COMPUTATIONAL FINANCE & RISK MANAGEMENT
UNIVERSITY of WASHINGTON
Department of Applied Mathematics

ROBUST COVARIANCES

Common Risk versus Specific Risk Outliers

R. Douglas Martin

Professor of Applied Mathematics
Director of Computational Finance Program
University of Washington
doug@amath.washington.edu

R-Finance Conference 2013
Chicago, May 17-18

R Packages and Code Used

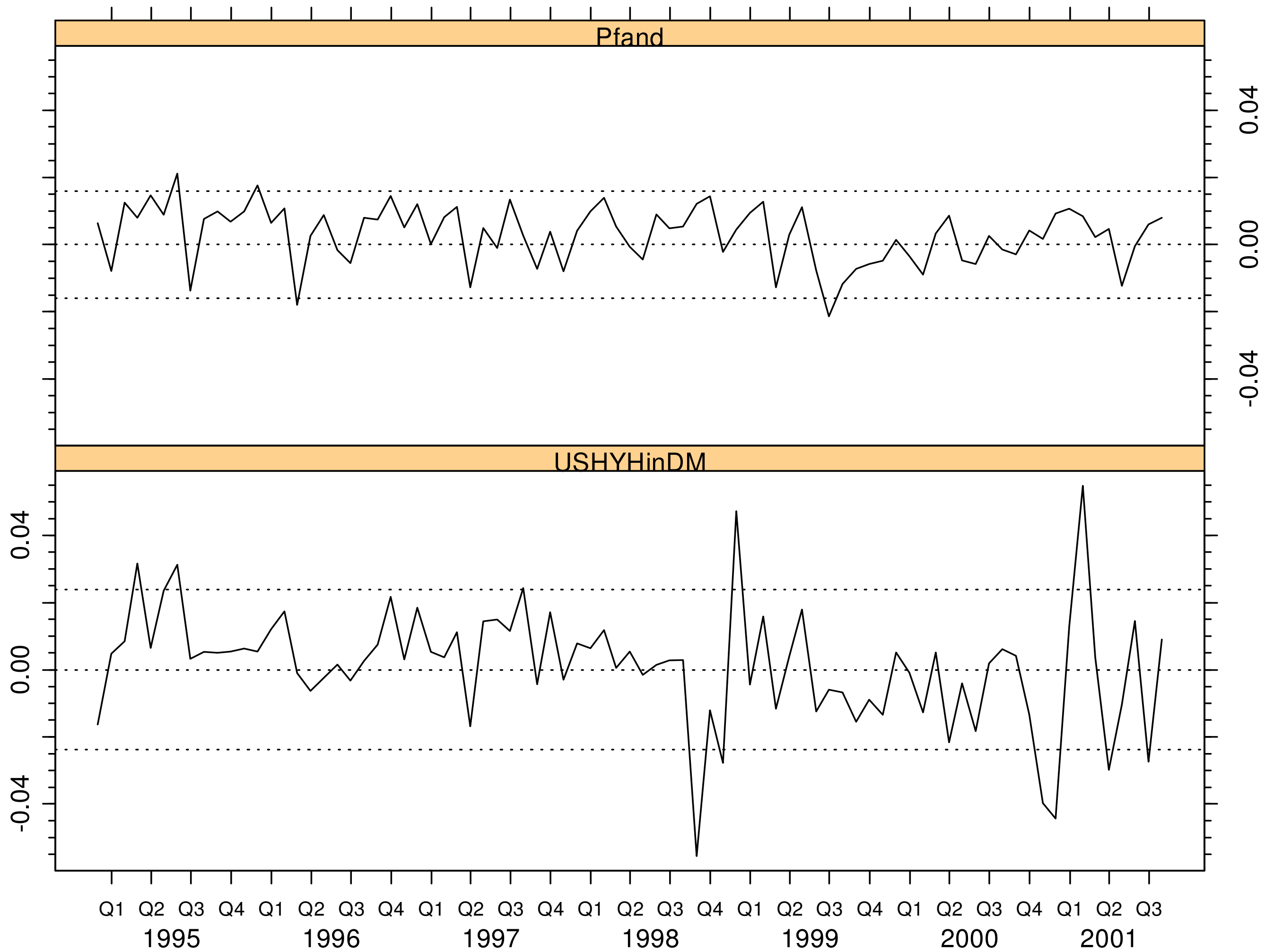
- R *robust* package
- *PerformanceAnalytics* package
- Global minimum variance portfolios with constraints
  - GmvPortfolios.r: gmv, gmv.mcd, gmv.qc, etc.
- Backtesting
  - btShell.portopt.r: btTimes, backtet.weight
  - gmvlo & gmvlo.robust.r

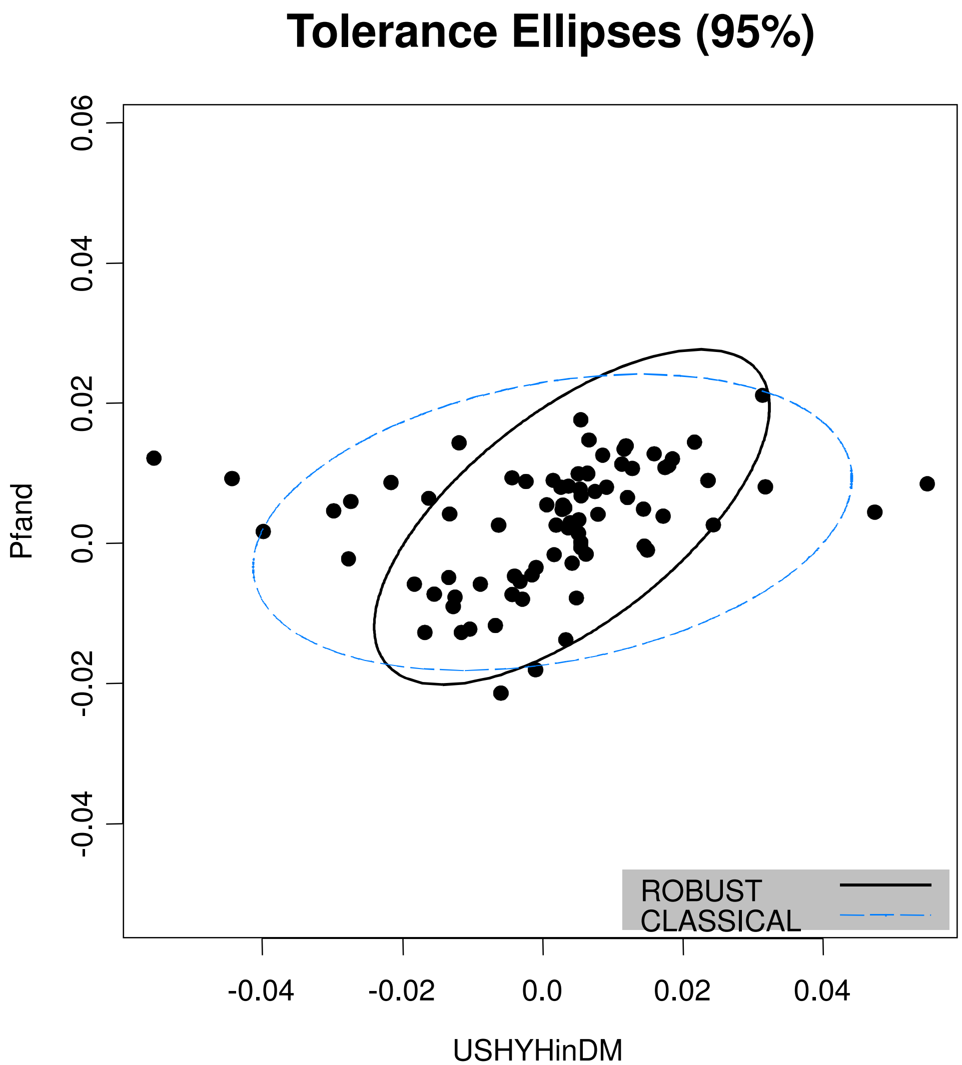
Robust Covariance Uses in Finance

- Asset returns EDA, multi-D outlier detection and portfolio unusual movement alerts
  - SM (2005), MGC (2010), Martin (2012)
- Data cleaning pre-processing
  - BPC (2008)
- Reverse stress testing
  - Example to follow
- **Robust mean-variance portfolio optimization**
  - Is it usefull ???? If so, which method ????

Robust vs. Classical Correlations

(Two assets in a larger fund-of-funds portfolio)





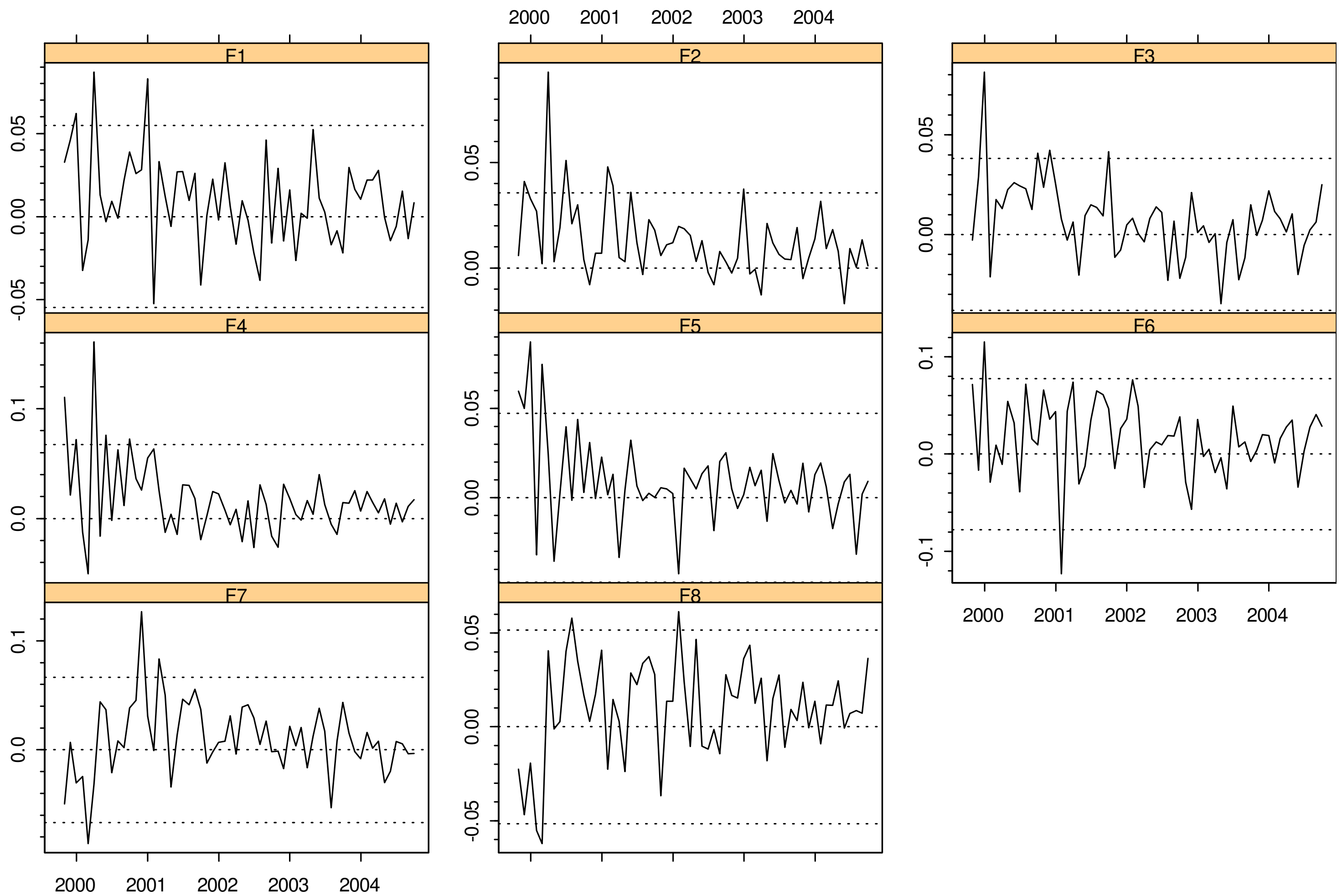
CLASSIC CORR. = .30

What you get from every stats package. Gives an overly optimistic view of diversification benefit!

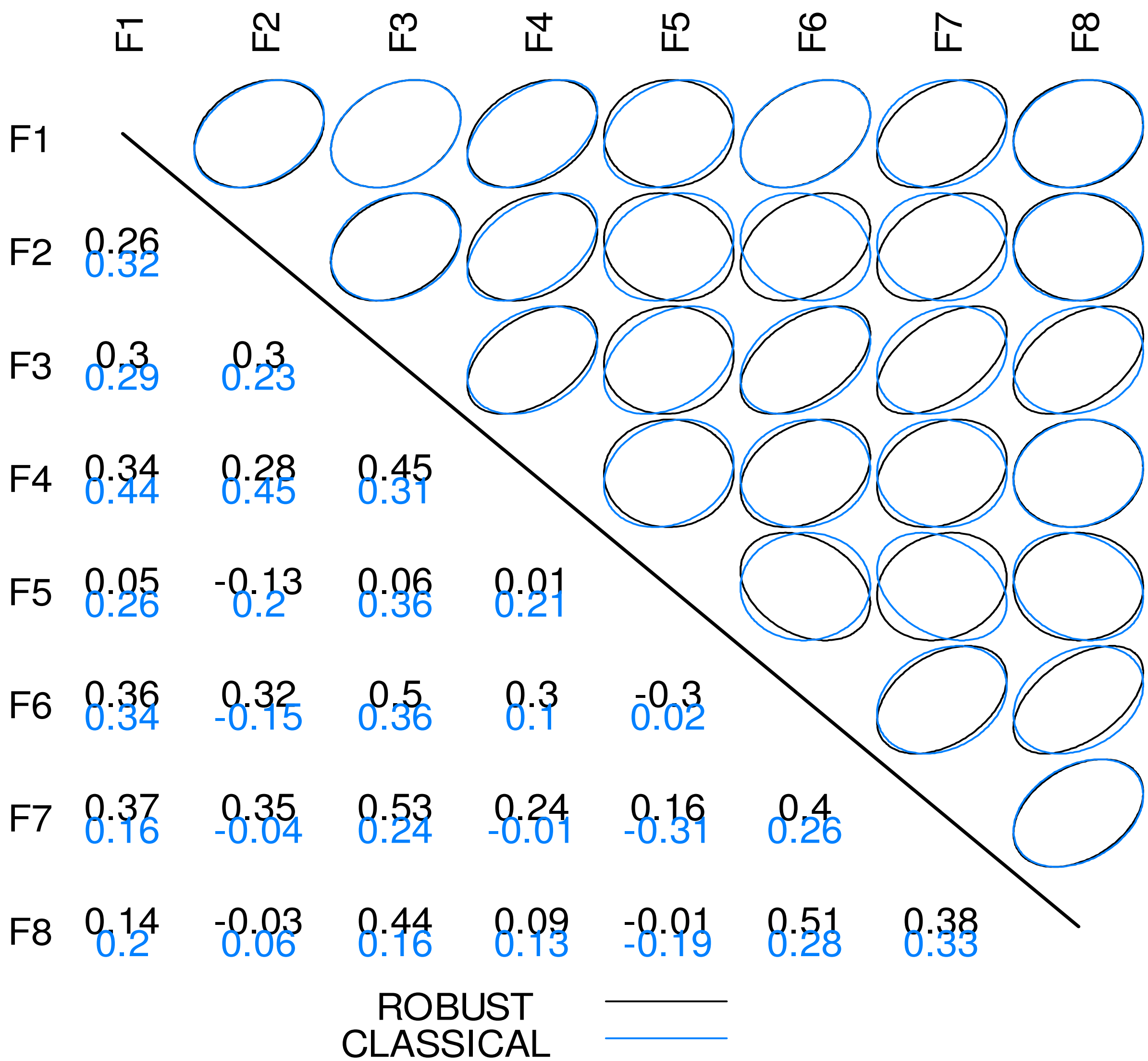
ROBUST CORR. = .65

A more realistic view of a lower diversification benefit!

Hedge Fund Returns Example



Hedge Fund Returns Example



Portfolio Unusual Movement Alerts

Mahalanobis Squared Distance (MSD)

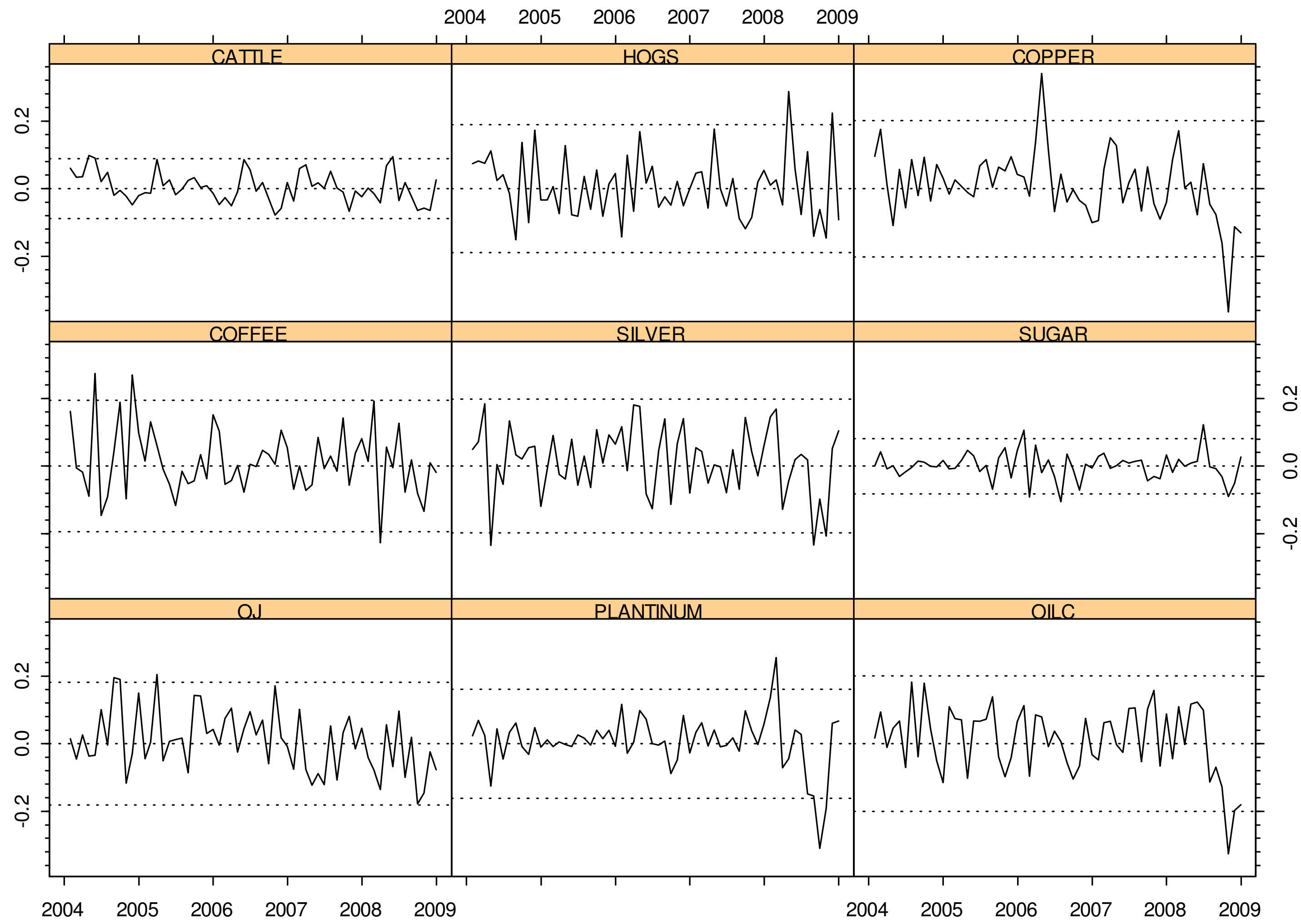
$$d_t^2 = (\mathbf{r}_t - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{r}_t - \hat{\boldsymbol{\mu}})$$

Crucial to use a robust covariance matrix estimate $\hat{\boldsymbol{\Sigma}}$!

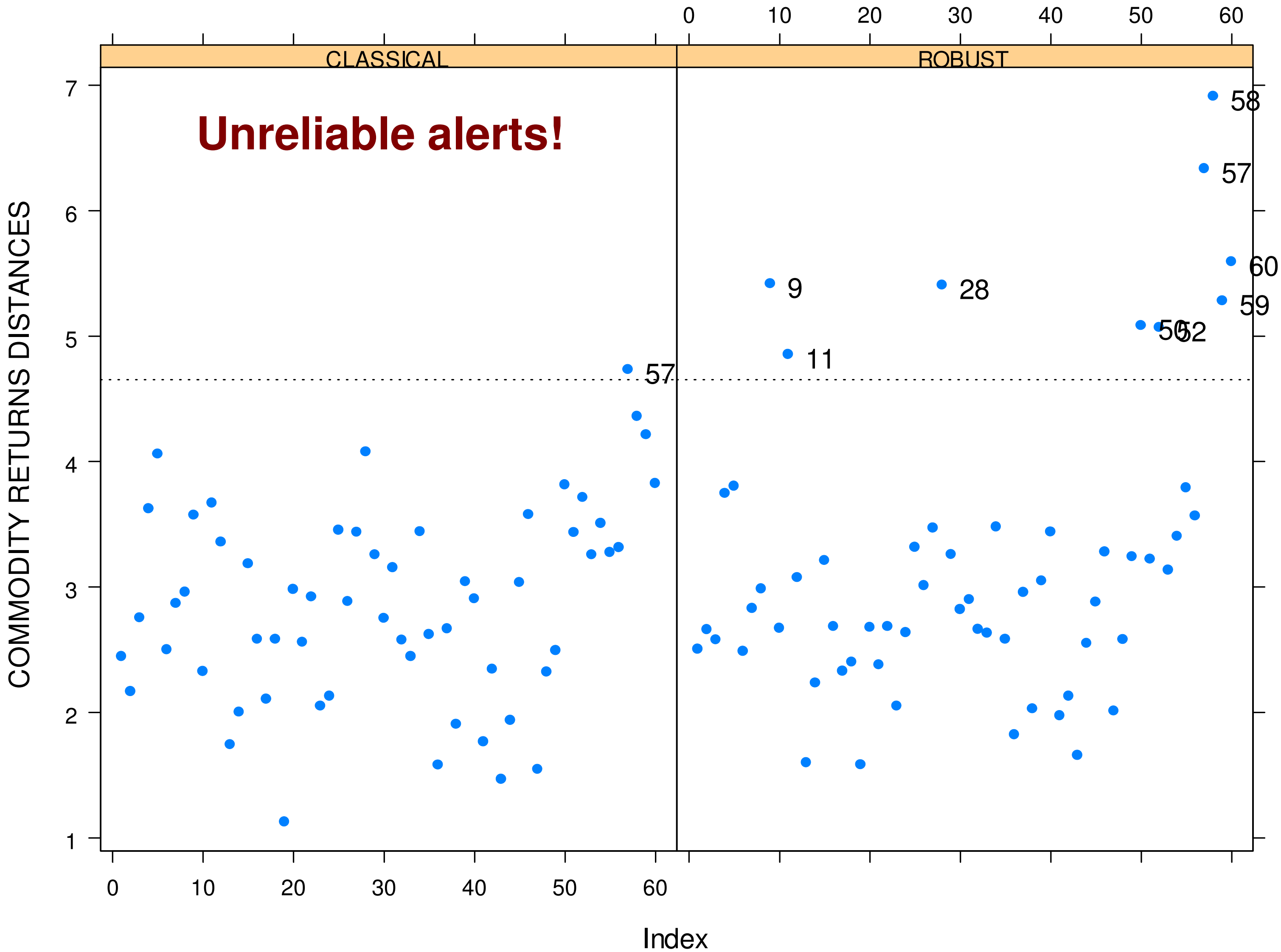
- Retrospective analysis
- Dynamic alerts

Commodities Example

(see Appendix A of Martin, Clark and Green, 2009)



Classical Alerts   Robust Alerts



```r
library(xts)
library(robust)
library(lattice)

ret = read.zoo("commodities9.csv",sep=",",header =
                T,format = "%m/%d/%Y")
ret = as.xts(ret)
ret = ret['2004-01-31/2008-12-31',]
xyplot(ret,layout = c(3,3))  # Not the same as slide
data = coredata(ret)

cov.fm <- fit.models(CLASSICAL = covMLE(data),
       ROBUST = covRob(data,estim = "mcd",quan = .7))
plot(cov.fm,which.plots = 3)
```

Robust Covariance Choices in R “robust”

- Min. covariance determinant (MCD)
- M-estimate (M)
- Donoho-Stahel (DS)

} affine equivariant

- Pairwise estimates (PW)
  - Quadrant correlation and GK versions
  - Positive definite (Maronna & Zamar, 2002)

} not affine equivariant

For details see the R robust package reference manual. See also Chapter 10.2.2 of Pfaff (2013) *Financial Risk Modeling and Portfolio Optimization* with R, Wiley.

The Usual Robustness Outliers Model

$\mathbf{R}$ $T \times n$ table of returns with rows $\mathbf{r}_t$

$$\mathbf{r}_t \sim F = (1-\gamma) \cdot N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \; + \; \gamma \cdot H$$

- **A natural model for common factor outliers**
  - Market crashes

1. Probability of a row $\mathbf{r}$ containing an outlier is independent of the dimension n, so the majority of the rows of $\mathbf{R}$ are outlier-free.
2. Fraction γ of rows that have outliers is unchanged under affine transformations, so use affine equivariant estimators, e.g., MCD

Independent Outliers Across Assets (IOA)

Let $B_i = 1\ (0)$ if asset i is (is not) an outlier. (AKMZ, 2002 and AVYZ, 2009)

Assume $B_1, B_2, \cdots, B_n$ are independent with $P(B_i) = \gamma_i$

- **A natural model for specific risk outliers**

Suppose for example that $P(B_i) = \gamma, \quad i = 1, 2, \cdots, n$. Then the probability of a row $\mathbf{r}_t$ not containing an outlier is $(1-\gamma)^n$, which decreases rapidly with increasing p . E.g., for $\gamma = .05$:

# of assets n	5	10	15	20
prob. clean row	.77	.60	.46	.36

N.B. Affine transformations increase the percent of rows with outliers, so no need to restrict attention to affine equivariant estimators.

Choice of Outliers Model and Estimator

Both are useful, but:

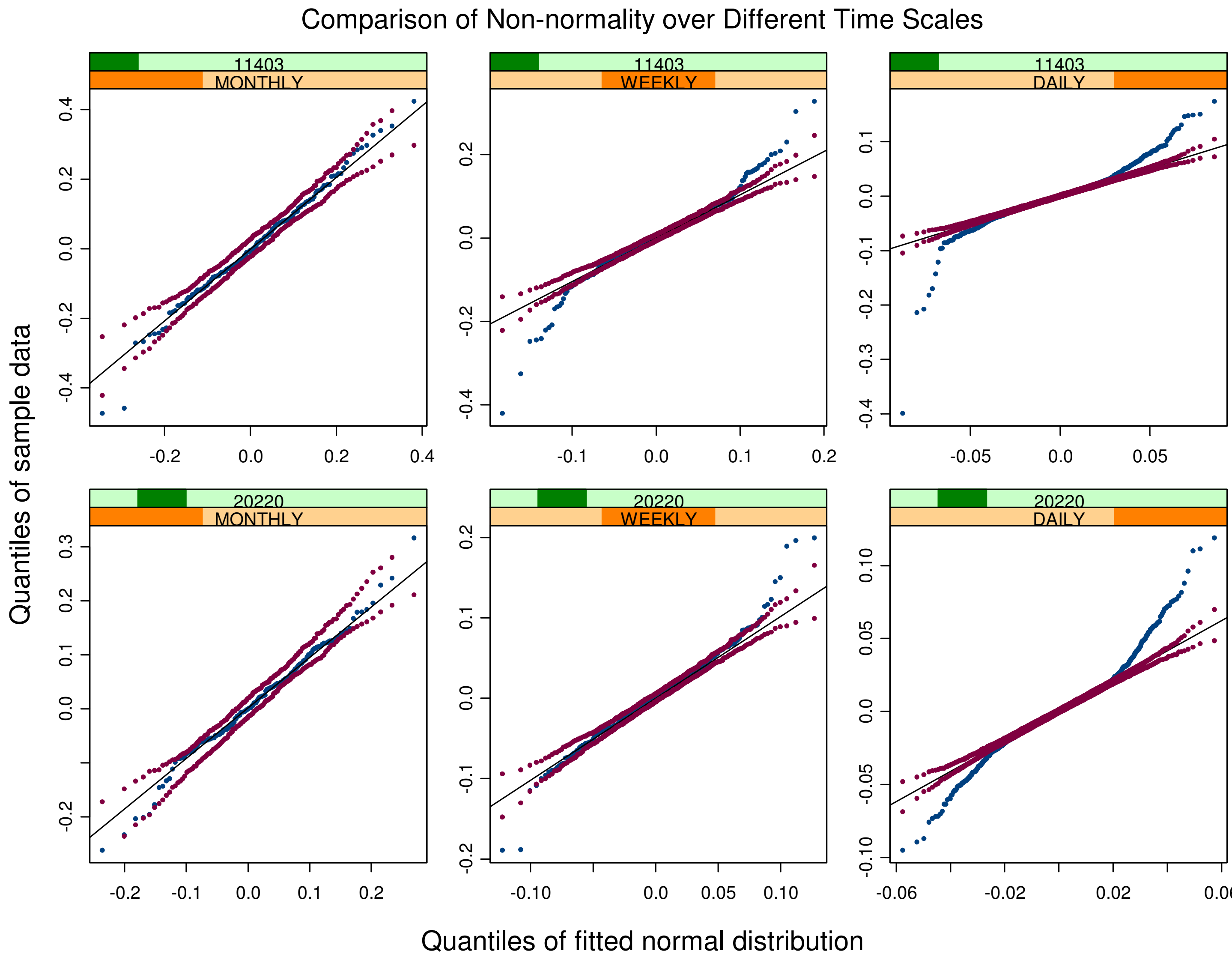
- The usual outliers model handles market events outliers and for these an affine equivariant robust covariance matrix estimator will suffice, e.g., MCD.
- The independent outliers across assets model is needed for specific risk outliers, and for these one may need to use a pairwise estimator to avoid breakdown!
- **Goal: Determine when pairwise robust covariance matrix estimator performs better than MCD, etc.**

Asset Class & Frequency Considerations

Specific risk outliers are more frequent in the case of:

- Higher returns frequency, e.g., weekly and daily
- Smaller market-cap stocks
- Hedge funds
- Commodities
- … ???

Non-Normality Increases with Frequency



Outlier Detection Rule for Counting

$\hat{\mu} =$ optimal 90% efficient bias robust location estimate*

$\hat{s} =$ associated robust scale estimate*

Outliers: returns outside of $(\hat{\mu} - \hat{s} \cdot 2.83, \ \ \hat{\mu} + \hat{s} \cdot 2.83)$

Probability of normal return being an outlier: 0.5%

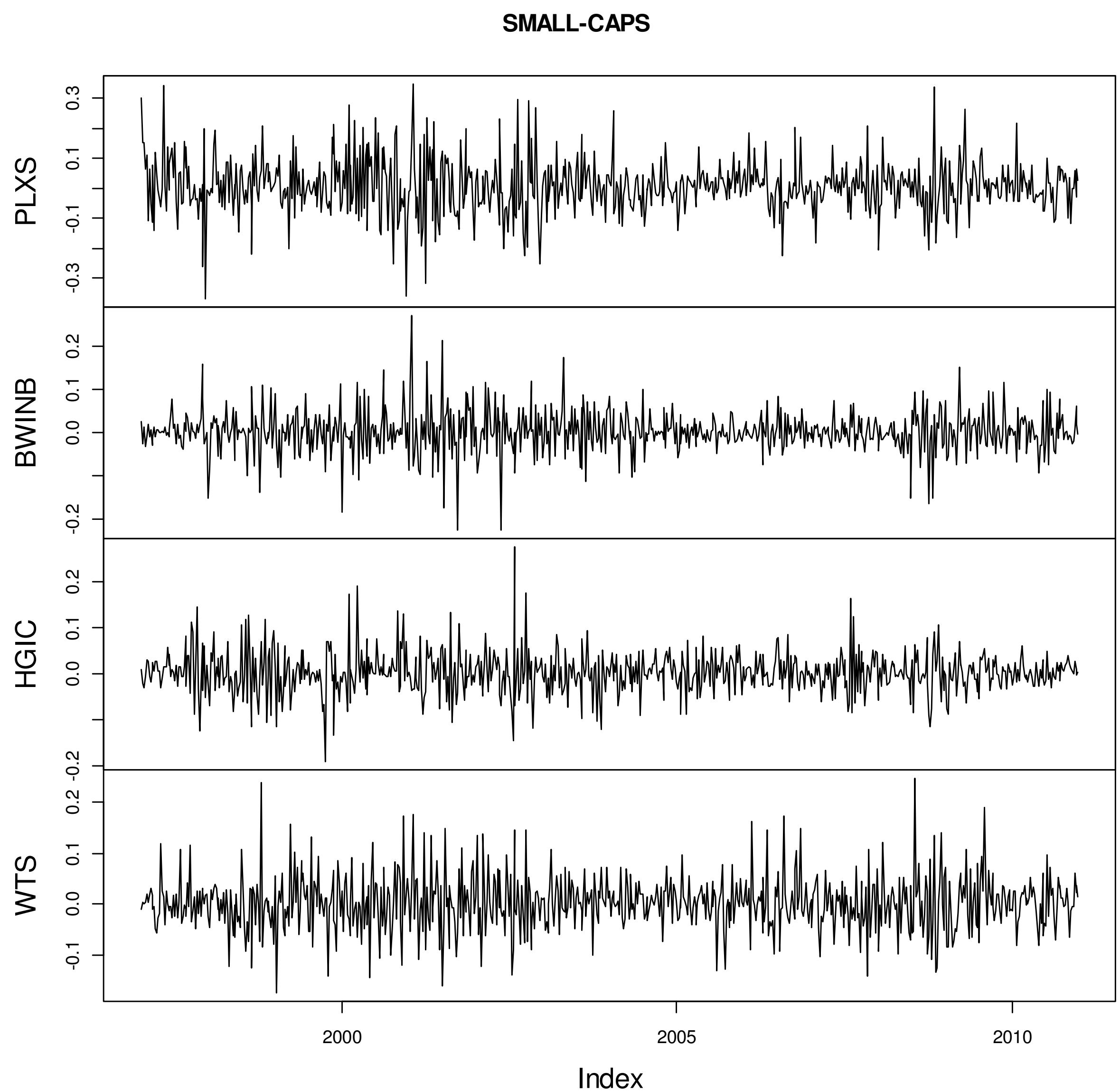
* Use lmRob with intercept only in R package *robust*

Empirical Study of IOA Model Validity

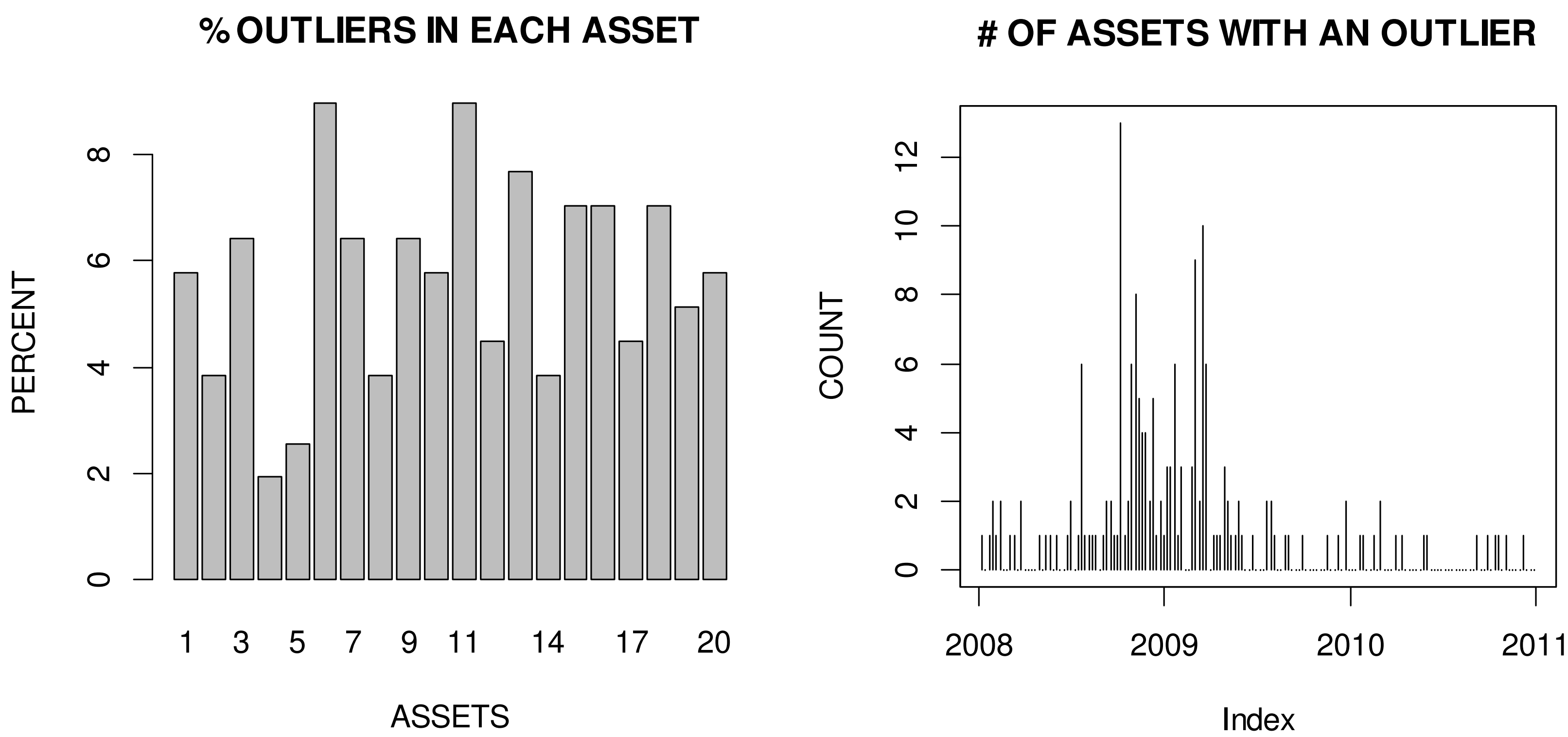
- Four market-cap groups of 20 stocks, weekly returns
- 1997 – 2010 in three regimes:
  - 1997-01-07 to 2002-12-31
  - 2003-01-07 to 2008-01-01
  - 2008-01-08 to 2010-12-28

1. Estimate outlier probability γ_i for each asset, and hence the probability $\prod_{i=1}^{n}(1-\gamma_i)$ that a row is free of outliers under the IOA model.
2. Directly estimate the probability γ_{row} that a row has at least one outlier.
3. Compare results from 1 and 2 across market-caps and regimes.

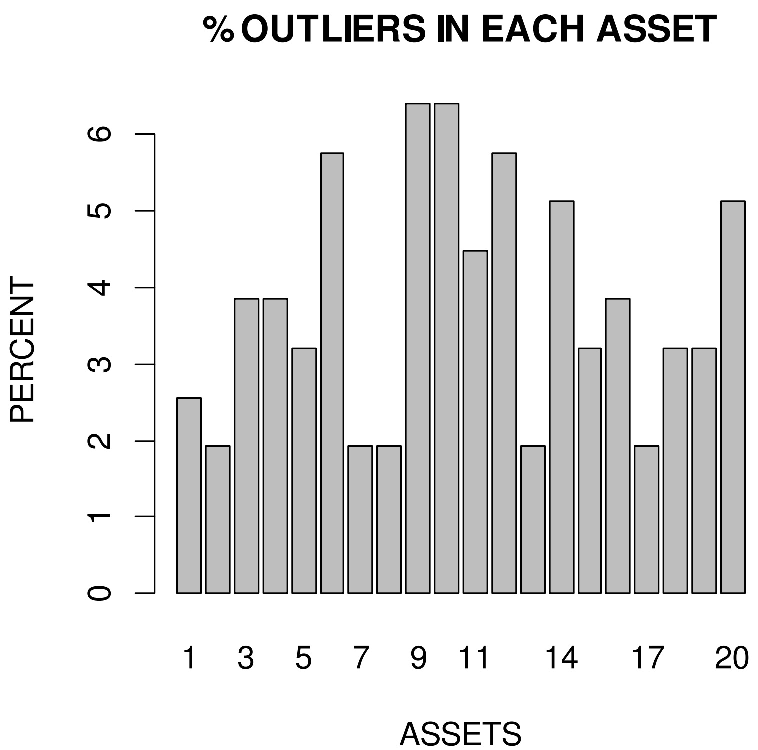
4 of the 20 Small-Caps for Entire History

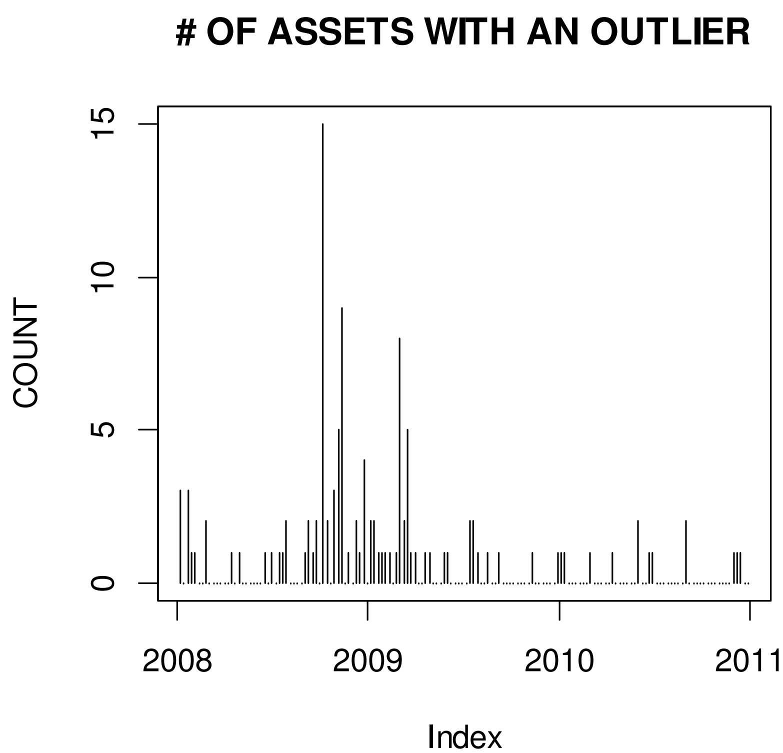


Small-Caps Outliers in Third Regime



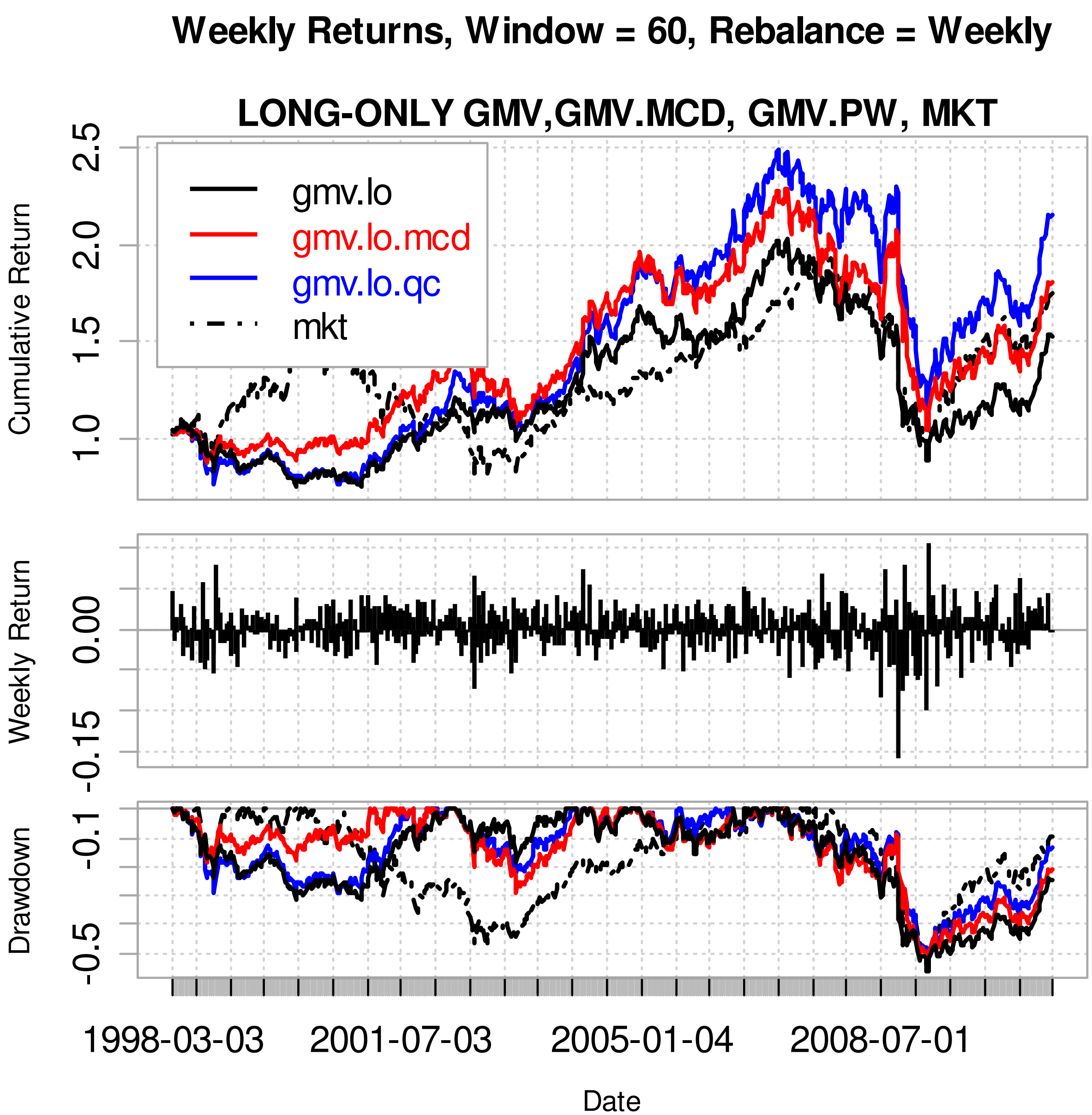
Large-Caps Outliers in Third Regime

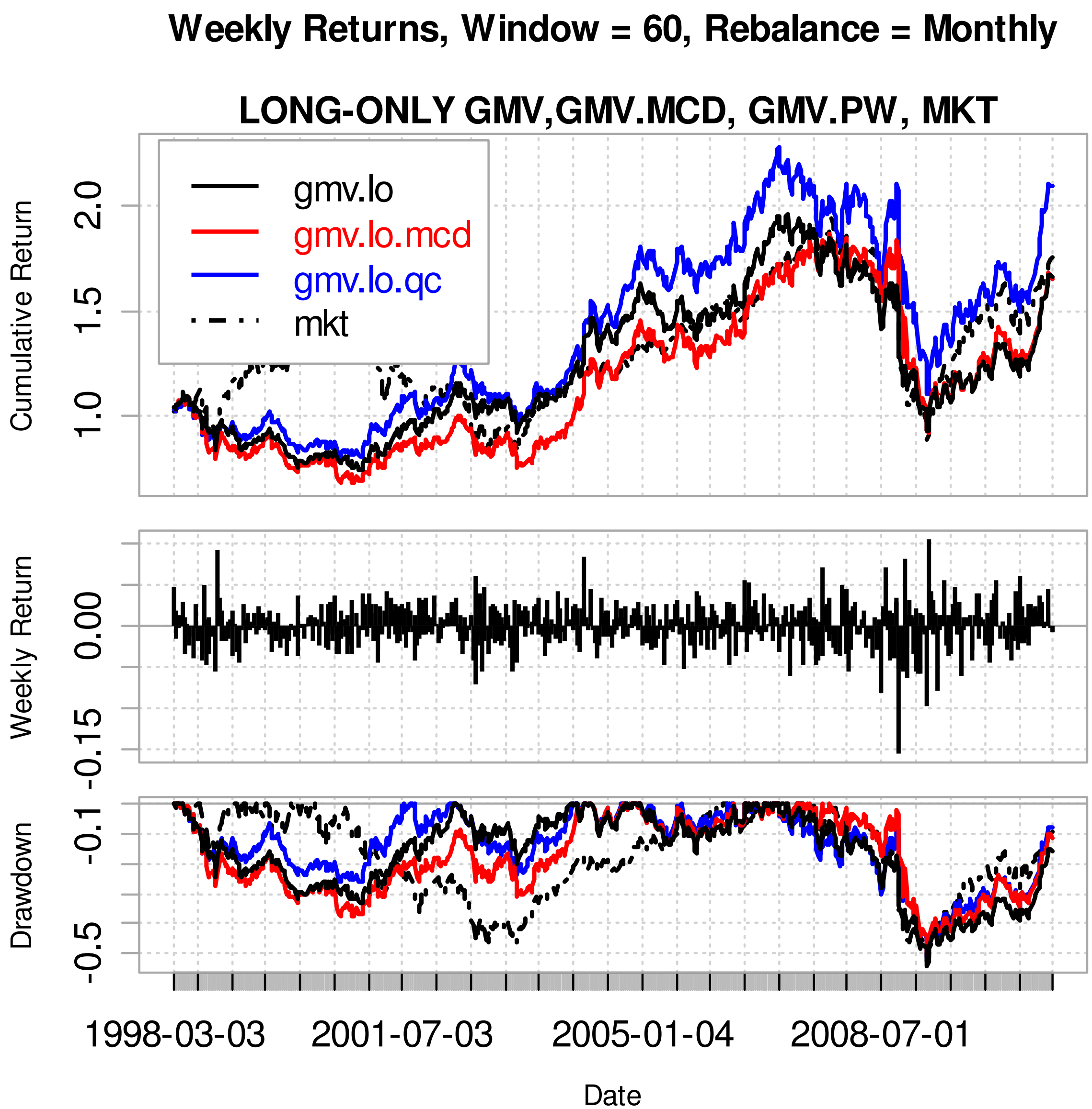




Evaluation of IOA Model for Weekly Returns

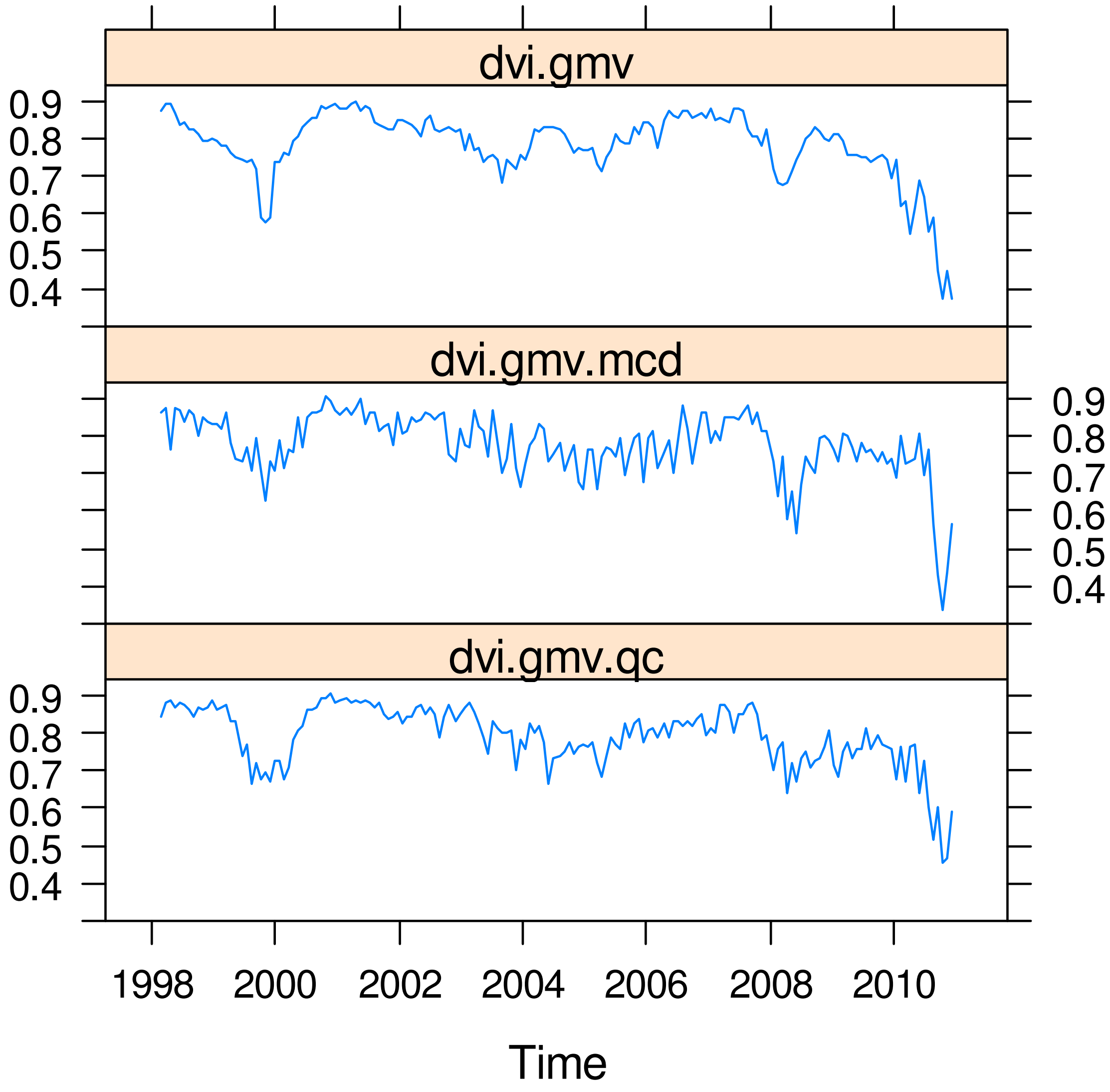
1997-01-07 to 2002-12-31	MICRO	SMALL	MID	LARGE
% Clean Rows IOA Model	32	39	46	59
% Clean Rows Direct Count	37	48	58	69
2003-01-07 to 2008-01-01	**MICRO**	**SMALL**	**MID**	**LARGE**
% Clean Rows IOA Model	33	46	52	57
% Clean Rows Direct Count	37	49	62	66
2008-01-08 to 2010-12-28	**MICRO**	**SMALL**	**MID**	**LARGE**
% Clean Rows IOA Model	23	31	39	46
% Clean Rows Direct Count	43	48	57	62


Weekly Returns, Window = 60, Rebalance = Weekly
LONG-ONLY GMV,GMV.MCD, GMV.PW, MKT
gmv.lo
gmv.lo.mcd
gmv.lo.qc
mkt
Cumulative Return
1.0
1.5
2.0
2.5
Weekly Return
0.00
-0.15
Drawdown
-0.1
-0.5
1998-03-03
2001-07-03
2005-01-04
2008-07-01
Date



Weekly Returns, Window = 60, Rebalance = Monthly
LONG-ONLY GMV,GMV.MCD, GMV.PW, MKT
gmv.lo
gmv.lo.mcd
gmv.lo.qc
mkt
Cumulative Return
1.0
1.5
2.0
Weekly Return
-0.15
0.00
Drawdown
-0.5
-0.1
1998-03-03
2001-07-03
2005-01-04
2008-07-01
Date

HHI Diversification Index (sum-of-squared wts.)



Back-Test Code

```
library(PerformanceAnalytics)
library(robust)
source("GmvPortfolios.r")
source("btShell.portopt.r")
source("btTimes.r")

# Diversification Index Function
dvi =function(x){1-sum(x^2)}

# Input returns
ret.all = read.zoo("smallcap_weekly.csv",sep=",",header =
                    T,format = "%m/%d/%Y")
mkt = ret.all[,"VWMKT"]
ret = ret.all[,1:20]
n.assets <- ncol(ret)

# get returns dates
all.date = index(ret)
```

```
# compute the backtest times
t.mw <- btTimes.mw(all.date, 4, 60)

# backtesting
weight.gmv.lo <- backtest.weight(ret, t.mw,gmv.lo)$weight
weight.gmv.lo.mcd <- backtest.weight(ret, t.mw,
                                     gmv.lo.mcd)$weight
weight.gmv.lo.qc <- backtest.weight(ret, t.mw,
                                    gmv.lo.qc)$weight
# The Diversification Index Plots
gmvdat = coredata(weight.gmv.lo)
gmvdat.mcd = coredata(weight.gmv.lo.mcd)
gmvdat.qc = coredata(weight.gmv.lo.qc)
dvi.gmv = apply(gmvdat,1,dvi)
dvi.gmv.mcd = apply(gmvdat.mcd,1,dvi)
dvi.gmv.qc = apply(gmvdat.qc,1,dvi)
dvi.all = cbind(dvi.gmv,dvi.gmv.mcd,dvi.gmv.qc)
dvi.all.ts = as.zoo(dvi.all)
index(dvi.all.ts) = index(weight.gmv.lo)
xyplot(dvi.all.ts, scales = list(y="same"))
```

```
# compute cumulative returns of portfolio
gmv.lo <- Return.rebalancing(ret, weight.gmv.lo)
gmv.lo.mcd <- Return.rebalancing(ret, weight.gmv.lo.mcd)
gmv.lo.qc <- Return.rebalancing(ret, weight.gmv.lo.qc)

# combined returns
ret.comb <- na.omit(merge(gmv.lo, gmv.lo.mcd, gmv.lo.qc, mkt,
                    all=F))

# return analysis
charts.PerformanceSummary(ret.comb,wealth.index = T,
  lty = c(1,1,1,4),colorset = c("black","red","blue","black"),
       cex.legend = 1.3,cex.axis = 1.3, cex.lab = 1.5, main =
      "Weekly Returns, Window  = 60, Rebalance = Monthly
       \n LONG-ONLY GMV,GMV.MCD, GMV.PW, MKT")
```

"Statistics is a science in my opinion, and it is no more a branch of mathematics than are physics, chemistry and economics; for if its methods fail the test of experience – not the test of logic – they will be discarded"

- J. W. Tukey

Thank You!

"Statistics is a science in my opinion, and it is no more a branch of mathematics than are physics, chemistry and economics; for if its methods fail the test of experience – not the test of logic – they will be discarded"

- J. W. Tukey

Thank You!

References

- Alqallaf, Konis, Martin and Zamar (2002). “Scalable robust covariance and correlation estimates for data mining”, Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 14-23. ACM.
- Scherer and Martin (2005). *Modern Portfolio Optimization*, Chapter 6.6 – 6.9, Springer
- Maronna, Martin, and Yohai (2006). *Robust Statistics* : *Theory and Methods*, Wiley.
- Boudt, Peterson & Croux (2008). “Estimation and Decomposition of Downside Risk for Portfolios with Non-Normal Returns”, *Journal of Risk*, 11, No. 2, pp. 79-103.
- Alqallaf, Van Aelst, Yohai and Zamar (2009). “Propagation of Outliers in Multivariate Data”, Annals of Statistics, 37(1). p.311-331.
- 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.
- Martin, R. D. (2012). “Robust Statistics in Portfolio Construction”, Tutorial Presentation, R-Finance 2012, Chicago,