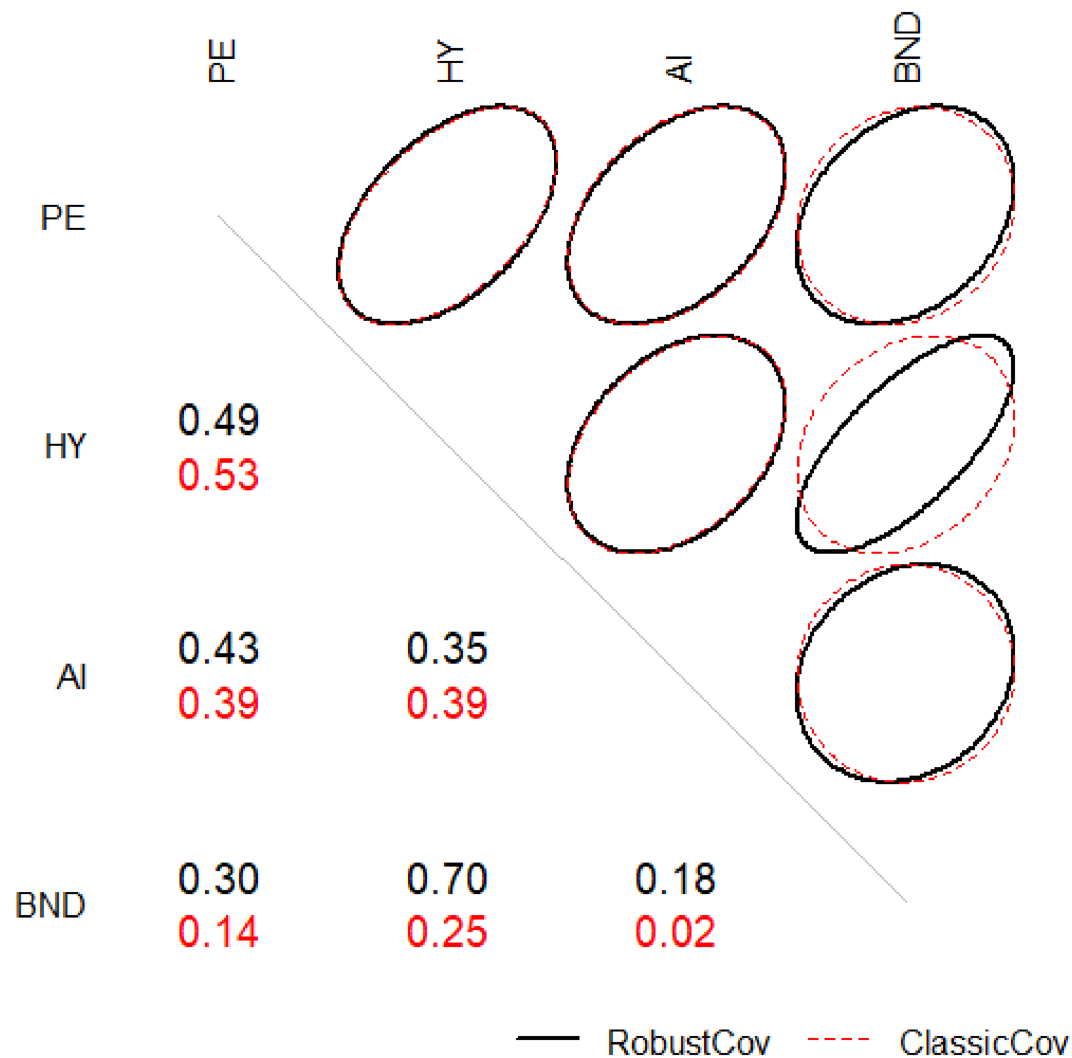

PE
HY
AI
BND
RETURNS (%)
5
0
-5
-15
4
2
0
-2
-4
-6
1
-1
-2
1997
1998
1999
2000
2001

# Hedge Funds Robust Correlations

```
data = data.frame(coredata(hfunds4))
dataMat = as.matrix(data)
cls <- covClassic(data)
rob <- MultiRobu(dataMat)
rob

# Make rob look like a covRob object
rob$cov <- rob$V
rob$center <- rob$mu
rob$corr <- FALSE
class(rob) <- "MultiRobu"

cov.fm <- fit.models(RobustCov = rob,
                        ClassicCov = cls)
plot(cov.fm, 4)
```


PE
HY
AI
BND
PE
HY
AI
BND
0.49
0.53
0.43
0.39
0.35
0.39
0.30
0.14
0.70
0.25
0.18
0.02
RobustCov
ClassicCov

# Reliable Detection of Portfolio Outliers

## Based on Robust Distances

So-called Mahalanobis distance

$$\hat{d}_t^2 = \left(\mathbf{r}_t - \hat{\boldsymbol{\mu}}\right)' \hat{\boldsymbol{\Sigma}}^{-1} \left(\mathbf{r}_t - \hat{\boldsymbol{\mu}}\right)$$

$$= \left(\mathbf{r}_t - \hat{\boldsymbol{\mu}}\right)' \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{\boldsymbol{\Sigma}}^{-1/2} \left(\mathbf{r}_t - \hat{\boldsymbol{\mu}}\right)$$

$$= \hat{\boldsymbol{z}}_t' \hat{\boldsymbol{z}}_t$$

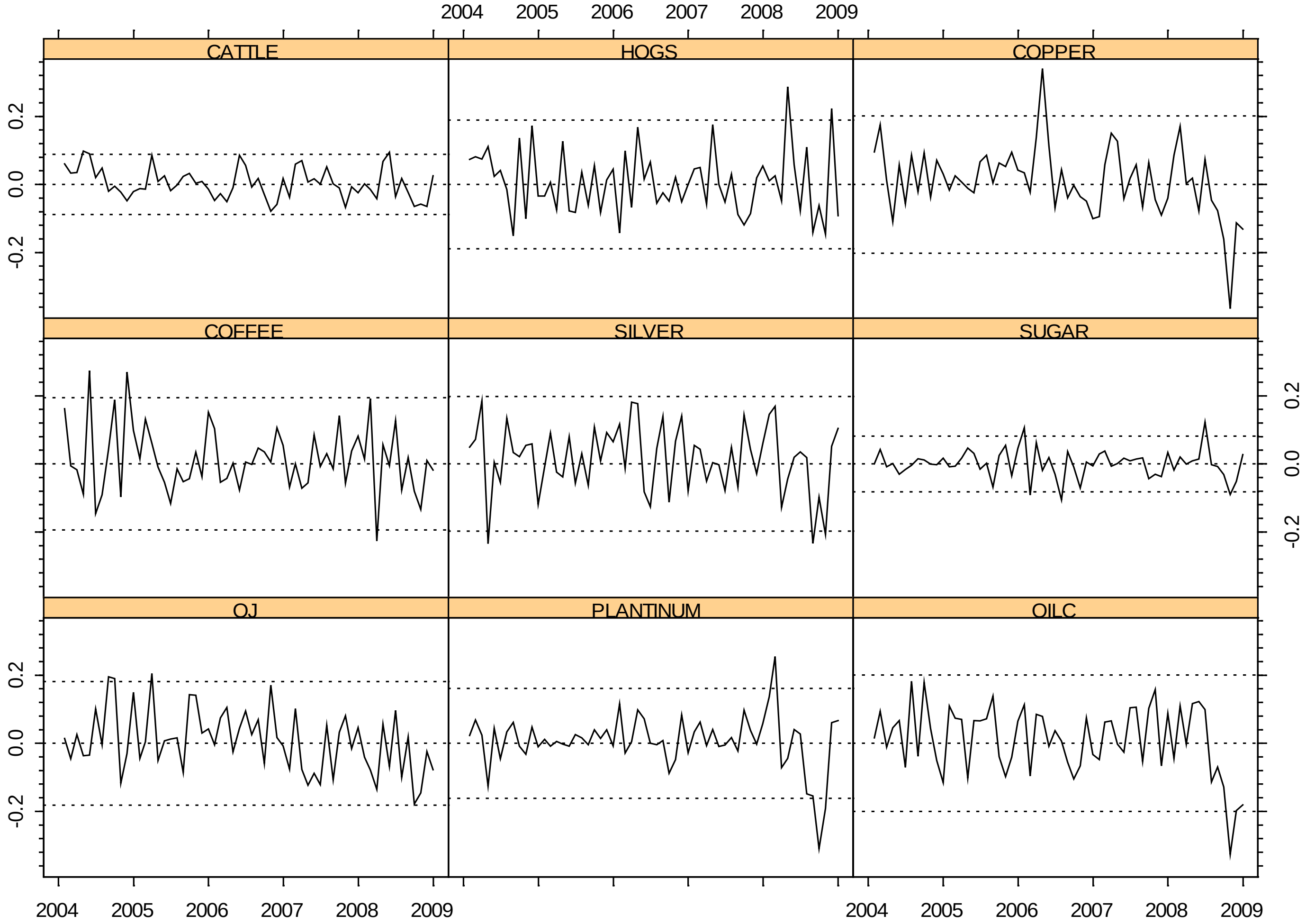
Classical version uses classical sample mean and sample covariance estimates.  Replace them with robust covariances computed with **`MultiRobu()`**

# Portfolio Unusual Movement Alerts

Useful for portfolios with not too many assets, “general” fund-of-funds (FoF), including manager-of-managers portfolios.

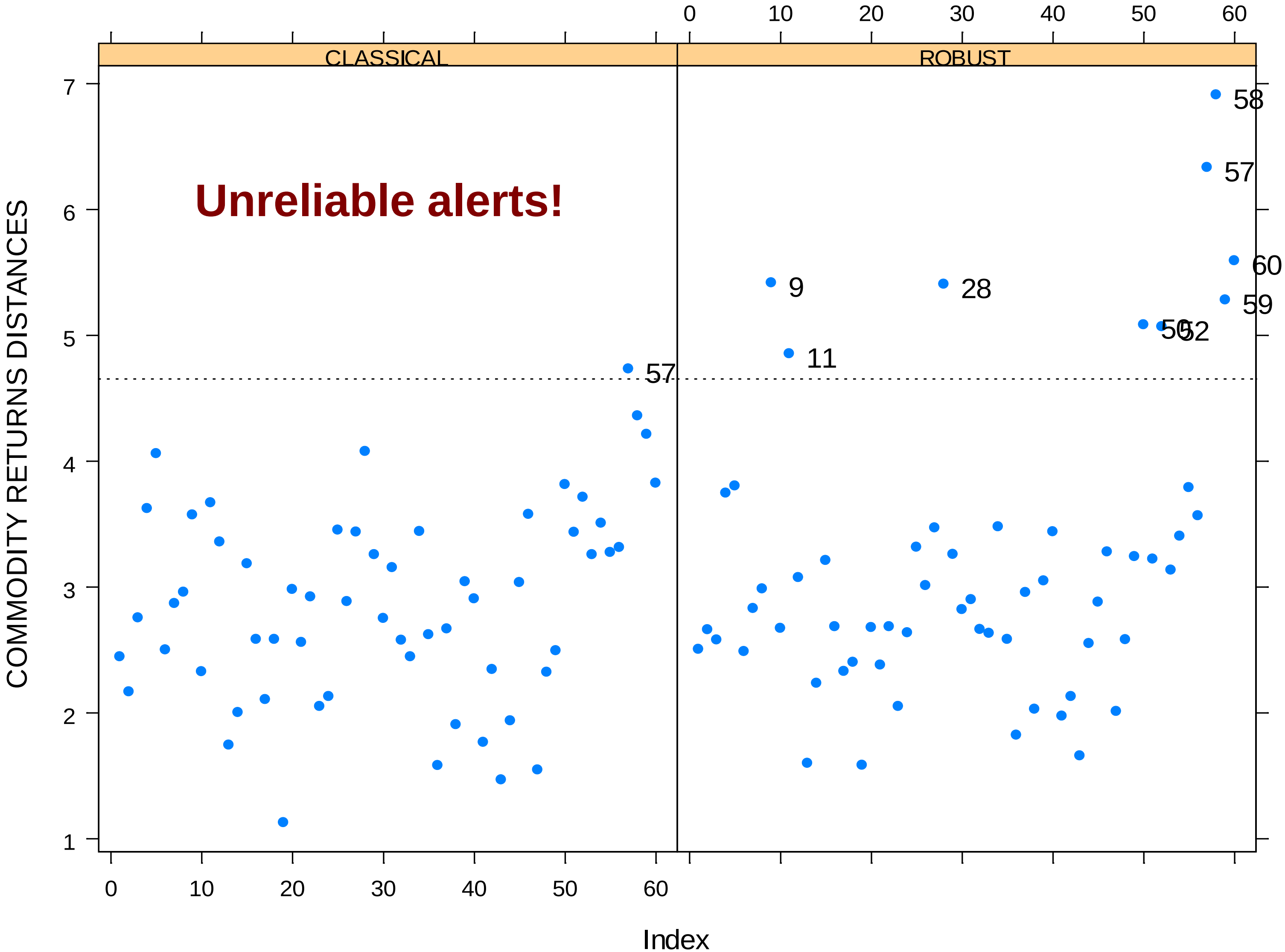
- Retrospective guidance in allocation decisions
- Dynamic unusual movement detection
- Entire fund and style sub-groups

# Commodities Example




(In Appendix A of Martin, Clark and Green, 2011)

# Classical Alerts    Robust Alerts



# 6. Other Important New Robust Methods

- Fast and reliable starting points for initial estimators:
- Robust regularized regression estimators
- Principal components analysis
- Covariance matrix estimators wit missing data
- Covariance matrices with independent outliers in variables
- Mixed linear models
- Generalized linear models
- Regularized estimators of inverse covariance matrix

# 7. Documents to Download

Visit the site **github.com/msalibian/RobStatTM** and look at the bottom of the page to find links to the following:

1. The MMYS book Preface.
2. Scripts that reproduced the examples in MMYS Chapters 4, 5 and 6.
3. MMSY Chapter 11 Descriptions of Data Sets

# References


- Maronna, R. A., Martin, R.D.,Yohai, V. J., and Salibian-Barrera, M. (2018). *Robust Statistics* : *Theory and Methods*, 2nd edition, Wiley.
- Martin, R. D., Clark, A and Green, C. G. (2010). "Robust Portfolio Construction", in *Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques*, J. B. Guerard, Jr., ed., Springer.
- Bailer, H., Maravina, T. and Martin, R. D. (2012). "Robust Betas for Asset Managers", in *The Oxford Handbook of Quantitative Asset Management*, Scherer, B. and Winston, K., editors, Oxford University Press.
- Chapter 6 of Scherer, B. and Martin, R. D. (2004). Modern Portfolio Construction, Springer