APPLICATIONS OF ROBUST STATISTICS IN FINANCE

R. Douglas Martin*

Current and Future Challenges in Robust Statistics
Banff Center, Canada
Wed. November 18, 2015
THREE TOPICS: TWO SHORT, ONE LONG

1. Robust detection of fundamental factor model exposures outliers (skipped)

2. Robust covariances for mean-variance portfolio optimization: Classic multivariate outliers model versus independent outliers in assets (IOA) model (short)

3. Non-parametric versus parametric expected shortfall: the influence function approach
1. ROBUST DETECTION OF FUNDAMENTAL FACTOR MODEL EXPOSURES OUTLIERS*

1. The fundamental factor model context
2. The data for this example
3. The MCD estimator
4. The results

* Joint work with Chris Green, Washington State Investment Board and UW Statistics PhD Candidate
1. Fundamental Factor Model

Work-horse of commercial portfolio optimization and risk management software (MSCI Barra, Axioma, Northfield).

\[ r_t = B_{t-1} f_t + \varepsilon_t \]

Exposures matrix \((N \times K)\) \quad Factor returns \((K \times 1)\)

\[ B_{t-1} = (b_1, \ldots, b_q \mid b_{q+1}, \ldots, b_K) \]

Continuous risk factor exposures \quad Industry exposures, each entry a 1 or 0

For example: firm size, E/P, B/M, momentum, leverage, etc.
2. The Data

Number of companies:

Fundamental data (exposures)

- Large-cap stocks (from CRSP database)
- **Size/ME** (log. mkt. cap.), B/M, E/P, momentum  (Compustat)
2. The MCD Estimator

• Cerioli’s IRMCD method
• Test for outliers using Cerioli approach and Bonferroni-corrected significance level of $0.025/325 \approx 0.00008$.
• Use a conservative version of MCD that uses approx. 95% of the data to estimate the robust dispersion matrix

• R package: CerioliOutlierDetection
  http://cran.r-project.org/web/packages/CerioliOutlierDetection/index.html
• Working Paper
3. The Results

![Graph showing changes in fraction of violations over time with two lines representing different distances and data alterations. The red line represents classical distances with unaltered data, and the blue line represents robust distances with unaltered data and MCD(0.95) estimator.]
2. ROBUST COVARIANCES FOR MEAN-VARIANCE OPTIMIZATION

Classic Multivariate Outliers Model vs. Independent Outliers in Assets (IOA) Model
The Two Outliers Models

\[ \mathbf{R} \quad T \times N \text{ table of returns with rows } \mathbf{r}_t \]

Model for common factor outliers/market crashes

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

Single Factor Market Model for a Portfolio of Assets:

\[ r_{t,i} = \alpha_i + \beta_i \cdot r_{M,t} + \epsilon_{t,i}, \quad t = 1, 2, \ldots, T, \quad i = 1, 2, \ldots, N \]

A market return outlier at time \( t \) causes an outlier in all assets (to various degrees), i.e., an outlying row.
Model for independent outliers across assets (IOA) 
Alqallaf, Van Aelst, Yohai and Zamar (2009)

Let $B_i = 1$ (0) if asset $i$ is (is not) an outlier. (AVYZ, 2009)
Assume $B_1, B_2, \ldots, B_N$ are independent with $P(B_i) = \gamma_i$

Probability of an outlier-free row $r_t$: $\prod_{i=1}^{N} (1 - \gamma_i)$

For $\gamma = .05$ and $N = 20$ the probability of a clean row is .36

Single Factor Market Model for a Portfolio of Assets:

$$r_{t,i} = \alpha_i + \beta_i \cdot r_{M,t} + \varepsilon_{t,i}, \quad t = 1, 2, \ldots, T, \quad i = 1, 2, \ldots, N$$

Asset specific (risk) time $t$ outliers tend to be independent
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 $\gamma_i$ 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 $1 - \gamma_{row}$ that a row is outlier free.

3. Compare results from 1 and 2 across market-caps and regimes.
4 of the 20 Small-Caps for Entire History
Outlier Detection Rule for Counting

\( \hat{\mu} = \text{optimal 90\% efficient bias robust location estimate}\)
\( \hat{s} = \text{associated robust scale estimate}\)

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
Small-Caps Outliers in Third Regime

% OUTLIERS IN EACH ASSET

# OF ASSETS WITH AN OUTLIER
Large-Caps Outliers in Third Regime

% OUTLIERS IN EACH ASSET

# OF ASSETS WITH AN OUTLIER
## Evaluation of IOA Model for Weekly Returns

<table>
<thead>
<tr>
<th>Period</th>
<th>MICRO</th>
<th>SMALL</th>
<th>MID</th>
<th>LARGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997-01-07 to 2002-12-31</td>
<td>32</td>
<td>39</td>
<td>46</td>
<td>59</td>
</tr>
<tr>
<td>% Clean Rows IOA Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Clean Rows Direct Count</td>
<td>37</td>
<td>48</td>
<td>58</td>
<td>69</td>
</tr>
<tr>
<td>2003-01-07 to 2008-01-01</td>
<td>33</td>
<td>46</td>
<td>52</td>
<td>57</td>
</tr>
<tr>
<td>% Clean Rows IOA Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Clean Rows Direct Count</td>
<td>37</td>
<td>49</td>
<td>62</td>
<td>66</td>
</tr>
<tr>
<td>2008-01-08 to 2010-12-28</td>
<td>23</td>
<td>31</td>
<td>39</td>
<td>46</td>
</tr>
<tr>
<td>% Clean Rows IOA Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Clean Rows Direct Count</td>
<td>43</td>
<td>48</td>
<td>57</td>
<td>62</td>
</tr>
</tbody>
</table>

Main Result: The differences % Clean Rows Direct Count - % Clean Rows IOA Model tend to be larger during the first and third time periods, which is apparently due to the fact that market crashes cause highly correlated outliers in most assets, for which the IOA model does not hold. However the IOA model seems to describe much of the behavior, particularly so for the middle time period.
Cumulative Return
Weekly Returns, Window = 60, Rebalance = Weekly
LONG-ONLY GMV, GMV.MCD, GMV.PW, MKT

Weekly Return

Date

Drawdown


-0.5 -0.1 -0.15

0.00

0.00

1.0 1.5 2.0 2.5
Weekly Returns, Window = 60, Rebalance = Monthly

LONG-ONLY GMV, GMV.MCD, GMV.PW, MKT

Cumulative Return

Weekly Return

Drawdown

Date

“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
References


- Scherer and Martin (2005). Modern Portfolio Optimization, Chapter 6.6 – 6.9, Springer
3. NON-PARAMETRIC VERSUS PARAMETRIC EXPECTED SHORTFALL:

The Influence Function Approach*

*Joint work with Shengyu Zhang, PhD, First Vice President, Homestreet Bank Seattle, WA
1. VALUE-AT-RISK

- **Valued-at-Risk (VaR)**
  - A capital weighted 1% or 5% tail probability quantile
    \[ \text{VaR}(\gamma; F) = -W \cdot q(\gamma; F) \quad q(\gamma; F) = F^{-1}(\gamma) \]

- **VaR is the de facto risk management standard**
  - Originated in the 1990’s at JP Morgan

- **VaR is simply not informative enough**
  - Doesn’t tell you anything about the size of losses beyond VaR
2. EXPECTED SHORTFALL (ES)

▪ What it is
  – The average of the losses beyond VaR
  – Much more informative than VaR
  – ES a coherent risk measure, VaR is not

▪ Basel III move to ES instead of VaR

▪ Alternative names for ES
  – Conditional Value-at-Visk (CVaR)
  – Expected Tail Loss (ETL)
ES vs VaR: Event Driven Hedge Funds Index
ES vs VaR: Event Driven Hedge Funds Index
Expected Shortfall Formula

Taking loss as negative for math but positive for some plots

\[ q(\gamma; F) = F^{-1}(\gamma) \]

\[ ES(\gamma; F) = E(r \mid r \leq q(\gamma; F)) \]

\[ ES(\gamma; F) = \frac{1}{\gamma} \int_{r \leq q(\gamma; F)} r \cdot dF(r) \]
Non-Parametric Estimators

Represent the asymptotic value of any estimator as a functional on a space of distribution functions:

\[ \theta = \theta(F) \quad \rightarrow \quad \hat{\theta} = \theta(F_n) \]

where \( F_n \) is the empirical distribution function (edf).

Expected shortfall estimator:

\[
ES(\gamma) = \frac{1}{\gamma} \int_{r \leq q(\gamma; F_n)} r \cdot dF_n(r) = \frac{1}{\lceil n\gamma \rceil} \sum_{i=1}^{\lceil n\gamma \rceil} r_{(i)}
\]
Parametric Estimators

First get a parametric representation of the risk measure:

Normal distribution ES

\[ ES_\gamma(\mu, \sigma) = \mu - \frac{\phi(z_\gamma)}{\gamma} \cdot \sigma \]

\( z_\gamma \) is the inverse cumulative distribution function of the normal distribution.

t-distribution ES

\[ ES_\gamma(\mu, s, \nu) = \mu - \frac{g_{\gamma,\nu}}{\gamma} \cdot s \]

\[ g_{\gamma,\nu} \triangleq \sqrt{\frac{\nu}{\nu - 2}} \cdot f_{\nu - 2} \left( \sqrt{\frac{\nu - 2}{\nu}} \cdot q_{\gamma,\nu} \right) \]

\( f \) is the density function and \( q \) is the quantile function of the t-distribution.

11/18/2015
Then substitute estimators for the unknown parameters

<table>
<thead>
<tr>
<th>ES Parametric Estimators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal Distribution</td>
</tr>
<tr>
<td>$\hat{\mu} - \frac{\phi(z_\gamma)}{\gamma} \cdot \hat{\sigma}$</td>
</tr>
</tbody>
</table>

**MLE parameter estimators $\rightarrow$ MLE risk estimator**
Normal vs Fat-Tailed Parametric ES/ETL Estimator

OXM Returns

0 100 200 300 400 500
-0.15 -0.10 -0.05 0.0 0.05 0.10
Normal vs Fat-Tailed Parametric ES/ETL Estimator

1% STABLE ETL vs. NORMAL VAR AND ETL: $1M OVERNIGHT

- Normal VaR = $47K
- Normal ETL = $51K
- Stable ETL = $147K
3. ES INFLUENCE FUNCTIONS

Non-Parametric ES IF Formulas

\[ F_p = (1 - p)F + p \cdot \delta_r, \quad 0 \leq p < 1 \]

\[ ETL(\gamma; F_p) = \frac{1}{\gamma} \int_{r \leq q(\gamma; F_p)} r \cdot dF_p(r) \]

\[ IF_{ETL}(r; F) = \left. \frac{d}{dp} ETL(F_p) \right|_{p=0} \]

Straightforward but tedious calculations give complex formulas. But the plots tell the story.
Non-Parametric VaR and ES Influence Functions

Straightforward formula calculations available in paper
Parametric Risk Estimators

Risk Measure
\[ \rho(\theta) = \rho(\theta_1, \theta_2, \ldots, \theta_K) \]

Risk Estimator
\[ \hat{\rho} = \rho(\hat{\theta}) = \rho(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K) \]

General IF Formula
\[ IF_{\hat{\rho}}(r; F) = \nabla \rho(\theta)' \cdot IF_{\hat{\theta}}(r) \]

MLE IF Formula
\[ IF(r; \hat{\rho}, F) = \nabla \rho(\theta)' \cdot I_{\theta}^{-1} \cdot \psi(r; \theta) \]
## ES IF Formulas for Normal and t-Distributions

### Normal Distribution

<table>
<thead>
<tr>
<th>Normal Distribution</th>
<th>t-Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (r - \mu) - \frac{\phi(z_\gamma)}{\gamma} \cdot \frac{(r - \mu)^2 - \sigma^2}{2\sigma} )</td>
<td>( \nu ) known</td>
</tr>
<tr>
<td>( \frac{1}{B} \cdot \psi_\mu(r) - \frac{g_{\gamma,v}}{\gamma C} \cdot \psi_s(r) )</td>
<td>( \nu ) unknown</td>
</tr>
</tbody>
</table>

### T-Distribution

\[
\psi_\mu(r) = \frac{(\nu + 1)(r - \mu)}{v S^2 + (r - \mu)^2}
\]

\[
\psi_s(r, \mu, s) = -\frac{1}{s} + \frac{(\nu + 1)(r - \mu)^2}{v S^3 + s(r - \mu)^2}
\]
Information Matrix for t-Distribution

\[ I_{v,\mu,s} = -E \begin{pmatrix} \frac{\partial^2 l}{\partial v^2} & \frac{\partial^2 l}{\partial v \partial \mu} & \frac{\partial^2 l}{\partial v \partial s} \\ \frac{\partial^2 l}{\partial \mu \partial v} & \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial s} \\ \frac{\partial^2 l}{\partial s \partial v} & \frac{\partial^2 l}{\partial s \partial \mu} & \frac{\partial^2 l}{\partial s^2} \end{pmatrix} \]

\[ = \begin{pmatrix} \frac{1}{4} \left[ \Omega' \left( \frac{v}{2} \right) - \Omega' \left( \frac{v+1}{2} \right) \right] & -\frac{1}{v} \left[ \frac{1}{v+1} - \frac{1}{2(v+3)} \right] & 0 & \frac{1}{s} \left[ \frac{1}{v+3} - \frac{1}{v+1} \right] \\ 0 & 0 & \frac{1}{s^2} \left[ 1 - \frac{2}{v+3} \right] & 0 \\ \frac{1}{s} \left[ \frac{1}{v+3} - \frac{1}{v+1} \right] & \frac{1}{s^2} \left[ 1 - \frac{2}{v+3} \right] & 0 & \frac{2}{s^2} \frac{v}{v+3} \end{pmatrix} \]
Score Vector for t-Distribution

\[
\begin{bmatrix}
\psi_v \\
\psi_\mu \\
\psi_s
\end{bmatrix} = \begin{bmatrix}
\frac{\partial l}{\partial v} \\
\frac{\partial l}{\partial \mu} \\
\frac{\partial l}{\partial s}
\end{bmatrix} = \\
\begin{bmatrix}
\frac{1}{2} \left( \Omega \left( \frac{v+1}{2} \right) - \Omega \left( \frac{v}{2} \right) \right) - \frac{1}{2v} - \frac{1}{2} \log \left( 1 + \frac{(r-\mu)^2}{vs^2} \right) + \frac{(v+1)(r-\mu)^2}{2v^2s^2 + 2v(r-\mu)^2}
\end{bmatrix}
\]

\[
- \frac{1}{s} + \frac{(v+1)(r-\mu)^2}{vs^3 + s(r-\mu)^2}
\]
IF’s of **Normal Distribution** VaR and ES MLE’s

These IF are **quadratically unbounded**. Furthermore:

- **Positive returns contribute to risk!**
- **A serious short-coming!**
IF’s of $t$-Distribution Risk MLE’s for Known d.o.f. = 5

These influence function are all bounded. But still:

Positive returns contribute to risk! A serious short-coming!
IF’s of t-Distribution Risk MLE’s for Unknown d.o.f. = 5

These influence function are unbounded, but only logarithmically.

Positive returns contribute to risk! A serious short-coming!
IF’s of ES Parametric Estimators (MLE’s)

2.5% tail probability (Basel III recommendation)

$t$-distribution d.o.f. = 5

$\sim \log |r|$ for large $|r|$

quadratic

bounded
4. ES ESTIMATOR ASYMPTOTIC VARIANCE

Influence function based asymptotic variance:

\[ V_\hat{\theta}(\theta) = \int IF_\hat{\theta}^2(r; \theta, F)dF(r) \]

\[ S.E.(\hat{\theta}_n) = \frac{V_\hat{\theta}^{1/2}(\hat{\theta})}{n} \]

Applicable to both non-parametric and parametric MLE estimators.

Complicated formulas are available in paper.
Non-Parametric ES S.E.’s for t-Distributions

![Diagram showing the relationship between tail probability and standard error for different degrees of freedom (d.o.f.) values and an infinite (normal) distribution.]
Parametric ES S.E.’s for t-Distributions

![Graph showing parametric ES S.E. for different degrees of freedom (d.o.f.) for t-distributions. The graph plots tail probability against the standard error (S.E.), with different lines representing d.o.f.=3, d.o.f.=5, d.o.f.=7, and d.o.f.=inf (normal).]
Non-Parametric vs Parametric ES S.E.’s

t-distribution with d.o.f. = 3
Ratio of S.E. of ES Parametric Estimators to S.E. of ES Non-Parametric Estimators

<table>
<thead>
<tr>
<th>Tail Probability</th>
<th>0.01</th>
<th>0.025</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parametric Known d.o.f. v.s. Non-parametric</td>
<td>15%</td>
<td>23%</td>
<td>31%</td>
</tr>
<tr>
<td>Parametric Unknown d.o.f. v.s. Non-parametric</td>
<td>60%</td>
<td>73%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Main Expected Shortfall Messages

- Use the influence function to analyze risk and performance estimators
  - Understand data behavior of the estimators
  - Get asymptotic standard errors
  - Can apply to other risk and performance measures

- Parametric estimators are more accurate than non-parametric estimators, but:
  - Defect is that positive returns contribute to downside risk
  - Variability price of estimating t-distribution d.o.f. is high
5. A WAY TO FIX THE PROBLEM

- The problem with parametric ES that positive returns contribute to risk is due to the symmetric nature of the standard deviation and scale estimators.
- A potential solution: Use a semi-scale estimator.
- Initial investigation for normally distributed returns is to use a semi-standard deviation estimator instead of the standard deviation (volatility) estimator.

\[
\sigma(F) = \left( \int_{-\infty}^{\mu} (\mu(F) - r)^2 dF(r) \right)^{1/2} \rightarrow SSD = \left[ \frac{1}{n} \sum_{r_i \leq \bar{r}} (r_i - \bar{r})^2 \right]^{1/2}
\]
$CVaR = ES$
CVaR = ES

Standard Error of CVaR under Normal Distribution

\[ \text{CVaR} = \text{ES} \]
6. INEFFECTIVENESS OF MODIFIED ES


- Modified VaR (mVaR) was proposed by Zangari (1996)* to improve normal VaR by adding skewness and kurtosis corrections

\[ mVaR(\gamma) = \mu + \sigma \cdot g_\gamma \]

where

\[ g_\gamma = z_\gamma + \frac{1}{6}[z_\gamma^2 - 1] \cdot S - \frac{1}{36}[2z_\gamma^3 - 5z_\gamma] \cdot S^2 + \frac{1}{24}[z_\gamma^3 - 3z_\gamma] \cdot K \]

- Reduces to usual normal VaR formula under normality

- Widely publicized and used in a commercial portfolio product called AlternativeSoft, but not a very good idea ....

Modified Expected Shortfall (mES)

- Modified ES (mES) was proposed and studied by Boudt et al. (2008)* as a way to improve normal distribution ES when returns are non-normal.

\[
mETL(\gamma) = \mu - \frac{\sigma}{\gamma} \cdot \phi(g_{\gamma}) \cdot \left[ 1 + c_1 \cdot SK + c_2 \cdot SK^2 + c_3 \cdot K \right]
\]

where \( c_1, c_2, c_3 \) are polynomials in \( g_{\gamma} \).

- Reduces to usual normal mES formula under normality

- This turns out to perform even worse than mVaR

Asymptotic Inefficiency of mES

- The parametric ES estimators for normal and t-distributions are maximum-likelihood estimators (MLE’s), hence they attain the information lower bound on asymptotic variance.

- Martin and Rohit (2015)* develop the large sample variance of mES, and calculate the asymptotic efficiency of mES estimator, defined as the ratio of the asymptotic standard deviation of the ES MLE to that of the mES estimator for normal and t-distributions.

- The results are on the next slide.

---

Asymptotic Inefficiency of mES

mES Efficiencies for Normal and t Distribution

Tail Probability: 1%, 2.5%, 5%
OPEN QUESTIONS & FUTURE RESEARCH

- Will robust semi-scale bound the undesirable influence of positive returns on parametric risk estimators?

- Need results for skewed t-distributions

- Need correct asymptotic variances and standard errors in the presence of serial correlation
  - Recall Lo (2002) results for Sharpe ratio
  - Incipient work on this with Xin Chen

- How do you back-test ES estimators?