Portfolio Optimization with Tail Risk Measures and Non-Normal Returns

Minfeng Zhu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2010

Program Authorized to Offer Degree:

Department of Statistics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Minfeng Zhu

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

_____

R. Douglas Martin

Reading Committee:

_____

R. Douglas Martin

_____

Andrew Clark

_____

Eric W Zivot

Date: _____

Signature _____

Date _____

University of Washington

Abstract

Portfolio Optimization with Tail Risk Measures and Non-Normal Returns

Minfeng Zhu

Chair of the Supervisory Committee:
Professor R. Douglas Martin
Department of Statistics

The traditional Markowitz mean-variance portfolio optimization theory uses volatility as the sole measure of risk. However, volatility is flawed both intuitively and theoretically: being symmetric it does not differentiate between gains and losses; it does not satisfy an expected utility maximization rationale except under unrealistic assumptions and is not a coherent risk measure. The past decade has seen considerable research on better risk measures, with the two tail risk measures Value-at-Risk (VaR) and Expected Tail Loss (ETL) being the main contenders, as well as research on modeling skewness and fat-tails that are prevalent in financial return distributions. There are two main approaches to the latter problem: (a) constructing modified VaR (MVaR) and modified ETL (METL) using Cornish-Fisher asymptotic expansions to provide non-parametric skewness and kurtosis corrections, and (b) fitting a skewed and fat-tailed multivariate parametric distribution to portfolio returns and optimizing the portfolio using ETL based on Monte Carlo simulations from the fitted distribution. It is an open question how MVaR and METL compare with one another and with empirical VaR and ETL, and also how much improvement can be obtained in fitting parametric distributions. In this dissertation, we first show that MVaR and METL are very sensitive to outliers, sometimes rendering complete failure of a portfolio. Then we propose new robust skewness and kurtosis estimates, study their statistical behavior and that of the resulting robust MVaR and METL through the use of influence functions, and show through extensive empirical studies that robust MVaR and METL can effectively curb the failure of the original estimates. We use the same experimental approach to show that the

simple empirical ETL optimization yields portfolio performance essentially equivalent to that of the much more complex method of fitting multivariate fat-tailed skewed distributions. Finally we address the following important problem: VaR and ETL based portfolio optimization do not have expected utility maximization rationales. Thus we establish a method of designing coherent spectral risk measures based on non-satiated risk-averse utility functions. We show that the resulting risk measures satisfy second order stochastic dominance and their empirical portfolio performances are slightly improved over ETL.

Table of Contents

List of Figures

List of Tables

Acknowledgements

The author wishes to express sincere appreciation to the Department of Statistics and the Boeing Capital Corp. for their support and especially to Dr. Murli Thatra, Dr. Dey Nishit Mr. Wayne Hill for their continuous inspiration and Professor R. Douglas Martin for his guidance. This thesis would never have been completed without the encouragement and devotion of my parents and friends.

Dedication

To Mom and Dad

Chapter 1
Introduction

In making portfolio investment decisions, there are two major steps in the process. The first is asset allocation which consists of deciding which asset classes such as equities, bonds, exchange traded funds (ETF's), hedge funds, mutual funds, pension funds, commodities and other alternative assets should be in the portfolio, e.g., 30% stocks, 30% bonds, 10% mutual funds, 5% exchange traded funds (ETF's), 15% hedge funds and 10% cash. There are many approaches to this first problem but most employ a substantial amount of economic and fundamental analysis. The second part of the process is to allocate money to each asset in each asset class. The fundamental analysis approach usually requires a full examination of the financial statements and picks the mispriced instruments by combining information from every financial aspect of the asset. It does not give out exact solutions as to how much capital should be invested in each candidate asset. The more recently developed quantitative analysis approach uses portfolio optimization techniques to exploit the relationships between different assets. These methods give exact weights (percentage of total investment) and aim to match the performance with the risk in the portfolio. Skilled portfolio managers are often good at the first task, i.e. picking a good asset allocation mix and selecting the right assets. However, they are not often so good at allocating capital at the individual asset level. A poorly balanced portfolio can substantially underperform even with the good assets. Therefore finding an appropriate portfolio optimization method is vital to the success of portfolio management.

In the search of finding an optimal solution to the portfolio optimization problem, a lot of theories have been developed in the past 60 years. In this chapter, we will first survey the major portfolio optimization theories and examine their connections. Next we will review the more recent developments with regard to the construction and choice of risk measures, especially downside and tail risk measures, as well as some standards developed in the past decade. Such standards not only meet the conventional wisdom of how people think of risks, but also connect different portfolio optimization methods. Following that, we will discuss the estimation techniques and the practical method of applying these risk measure in the portfolio optimization, in particular when the underlying distribution is skewed and fat-tailed. In the last part of this Chapter we will discuss the organization of the remainder of this dissertation.

1.1 Portfolio Optimization Theories

1.1.1 Mean-Risk Optimization

Markowitz's (1952) seminal work of Mean-Variance portfolio optimization theory is the cornerstone of the *Modern Portfolio Theory* (MPT). It tells us how investors can choose portfolios using an optimal combination of portfolio expected return and volatility (i.e., standard deviation of returns) risk. His optimization problem is as follows,

$$\begin{aligned} \max \quad & \mu_p = \mathbf{w}^{\mathrm{T}}\boldsymbol{\mu} \\ \text{s.t.} \quad & \sigma_p^2 = \mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w} < \sigma_0^2 \end{aligned}, \tag{1.1}$$

where $\mathbf{w} = \left(w_1, w_1, \ldots, w_N\right)^T$ is the weight vector for all N assets. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean return vector and the covariance matrix for all assets, respectively. σ_0^2 is the level of risk we have committed to control our portfolio for. Under this setting, Markowitz tried to

maximize the mean return while limiting the total volatility risk. This is possible due to the fact that the expected returns of the assets in the portfolio are additive, but their volatility risks are not.

The above setting didn't put any constraints, such as short-selling, on the investments. Under this circumstance, the analytical solution constitutes a hyperbolic curve (more details below) as in Fig. 1.1. If the portfolio manager has committed to the long-only strategy and fully invest his capital in the portfolio without leaving any cash, the following two linear constraints are often used

$$\mathbf{w}^{\mathrm{T}}\mathbf{1} = 1, \qquad w_i \geq 0 \qquad \forall i \in 1 \ldots N\,, \tag{1.2}$$

or in more general form of linear constraints

$$A^{\mathrm{T}}\mathbf{w}^{\mathrm{T}} \leq b\,. \tag{1.3}$$

Markowitz's work is a diversification plan. However, diversification is not his ultimate goal. His optimization method also achieves efficiency, i.e. for each possible portfolio with the same mean return, the optimized portfolio has the minimum risk (the dual problem of the above optimization problem) and for each portfolio with the same amount of the risk, the optimized portfolio has the maximum mean return.

If we represent each portfolio by a pair of portfolio mean return and volatility risk, we can plot the entire portfolio universe on a two-dimensional mean vs. risk diagram as below.

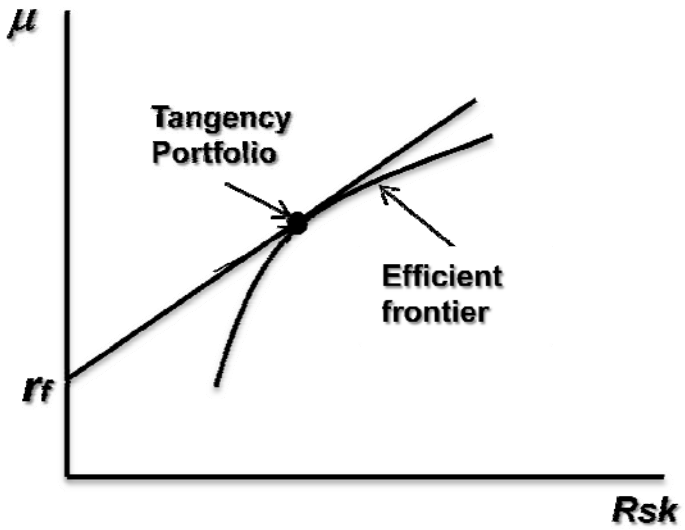

Figure 1.1 Diagram of Mean vs. Risk

The upper boundary (solid curve) of the universe represents the so-called “efficient frontier”. Each optimized portfolio on the frontier is well diversified and efficient in the sense discussed above. If we further include a risk free asset in the universe, the efficient frontier evolves into a straight line that starts from the risk free asset and passes a tangent point (tangency portfolio) on the efficient frontier. This is because by borrowing or lending a risk free asset (cash), we can scale our portfolios outside of the original universe. The portfolios on this line (dashed in the Fig 1.1) are linear combinations of the tangency portfolio and the risk free asset.

The risk measure (variance) in Markowitz’s framework measures the uncertainty of achieving the expected return. Naturally the larger this uncertainty is, the more risky we feel. However, it is counter-intuitive why the upper side of the return distribution, despite of its randomness, constitutes risk. Achieving more than what we expected is not a crime. In fact, Markowitz (1952b) proposed semi-variance as a more appropriate risk measure in the mean vs. risk paradigm. The semi-variance is formally defined as

$$\sigma_-^2 = \int_{-\infty}^{t} (t-r)^2 \, dF(r), \tag{1.4}$$

where t is the target return. The semi-variance calculates the variance for the returns below the target return. The target return can be a fixed number or distribution dependent, e.g. $t = \mu$. It is only because of computational complexity, Markowitz settled for variance as risk in his famous work. Recent literatures have developed many more risk measures of this sort that only counts the down-side as risks. They are collectively called the down-side risk measures or tail risk measures.

Swapping each of these risk measures with variance in the original optimization problem, we are still able to get a "mean vs. risk" efficient portfolio and get the efficient frontier as in Fig. 1.1. The only caveat is that we need a convex risk measure, i.e.

$$Risk(\lambda x + (1-\lambda) y) \le \lambda Risk(x) + (1-\lambda) Risk(y) \qquad \lambda > 0, \tag{1.5}$$

in order to obtain a smooth efficient frontier. If the risk measure is not convex, we may not have a diversification plan, i.e. combining two assets may lead to elevated risk. We will discuss more details on this topic when we review different risk measures.

1.1.2 Expected Utility Maximization

Besides Markowitz's way of optimizing portfolios in the mean vs. risk world, there are other approaches. Economists traditionally use utility functions $U(W)$ as a measure of relative satisfaction given a certain level of wealth. The utility function is always increasing in W , the wealth level, which shows investors perpetual preference for "more". However, such preference decreases marginally as the wealth level increases, i.e.

$$\begin{matrix} U'(W) \geq 0 \\ U''(W) \leq 0 \end{matrix}. \tag{1.6}$$

This defines one special class of utility functions: the risk-averse utility functions. Graphically they can all be represented by the following diagram of a concave function.

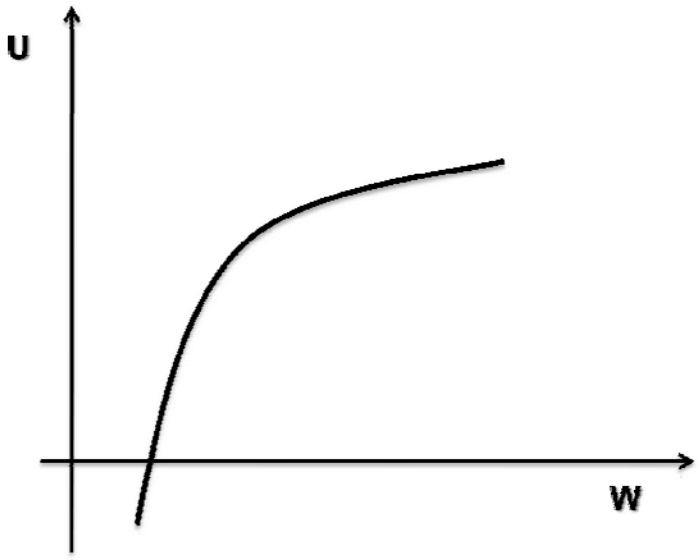

Figure 1.2 Diagram of a Risk-averse Utility Function

With the random nature of future asset returns, the wealth level of tomorrow is random, too. The best an investor can hope for is to maximize his expected utility, i.e. to maximize $E\left[U(w)\right]$. Under the risk-averse assumptions, this goal is different than just directly maximize the expected wealth, because

$$E\left[U(w)\right] \leq U\left(E\left[w\right]\right) \tag{1.7}$$

by Jensen's inequality. This implies it is possible to pursue a better expected utility even when expected return remains the same.

The goal of expected utility maximization is relatively simple, however, it is yet to be agreed among economists which utility function best fits the reality. There have been numerous proposals of utility functions, including but not limited to negative exponential, logarithm, power utility functions and etc. There are also proposals for non-risk-averse utility functions, such as the S-shaped utility functions. It argues that investors can be risk-pursuing when their wealth level is low, such as in gambling, and become risk averse when their wealth level is higher. Besides the difficulty of agreeing on one utility function, it is also a hard problem to calibrate a utility function. We will have more details on the category of utility functions when we formally develop risk measures based on utility functions in Chapter 6.

Expected utility maximization is appealing in the sense that it does not depend on any particular risk or reward measure. It is purely based on the wealth level or asset returns. However, it does not offer a formulated recipe as mean-risk optimization method does. It would be ideal to find the cross-section of both worlds and bridge the gap. Unfortunately, only under certain restrictive special cases, e.g. if the asset returns follow normal distributions and the investor chooses a negative exponential or quadratic utility, would these two approaches generate identical solutions. In more general cases, they do not agree with each other.

1.1.3 Second-Order Stochastic Dominance

By comparison of both mean-risk optimization paradigm and the expected utility theories, we may realize that both represent a method of ordering portfolios by preferences. For the mean-risk paradigm, it is the combination of mean vs. risk pairs that imparts the preference of one portfolio over the other. For the expected utility theory, it is the utility function and the underlying portfolio return distribution that creates such an order. As we

mentioned, generally, these ordering do not agree with each other. The problem comes from the definition of risk measures. It is only possible through the proper definition of a good risk measure that these two paradigms will generate the same ordering.

The stochastic dominance theory, especially the second-order stochastic dominance (SSD) theory, developed by Hanoch and Levy (1969) and Rothschild and Stiglitz (1970) provides some insight to this problem.

The formal definition of first order stochastic dominance (FSD) is as follows:

$$\int_{-\infty}^{t} dF_A(x) \le \int_{-\infty}^{t} dF_B(x) \qquad \forall t \in R, \tag{1.8}$$

if A dominates B or $A \succ_{FSD} B$. It states that the probability of A being lower than any given target is always no larger than the probability of B being lower than the same target. Clearly if this happens, we would prefer portfolio A to B under any circumstance. It simply has less chance to lose. However, portfolio choice is seldom this easy. Most portfolios will not be able to be differentiated by the FSD, i.e. neither $A \succ_{FSD} B$ nor $B \succ_{FSD} A$.

In light of this situation, we can look into the second-order stochastic dominance (SSD), which is

$$\int_{-\infty}^{t} F_A(x)\,dx \le \int_{-\infty}^{t} F_B(x)\,dx \qquad \forall t \in R \tag{1.9}$$

Portfolio A is allowed to have a larger chance to lose for some levels, but overall the cumulative winning odds are still favorable on A for any level. It is worthy to note that SSD is a less stringent condition than FSD.

It was proven (Levy, 1998) that portfolio A dominates portfolio B by SSD if and only if $E_A\left[h(x)\right] \geq E_B\left[h(x)\right]$ for all increasing and concave function h. This is a very important theorem that basically encompasses every risk-averse utility function. It states that if $A \succ_{SSD} B$ SSD, portfolio A will be preferred to portfolio B by all risk-averse expected utility maximizing investors. With the help of this theorem, we no longer need to check any particular utility function. Instead we only need to compare the preference under SSD.

If we could apply SSD directly in asset allocation, it would be a home-run for all risk-averse investors. However, SSD is still only a partial ranking between two competing portfolios, which means if A does not dominate B, it doesn't necessarily imply B dominates A. Some pairs just can't be ranked. Recall that dominance requires the inequality be true for all target levels. It is possible to rank portfolios against a particular benchmark portfolio by dominance, but this does not guarantee it is the "efficient" (not dominated by any other) portfolio in the whole universe. We do not yet have an available recipe to get the global SSD efficient portfolio.

The failure to obtain a SSD efficient portfolio directly wouldn't stop us from applying this rule in the mean-risk paradigm. After all, it satisfies all risk-averse expected utility maximizing investors and mean-risk paradigm has a complete recipe. The goal is to find a reward-risk pair that is consistent with the SSD criteria. By consistency we demand

that every pair ordering in the mean-risk paradigm be preserved under the SSD criteria, i.e.

$$A \succ_{SSD} B \Rightarrow A \succ_{M-R} B. \tag{1.10}$$

It is difficult to demand the opposite, because the mean-risk pair usually does not give information for more than a fixed number of moments of the underlying distribution, whereas the SSD demands information on the whole distribution. However, under this consistency, we can easily see that if one portfolio is chosen by the mean-risk pair dominance, it would certainly be chosen under the SSD dominance. This provides an alternative way of finding SSD efficient portfolios but limiting the search under the mean-risk paradigm.

Based on this view, Giorgi (2005) had a thorough discussion of finding SSD consistent mean-risk pairs. It was found that mean returns are consistent with SSD under any distribution. After all, if A dominates B by SSD, it is certain $E[A] \geq E[B]$. However, for risk measures, it is much trickier, unless we would use mean as a risk measure as well.

Hanoch and Levy (1969) showed variance as a risk measure is not consistent with SSD for an arbitrary distribution except the normal distribution. Therefore the mean-variance optimal result is not desirable to risk-averse expected utility maximizing investors. It wasshown by Copeland and Weston (1988) that it is possible to find a portfolio with larger variance that is not optimal under the mean-variance trade-off, but favored by a risk-averse expected utility maximizing investor.

Fishburn (1977) proposed the lower partial moment (LPM) as a risk measure in this famous $\alpha - t$ model, which is defined as

$$\rho_t^\alpha = \int_{-\infty}^{t} (t-r)^\alpha dF(r) \qquad \alpha \geq 0. \tag{1.11}$$

His $\alpha - t$ model (see also Ogryczak and Ruszczynski, 1999), where risk defined as above is the α power of semi-deviation below a target return t, satisfies the SSD condition when $\alpha \geq 1$. Giorgi (2005) discussed how to construct SSD consistent risk measures in general and gave a recipe for such construction. We will establish a new way to design coherent spectral risk measures based on any non-satiated risk-averse utility function in Chapter 6.

1.2 Risk Measures

Besides the theoretical arguments above, there are other factors to consider when picking risk measures. Variance for example is not a qualified risk measure under SSD criteria. In addition, more obvious and practical criticism for variance comes from the real world. The economists (Mandelbrot, 1963) and other financial practitioners have long recognized that stock market returns have non-normal behaviors, namely asymmetry and fat-tails. They do not follow the smooth and symmetric normal distribution in general, especially in the world of small-cap stocks or hedge fund returns and for shorter sampling periods. Variance is a dispersion measure that fails to discriminate between the upside and downside. Under normality any down-side measure will reduce to a function of the first two moments. Therefore using variance directly is not a problem. Under skewed and fat-tailed distributions, variance could overestimate risks for positive skewed distributions and underestimate for negative skewed distributions. Failure to capture non-normality would result in loss in risk/reward efficiency in general.

As more attention has been drawn to focus on the downside of the return distribution and the necessity to address the non-normality issues, various new risk measures have been developed. We now discuss in details of these risk measures and relevant theories in the following sections.

1.2.1 Value-at-Risk (VaR) and Modified VaR

Value-at-Risk (VaR) as a popular measure among practitioners is the negative of a certain quantile on left tail of the return distribution. The importance of VaR is well established for the purposes of risk management in the banking industry in the 1990s. It is a goal post for risk managers to prepare enough equity to cover for losses up to a certain confidence level. It is not influenced by any prospect of profit on the upper side. We give the formal definition of VaR as below

$$VaR_{1-\alpha}(r) = -\inf_{r}\left\{P(x \le r) \ge \alpha\right\}. \tag{1.12}$$

For example, if among 100 scenarios, the 5th worst loss is \$100, VaR is 100 at the 95% confidence level. As a tail risk measure, VaR gives information about what's to expect at the pre-specified confidence level. It is simple to estimate and easy to back test. It gives portfolio managers and executives a feeling of the magnitude of the potential losses. It can be implemented across different asset classes.

However, since the invention of VaR, there have been plenty of critics of this risk measure despite its popularity in industry.

1. In reality, it is hard to stick to one such probability that triumphs all situations. Actually loss beyond the point that people have been focusing on is usually what causes big problems. Binding to one particular target may defeat the purpose of

a risk measure even when the reality is only slightly worse than the pre-specified confidence level.

2. It is not a coherent measure (more details below). As such, risks measured under VaR are not sub-additive or convex. Combining two assets may even increase risks under VaR, which is contrary to the conventional wisdom of diversification.

3. VaR is a point estimate on the tail, which implies it demands a lot more data to get an accurate estimate than variance.

4. Since VaR is not a convex function of portfolio weights, it is hard to implement its minimization. It can have many local optima that trap the optimization procedure.

Despite all these critics, it is still important to note that VaR gives people important information. It is a downside risk measure and does not depend on normality. Although there's no simple method to make VaR coherent, it is possible to estimate VaR using the entire distribution instead of a simple quantile. The idea is to first build on quantiles of the normal distribution and compensate slight departures from normality by making adjustments using higher moments. There has also been a recent proposal (Jaschke, 2001) to use skewness and kurtosis in a Cornish-Fisher expansion (1938 and 1960) to improve the approximation of VaR for use in portfolio optimization. The general format of the CF expansion is

$$x = \mu + \sigma\left(z + \frac{1}{6}[z^2 - 1]S + \frac{1}{24}[z^3 - 3z]K - \frac{1}{36}[2z^3 - 5z]S^2 \right) = \mu + \sigma \cdot \delta_z \qquad (1.13)$$

It represents the quantile of an arbitrary distribution in terms of the quantiles on the normal distribution and makes adjustments with higher moments. One advantage of this format is that it is a smoother function. It also gives us a convenient tool so that we can

easily estimate VaR without having to specify the exact distribution of historical returns and sufficiently characterize its non-normality. From a practical point of view the use of the first four moments is the most workable. Since higher moments are quite vulnerable to outliers, appropriate robust methods should prove to be helpful. Jaschke (2001) discussed the behavior of CF expansion under specific distribution families. Kim and White (2004) documented the use of several robust measures for skewness and kurtosis on the S&P 500 index. Martellini, Vaissié and Ziemann (2005) tested the idea of modifying VaR and using structured comoments matrix when applying Cornish-Fisher (CF) expansion. Martin and Zhu (2008) further explored the idea of using tail trimmed skewness and kurtosis in CF-modified VaR. We will further discuss the robust aspects of modified VaR in Chapter 3.

1.2.2 Expected-Tail-Loss (ETL) and Coherent Risk Measures

Although CF-modified VaR is appealing by introducing non-normality into the equation, it still suffers from the same culprits that VaR has, namely non-convexity. In this sense, Expected tail loss (ETL) (a.k.a Conditional VaR (CVaR)) is a more appealing risk measure. It is the average of the losses beyond VaR. We can formally define

$$ETL_{1-\alpha}(r) = -E(r \mid r \le VaR_{1-\alpha}). \tag{1.14}$$

It is well known that ETL is a convex risk measure by the effort of Rockafellar and Uryasev (2002) and Pflug (2000). It is a convex measure and will always encourage diversification. Its minimization can be nicely mapped into a convex optimization problem solved via linear programming. Therefore ETL is a much better behaved risk measure than VaR.

Artzner's (1999) introduced the concept of coherent risk measures and established a set of axioms with which a good risk measure should satisfy. A coherent risk measure has the following properties: monotonicity, sub-additivity, positive homogeneity and translation invariance. We can formally define a coherent risk measure by the following definition.

Consider a set V of real-valued random variables. A function $\rho : V \to R$ is called a coherent risk if it is

1. Monotonous: $X, Y \in V, Y \geq X \Rightarrow \rho(Y) \leq \rho(X)$,
2. Subadditive: $X, Y, X + Y \in V \Rightarrow \rho(X + Y) \leq \rho(X) + \rho(Y)$,
3. Positively homogeneous: $X \in V, h > 0, hX \in V \Rightarrow \rho(hX) = h\rho(X)$,
4. Translation invariant: $X \in V, a \in R \Rightarrow \rho(X + a) = \rho(X) - a$.

Monotonicity ensures that an investment that pays off better than the other in a stochastic sense has smaller risks. Sub-additivity makes sure diversification is always preferable. Positive homogeneity states that risk is scalable as investments. Changing the size of the investment will not benefit the investors in terms of the characteristics of their risks. Translation invariance shows adding cash or risk free asset to the portfolio reduces the leverage hence the risk. Therefore a coherent risk measure is preferred in various ways over non-coherent risk measures. Based on these standards, Acerbi (2001) further solidified why ETL is superior to VaR and variance as a risk measure, because it is the only candidate among variance, VaR and ETL that is coherent. It is also proved that ETL is a measure that agrees with SSD. Therefore it is preferred by all risk-averse expected utility maximizing investors, too.

In parallel to CF-modified VaR, Boudt and etl.(2008) applied CF expansion to ETL and created the CF-modified ETL measure as below (see more details on the parameters in Chapter 4)

$$\begin{aligned}\mathrm{mES}(1-\alpha) &= -\mu_p + \frac{1}{\alpha}\sigma_p \cdot \phi(g_\alpha) \\ &\quad + \frac{1}{\alpha}\sigma_p \cdot \left(\begin{array}{c} \frac{1}{24}\left[I^4 - 6I^2 + 3\phi(g_\alpha) \right] \cdot K_p + \frac{1}{6}\left[I^3 - 3I^1 \right] \cdot S_p \\ + \frac{1}{72}\left[I^6 - 15I^4 + 45I^2 - 15\phi(g_\alpha) \right] \cdot S_p^2 \end{array} \right), \\ &\triangleq -\mu_p - \sigma_p \cdot E_G \end{aligned} \tag{1.15}$$

It is another attempt to use the whole distribution instead of the few points on the tail to estimate the risk measure. The CF-modified ETL is between a fully parametrically specified ETL and a non-parametric ETL. It doesn't require specification of the return distribution, but nicely includes skewness and kurtosis in its form. However, they pointed out that such expansion will become unstable when the threshold is too far in the tail, hence necessary robust methods are needed.

1.2.3 Lower Partial Moment (LPM)

While VaR and ETL are both downside measures, they are also both of first order. One criticism of ETL is that it doesn't give more weight to more severe losses. One way of doing so is to increase the power on the losses. Formerly, Fishburn (1977)'s lower partial moment as defined in (1.11) is one of this kind. Markowitz's semi-variance is one special case of LPM.

LPM is a SSD consistent measure when the power is greater than or equal to 1. However, it is not a coherent risk measure. There's discussion in Giorgi (2005) that states most coherent risk measures are not SSD consistent, except for some special cases, such as ETL.

1.2.4 Spectral Risk Measures (SRM) and SMCR

Acerbi (2002, 2007) proposed Spectral risk measures (SRM) as a class of coherent risk measures. It is based on convex combinations of ETL. As shown by Acerbi, any convex combinations of coherent risk measures are still coherent. Acerbi (2002) proved the following rules that SRMs have to follow to stay coherent.

Let $M_\phi(X)$ be defined by

$$M_\phi(X) = -\int_0^1 F_X^{-1}(p)\phi(p)\,dp \tag{1.16}$$

where $\phi \in L^1([0,1])$ and satisfies the following conditions

1. ϕ is positive,
2. ϕ is decreasing,
3. $\|\phi\| = \int_0^1 |\phi(p)|\,dp = 1$,

then $M_\phi(X)$ is a coherent risk measure.

Acerbi's construction is quantile based and assigns more weight to the loss side. In addition, the three conditions for its weight function ensure all weights are positive and normalized to 1. Following his approach, Cotter and Dowd (2006) applied this construction to various stock market indices and compared SRM with other risk measures like VaR and ETL. Although Acerbi gave the basic formula for SRM, it remains an interesting question to find the appropriate weight function that makes economic sense. We will discuss this topic in Chapter 6.

Besides Acerbi's work, Krokhmal (2007) introduced a new family of higher-moment coherent risk measures (HMCR), which uses the power of tail losses in risk calculation. The basic construction is

$$\mathrm{HMCR}_{p,\alpha} = \min_{\eta \in R} \eta + \frac{1}{\alpha} \left\| \left(X - \eta \right)^{+} \right\|_{p} \tag{1.17}$$

where $p \geq 1, \alpha \in (0,1)$. Here α is the threshold and p designates the power. The p th order norm is formally defined by

$$\left\| X \right\|_{p} = \left(E\left(\left| X \right|^{p} \right) \right)^{1/p} \tag{1.17}$$

ETL is a special case of HMCR, where $p = 1$. When $p = 2$, the above becomes second moment coherent measure (SMCR). It is similar to the construction of the lower partial moment. However, when $p > 1$, η no longer holds the traditional sense of VaR that corresponds to the level $1 - \alpha$. The HMCR measure is a stochastic programming concept. η is only the optimal point that makes the whole measure minimized.

Krokhmal also showed that the HMCR measures under certain conditions are compatible with the second order stochastic dominance. Therefore HMCR constructed by Krokhmal should also be favored by risk-averse expected utility maximizing investors.

1.3 Estimating Risk Measures with Higher Moments or Parametric Distributions

Through the discussion above, we realize tail risk measures are more appropriate risk measures. However, we shall not write variance off as a risk measure just yet. Though undesirable under various circumstances, variance is simple and stable to estimate as

proved in various empirical studies. When we deal with non-normality and tail risk measures, the estimates may become unreliable. There are generally two approaches to the estimation of downside risk measures under non-normal distributions besides the non-parametric estimates by definition. One is through the high moments adjustments as in modified VaR and ETL. The other is to fit skewed and fat-tailed distributions like skewed-t or skewed-stable distributions to the marginal return distribution and then estimate risk measures using simulation from the fitted distribution.

The first approach relies on the reliability of the higher moment estimates. High moments are usually estimated by scenario based approaches, which either directly use the historical data or generate scenarios (bootstrapping) from the historical data. This is often under the influence of outliers, which greatly impact the estimates as the power on the moment estimators goes higher. It can be addressed in two ways. First, if we estimate high moments through the multivariate co-skewness or co-kurtosis matrices, there is an explosion of number of parameters to estimate as the number of assets increases. Martellini and Ziemann (2008) proposed a robust method that uses a single factor model and shrinkage method for estimating these co-moment matrices. Boudt et. al. (2008) proposed a cleaning method that scales back outliers based on multivariate normal assumptions. Second, we can deal with outliers after we have applied weights to each asset to form a portfolio. At this time, the portfolio return distribution is a univariate distribution that various robust methods can be applied. This will be addressed in Chapter 2 through an influence function approach.

The second approach, using parametric distributions, usually involves first fitting a skewed and fat-tailed distribution to the marginal asset return distribution by methods such as maximum likelihood. The fitted distribution is then used to generate samples

with much larger sample size than the original sample. The generated scenarios are used to calculate the risk measures using the original nonparametric definitions. Rachev et al. (2007) used stable distributions to fit the marginal distributions of asset returns and the skewed t-Coupla to "glue" them together. The Cognity® software by *FinAnalytica* is the only commercial product to date that offers such capability. The parametric solutions are based on model assumptions. One advantage is that a probability model will assign probabilities beyond the scope of the original sample and therefore is more versatile. However, they are also estimated by the original sample. Any bias in the original sample will be broadcasted to the simulated samples, let alone the variability across different simulated samples. Robust estimation of highly skewed and fat-tailed distribution is difficult.

1.4 Remainder of Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, we will discuss the robust methods for skewness and kurtosis. We will derive influence functions for skewness and kurtosis as well as their robust versions. We will also discuss the finite sample variances of these estimates. In Chapter 3, we will derive the influence function for modified VaR and discuss its implications. We will carry out empirical studies of portfolio optimization using hedge fund data. We will perform a horse race among the volatility, VaR and modified VaR risk measures on 100 random portfolios of hedge funds. We will discuss results and compare the performance of different risk measures under financial crisis of 2008-2009. We will also discuss impacts of different robust methods. In Chapter 4, we will perform parallel experiments studying modified ETL and repeat the same portfolio optimization using ETL and modified ETL measures. In Chapter 5, we will seek a parametric model for estimating ETL. It will be compared with the historical data

based methods, especially under the big swings of financial crisis. We will see which method holds better under these circumstances. In Chapter 6, we will introduce utility based spectral risk measures. It is an extension of work of Acerbi (2002) and Giorgi (2005). We will derive the basic format of such risk measures under different utility functions. We will also demonstrate that it is consistent with the definition of SRM and SSD. We will explain different economic implications from various perspectives. Finally, in Chapter 7, we will repeat the portfolio optimization experiment using the spectral, SMCR and SSD risk measures.

Chapter 2
Influence Functions and Robust Skewness and Kurtosis

2.1 Skewness and Kurtosis and Robust Versions

Classical skewness and kurtosis are defined as the following:

$$S = \frac{E[x-\mu]^3}{\left(E[x-\mu]^2\right)^{3/2}} \qquad K = \frac{E[x-\mu]^4}{\left(E[x-\mu]^2\right)^2} - 3 \tag{2.1}$$

where $\mu = E[x]$, provided the third and fourth moments, respectively exist. Skewness is a measure of asymmetry of a probability distribution. Positive or negative skewness points to more stretched tails in the right or left side of the distribution, respectively. Kurtosis is a measure of "peakedness" or "fatness" of the tails of a probability distribution. Larger kurtosis points to a fatter tailed distribution. For symmetric distributions, $S = 0$. For normal distributions, $K = 0$. The sample estimates of skewness and kurtosis are

$$\hat{S}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}_n\right)^3}{\hat{\sigma}_n^{\,3}} \qquad \hat{K}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}_n\right)^4}{\hat{\sigma}_n^{\,4}} - 3 \tag{2.2}$$

where $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\mu}_n\right)^2}$. Since sample skewness and kurtosis are based on cubic and quadratic functions of the data these estimates are not at all "robust" in the sense that they can be very highly influenced by one or more outliers.

The influence of outliers on estimators is conveniently shown by computing influence functions (Hampel, Ronchetti, Rousseeuw and Stahel, 1986 and Maronna, Martin and Yohai, 2006). We show in the next section that the influence functions of the classical

skewness and kurtosis estimates are unbounded, reflecting the fact that a single outlier can have arbitrarily large influence on the estimates.

The statistics literature contains several proposals for robust skewness and kurtosis. Hinkley (1975) introduced the quantile based coefficient of skewness

$$S_q = \frac{\left(q(1-\alpha)-q(0.5)\right)-\left(q(0.5)-q(\alpha)\right)}{q(1-\alpha)-q(\alpha)} \tag{2.3}$$

where $q(\cdot)$ is the quantile function

$$q(\alpha) = F_X^{-1}(\alpha) = \inf\left\{x \mid F_X(x) \geq \alpha\right\}, \quad 0 < \alpha < 1 \tag{2.4}$$

This was developed based on the coefficient of skewness defined by Bowley (1920), where α takes 0.25 in (2.3), as a special case of Hinkley's definition. In order to remove the dependence of Hinkley's definition on α, Groeneveld and Meeden (1984) made a new definition by integrating the numerator and denominator with respect to α, respectively, i.e.

$$S_{GM} = \frac{\mu - q(0.5)}{E\left(\left|x - q(0.5)\right|\right)}. \tag{2.5}$$

Note the definition in (2.5) uses $E\left(\left|x-q(0.5)\right|\right)$ as a measure of dispersion. In Pearson coefficient of skewness (see Kendall and Stuart, 1977), this was replace by the standard deviation, i.e.

$$S_{GM} = \frac{\mu - q(0.5)}{\sigma}. \tag{2.6}$$

Kim and White (2005) compared the definitions above along with the classical skewness measure. They used the returns of S&P 500 index as the underlying sample for their experiments. They found if one does not want to have any impact of s ingle outlier when measuring skewness, definition (2.3) is the best candidate. Definitions (2.5) and (2.6) involved μ or σ in their expression and are less robust than (2.3).

Brys et al. (2003, 2004) developed the medcouple concept for measuring skewness. The formal definition is

$$MC_n = \underset{x_i \le m_n \le x_j}{med} \frac{\left(x_j - m_n\right) - \left(m_n - x_i\right)}{x_j - x_i}, \tag{2.7}$$

where med is the median function and m_n is the sample median. It takes every pair of samples with one above the median and one below and then calculates the respective difference similar to (2.3). After all pairs of differences are calculated, the final skewness is defined as the median of these differences.

Besides the idea of using quantiles for robust skewness, trimming the tails is another option. Dolmas (2005) applied trimming to the tails when studying the price index for Personal Consumption Expenditures (PCE). He found the only robust skewness measures can provide evidence of more systematic skewness in the distribution of PCE inflation rate in recent years.

For kurtosis, Moors (1988) defined the following octile based kurtosis

$$K_M = \frac{\left(q\left(\frac{7}{8}\right) - q\left(\frac{5}{8}\right)\right) - \left(q\left(\frac{3}{8}\right) - q\left(\frac{1}{8}\right)\right)}{\left(q\left(\frac{6}{8}\right) - q\left(\frac{2}{8}\right)\right)}. \tag{2.8}$$

Hogg (1972,1974) used the following definition

$$K_H = \frac{U_{0.05} - L_{0.05}}{U_{0.5} - L_{0.5}} \tag{2.9}$$

where $U_\alpha = \frac{1}{\alpha}\int_{1-\alpha}^{1} q(t)\,dt$ and $L_\alpha = \frac{1}{\alpha}\int_{0}^{\alpha} q(t)\,dt$ are the average of the upper or lower α quantiles, respectively.

Crow and Siddiqui (1967) has used the following definition

$$K_{CS} = \frac{q(1-\alpha) - q(\alpha)}{q(0.75) - q(0.25)}. \tag{2.10}$$

where α is taken to be 0.025.

In the study of Kim and White (2005), they found conventional kurtosis has very poorly stability even with a sample size of 5,000 when the underlying is a heavy tailed distribution. Hogg's definition is better but still under the influence of outliers, because of it relies on the average of the tails. Both definitions (2.8) and (2.10) are robust to the impact of outliers. But Moors's definition is only affected when the outlier is inside the region between the first and last octiles. There is very little sensitivity left to the tails. In this sense, Crow and Siddiqui's definition is better.

Seier and Bonett (2003) studied the following two definitions for kurtosis,

$$K_1(b) = E\left[ab^{-|z|}\right] \qquad K_2(b) = E\left[a\left(1-|z|^b\right)\right].$$

where a, b are constants and z is the standardized random variable. Both measures focus on the center part of the distribution to measure the "peakedness" in the distribution.

Brys et al. (2006) extended their concept of medcouple to kurtosis definitions as well. They've shown that medcouple has a high break-down point and bounded influence function. The only drawback is its computation expense is high.

In our study, we would mainly use skewness and kurtosis in our construction of risk measures. Therefore sensitivity to the fatness in the tail and robustness to the outliers are both important to us. Kim and White (2005) showed that even the experiments on the returns of S&P 500 index would demand good robust measures for skewness and kurtosis. One might expect that the influence of outliers could be even more severe in other indexes such as the Russell 2000 small-cap, micro-cap indices (name one), or hedge fund indices such as provided by HFR and HedgeFund.net. We pick the trimming method and one of the quantile based method for robust skewness and kurtosis to study in this chapter. The trimming method would inherit the characteristics for the classical definitions for most part of the underlying distribution. The quantile based method are (2.3) and (2.10) with $\alpha = 0.025$. The other methods are either not sensitive the far tails or less robust because involvement of μ or σ.

Trimming Method

The first is a very simple symmetric trimming method. We trim the sample symmetrically on both tails and calculate the skewness based on the rest of the sample. We first defined the following trimmed sample moments,

$$
\begin{aligned}
n_{trim} &= \sum_{i=1}^{n} I\left(x_i \in \left[q_n(\alpha), q_n(1-\alpha)\right]\right) \\
\hat{\mu}_{trim,n} &= \frac{1}{n_{trim}} \sum_{x_i \in \left[q_n(\alpha), q_n(1-\alpha)\right]} x_i \\
\hat{\mu}^m_{trim,n} &= \frac{1}{n_{trim}} \sum_{x_i \in \left[q_n(\alpha), q_n(1-\alpha)\right]} \left(x_i - \hat{\mu}_{trim,n}\right)^m,
\end{aligned}
\tag{2.11}
$$

where the empirical quantile $q_n(\alpha)$ is defined by interpolating between two closest order statistics

$$q_n(\alpha) = (1-\lambda)\, x_{(j)} + \lambda x_{(j+1)}$$

where $x_{(i)}$ is the i th order statistic of the sample $\{x_1, x_2, ..., x_n\}$. The constant λ is defined by convention as in Hyndman and Fan (1996) [1].

Now we can define the respective finite sample estimates for (2.11) as:

$$
\hat{S}_{trim} = \frac{\hat{\mu}^3_{trim,n}}{\left(\sqrt{\hat{\mu}^2_{trim,n}}\right)^3} \qquad \hat{K}_{trim} = \frac{\hat{\mu}^4_{trim,n}}{\left(\sqrt{\hat{\mu}^2_{trim,n}}\right)^4} - C_{trim} \tag{2.12}
$$

The definition (2.12) will consistently converge to the following when $N \to \infty$

[1] We choose j that satisfies $\frac{j-0.5}{n} \leq \alpha < \frac{j+0.5}{n}$ and set λ equal to the fractional part of $n\alpha - (j - 0.5)$

$$S_{trim} = \frac{E_{\tilde{F}}\left[x-\tilde{\mu}\right]^3}{\left(E_{\tilde{F}}\left[x-\tilde{\mu}\right]^2\right)^{3/2}}, \qquad K_{trim} = \frac{E_{\tilde{F}}\left[x-\tilde{\mu}\right]^4}{\left(E_{\tilde{F}}\left[x-\tilde{\mu}\right]^2\right)^2} - C_{trim}, \tag{2.13}$$

where the measure $\tilde{F}$ is the probability measure after trimming,

$$\tilde{F}_X(x) = \frac{1}{1-2\alpha} F_X(x) \cdot I\left(q(\alpha) < x < q(1-\alpha)\right). \tag{2.14}$$

The $I(\cdot)$ function is the indicator function

$$I(x \in A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{2.15}$$

and we abbreviate $I\left(q(\alpha) < x < q(1-\alpha)\right)$ to $I_{\alpha,1-\alpha}(x)$ for simplicity in the future. The trimmed mean $\tilde{\mu}$ in (2.11) is defined as $\tilde{\mu} = E_{\tilde{F}}\left[x\right]$ and $C_{trim} = \frac{E_{\tilde{\Phi}}\left[x-\tilde{\mu}\right]^4}{\left(E_{\tilde{\Phi}}\left[x-\tilde{\mu}\right]^2\right)^2}$.

The classical definition (2.1) for skewness and kurtosis is only finite when the third and fourth moments exist for the distribution, respectively. Their sample variances are finite only when the sixth and eighth moments exist, respectively. However, the trimmed definition in (2.13) would exist even if the lower order moments do not exist.

Robust Quantile Method

The second method is a quantile based robust method. The definition in (2.3) is extended to kurtosis and takes the following form,

$$S_r = \frac{q(\alpha) + q(1-\alpha) - 2q(0.5)}{q(1-\alpha) - q(\alpha)} \qquad K_r = \frac{q(1-\alpha) - q(\alpha)}{q(0.75) - q(0.25)}. \tag{2.16}$$

The sample based estimates of the robust quantile estimates are given by:

$$\hat{S}_q = \frac{q_n(\alpha) + q_n(1-\alpha) - 2q_n(0.5)}{q_n(1-\alpha) - q_n(\alpha)}$$
$$\hat{K}_q = \frac{q_n(1-\alpha) - q_n(\alpha)}{q_n(0.75) - q_n(0.25)}.$$

The α 's in the above definitions determine the trade-off between control of bias due to outliers and variance of the estimates (larger α yields more bias control but larger variance, and vice versa). Generally speaking we believe that fairly small values of α , such as 2.5% or 5%, will due.

2.2 Influence Functions for Skewness and Kurtosis

We briefly introduce empirical and asymptotic influence functions for the purposes of comparing the classical skewness and kurtosis estimates with two proposed robust estimates of skewness and kurtosis.

In order to introduce the asymptotic (theoretical) influence function we consider the mixture distribution $F_\gamma = (1-\gamma)F + \gamma\delta_r$ where F is the nominal distribution of the data and δ_r is the delta Dirac function at r . Most estimators can be represented asymptotically as a functional $\theta(F)$ of a general probability distribution F . For example the sample mean may be represented asymptotically as the functional

$$\theta(F) = \int x dF(x)$$

for all distributions F for which the mean exists. Similarly the skewness and kurtosis estimates may be represented as functionals for all distributions for which the mean and kurtosis exist. With the functional representation of an estimator in hand, the asymptotic influence function is defined by

$$\mathrm{IF}\left(r;\theta,F\right)=\lim_{\gamma\downarrow 0}\frac{\theta(F_\gamma)-\theta(F)}{\gamma}=\frac{d}{d\gamma}\theta(F_\gamma)\Big|_{\gamma=0} \tag{2.17}$$

and the influence functions for skewness and kurtosis are both unbounded, as we show in the next section.

For finite samples, let $\hat{\theta}_n=\hat{\theta}_n(\mathbf{x})$ be a univariate estimator based on the sample $\mathbf{x}=\left(x_1,x_2,...,x_n\right)$. The empirical influence function is defined by

$$EIF\left(z;\hat{\theta}_n,\mathbf{x}\right)=(n+1)\left(\hat{\theta}_n(z,\mathbf{x})-\hat{\theta}_n(\mathbf{x})\right) \tag{2.18}$$

2.2.1 Classical Skewness and Kurtosis

Before we show the influence functions of the classical skewness and kurtosis, we work out the influence function for centered moments. Define

$$\begin{aligned}\mu&=\int r dF_R\left(r\right)\\ \mu_n&=\int\left(r-\mu\right)^n dF_R\left(r\right).\end{aligned} \tag{2.19}$$

We have

$$\begin{aligned}\mathrm{IF}\left(r;\mu_n,F\right)&=\frac{d}{d\gamma}\mu_n(F_\gamma)\Big|_{\gamma=0}\\ &=\frac{d}{d\gamma}\int\left(x-\mu_\gamma\right)^n dF_\gamma\left(x\right)\Big|_{\gamma=0}\end{aligned} \tag{2.20}$$

Note that we single the mean μ out because it is not a centered moment. The influence function for μ is trivial

$$\begin{aligned} \mathrm{IF}(r;\mu,F) &= \frac{d}{d\gamma}\int x dF_\gamma(x)\Big|_{\gamma=0} \\ &= r-\mu \end{aligned} \tag{2.21}$$

It is certainly unbounded and but only linearly increasing with the outlier.

The first order centered moment is $\mu_1 = 0$ regardless of the actual distribution, hence its influence function is also 0 everywhere.

For higher order central moments, we can write

$$\begin{aligned} \mu_n &= \int \left(x-\mu_\gamma\right)^n dF_\gamma(x) \\ &= \int \sum_{i=0}^{n}\left\{\binom{n}{i} x^i \left(-\mu_\gamma\right)^{n-i}\right\} dF_\gamma(x) \\ &= \sum_{i=0}^{n}\binom{n}{i}\int x^i \left(-\mu_\gamma\right)^{n-i} dF_\gamma(x). \end{aligned} \tag{2.22}$$

The influence function is

$$\begin{aligned} \mathrm{IF}(r;\mu_n,F) &= \frac{d}{d\gamma}\int \left(x-\mu_\gamma\right)^n dF_\gamma(x)\Big|_{\gamma=0} \\ &= -\int n\left(x-\mu\right)^{n-1}\cdot \mathrm{IF}(r;\mu,F)\cdot dF(x) + \int \left(x-\mu\right)^n \cdot d\left[\delta_r(x)-F_X(x)\right] \\ &= -n\left(r-\mu\right)\mu_{n-1} + \left(r-\mu\right)^n - \mu_n \end{aligned} \tag{2.23}$$

We use (2.23) to get the influence function for σ^2

$$\begin{aligned}
&\mathrm{IF}\left(r;\sigma^2,F\right)=\mathrm{IF}\left(r;\mu_2,F\right)\\
&=\left(r-\mu\right)\left\{\left(r-\mu\right)^{2-1}-2\mu_{2-1}\right\}-\mu_2\\
&=\left(r-\mu\right)\left\{\left(r-\mu\right)-\mu_1\right\}-\mu_2\\
&=\left(r-\mu\right)^2-\sigma^2
\end{aligned} \tag{2.24}$$

Now we will work out the influence functions for the third and forth centered moments respectively,

$$\begin{aligned}
&\mathrm{IF}\left(r;\mu_3,F\right)\\
&=\left(r-\mu\right)\left\{\left(r-\mu\right)^{3-1}-3\mu_{3-1}\right\}-\mu_3\\
&=\left(r-\mu\right)\left\{\left(r-\mu\right)^2-3\mu_2\right\}-\mu_3\\
&=\left(r-\mu\right)^3-3\left(r-\mu\right)\sigma^2-\mu_3
\end{aligned} \tag{2.25}$$

and

$$\begin{aligned}
&\mathrm{IF}\left(r;\mu_4,F\right)\\
&=\left(r-\mu\right)\left\{\left(r-\mu\right)^{4-1}-4\mu_{4-1}\right\}-\mu_4\\
&=\left(r-\mu\right)\left\{\left(r-\mu\right)^3-4\mu_3\right\}-\mu_4\\
&=\left(r-\mu\right)^4-4\left(r-\mu\right)\mu_3-\mu_4
\end{aligned} \tag{2.26}$$

In order to get the influence functions for skewness and kurtosis, we also need the following influence function for any power function of the variance, i.e.

$$\begin{aligned}
&\mathrm{IF}\left(r;\left(\sigma^2\right)^{n/2},F\right)=\mathrm{IF}\left(r;\left(\mu_2\right)^{n/2},F\right)\\
&=\frac{n}{2}\left(\mu_2\right)^{n/2-1}\mathrm{IF}\left(r;\mu_2,F\right)\\
&=\frac{n}{2}\sigma^{n-2}\left[\left(r-\mu\right)^2-\sigma^2\right]
\end{aligned} \tag{2.27}$$

Now we are ready to derive the influence function for skewness and kurtosis.

For skewness we have

$$\begin{aligned}
&\mathrm{IF}\left(r;S,F\right)=\mathrm{IF}\left(r;\frac{\mu_3}{\left(\mu_2\right)^{3/2}},F\right)\\
&=\frac{\mathrm{IF}\left(r;\mu_3,F\right)\cdot\left(\mu_2\right)^{3/2}-\mu_3\cdot\mathrm{IF}\left(r;\left(\mu_2\right)^{3/2},F\right)}{\left(\mu_2\right)^3}\\
&=\frac{\left(\left(r-\mu\right)^3-3\left(r-\mu\right)\sigma^2-\mu_3\right)\cdot\sigma^3-\frac{3}{2}\mu_3\cdot\sigma\left[\left(r-\mu\right)^2-\sigma^2\right]}{\sigma^6}\\
&=\frac{\left(r-\mu\right)^3-3\left(r-\mu\right)\sigma^2+\frac{1}{2}\mu_3}{\sigma^3}-\frac{3\mu_3}{2\sigma^5}\left(r-\mu\right)^2
\end{aligned}\tag{2.28}$$

In a simplified case, i.e. for any symmetric distribution where we have $\mu_3=0$, (2.28) reduces to

$$\begin{aligned}
&\mathrm{IF}\left(r;S,F\right)\\
&=\frac{\left(r-\mu\right)^3-3\left(r-\mu\right)\sigma^2}{\sigma^3}\\
&=\frac{\left(r-\mu\right)^3}{\sigma^3}-3\frac{r-\mu}{\sigma}
\end{aligned}\tag{2.29}$$

It is a cubic function. If we further assume the distribution is standardized, i.e. $\mu=0,\sigma^2=1$, we have

$$\mathrm{IF}\left(r;S,F\right)=r^3-3r.\tag{2.30}$$

For kurtosis, we have

$$
\begin{aligned}
\mathrm{IF}\left(r;K,F\right) &= \mathrm{IF}\left(r;\frac{\mu_4}{\left(\mu_2\right)^2}-3,F\right) \\
&= \frac{\mathrm{IF}\left(r;\mu_4,F\right)\cdot\left(\mu_2\right)^2-\mu_4\cdot\mathrm{IF}\left(r;\left(\mu_2\right)^2,F\right)}{\left(\mu_2\right)^4} \\
&= \frac{\left(\left(r-\mu\right)^4-4\left(r-\mu\right)\mu_3-\mu_4\right)\cdot\sigma^4-2\mu_4\cdot\sigma^2\left[\left(r-\mu\right)^2-\sigma^2\right]}{\sigma^8} \\
&= \frac{\left(r-\mu\right)^4-4\left(r-\mu\right)\mu_3+\mu_4}{\sigma^4}-2\frac{\mu_4}{\sigma^6}\left(r-\mu\right)^2
\end{aligned}
\tag{2.31}
$$

It is a quartic function. If we have a symmetric distribution where $\mu_3=0$,

$$
\mathrm{IF}\left(r;K,F\right)=\frac{\left(r-\mu\right)^4}{\sigma^4}-2\frac{\mu_4}{\sigma^6}\left(r-\mu\right)^2+\frac{\mu_4}{\sigma^4}. \tag{2.32}
$$

If we further assume the distribution is standardized with $\mu=0,\sigma^2=1$ as above, we have

$$
\mathrm{IF}\left(r;K,F\right)=r^4-2\mu_4 r^2+\mu_4 \tag{2.33}
$$

For the case of standard normal distribution, we have $\mu_4=3$. Hence

$$
\mathrm{IF}\left(r;K,F\right)=r^4-6r^2+3 \tag{2.34}
$$

We show the plots (Fig. 2.1) of influence functions for skewness and kurtosis under standard normal distribution. The black curve is the theoretical curve derived in (2.25) and (2.34). The red dashed line represents the empirical influence functions. They agree very well.

The empirical influence functions are obtained by first using 10,000 quantiles from 0.01% to 99.99% in steps of .01% to represent the original distribution. A contamination point is then added to the original 10,000 points to form the mixture distribution. The empirical influence function is calculated by (2.18).

It is obvious from the plots that both influence functions are unbounded. Outliers on either tail can make significant impact on skewness and kurtosis. We also note that due to the cubic nature of the influence function for skewness, positive r values do not always have positive influence on the skewness estimate. For $\sqrt{3}$ standard deviations below the mean, the new data addition would result in higher skewness. The opposite is that for $\sqrt{3}$ standard deviations above the mean, the new data addition result in lowered skewness. In the portfolio optimization scenario, if we are to maximize skewness as our sole objective, we may prefer a portfolio with an additional return slightly below the mean to an additional return slightly above the mean holding everything else constant. This is a counter-intuitive fact as a consequence of the structure of the classical measure of skewness. Different distributions may show different type of cubic influence functions (see next few sections), but they will more or the less have this characteristic. This gives us doubt for directly maximizing skewness as a portfolio optimization objective function.

We observe the same for kurtosis. Kurtosis is supposed to measure the tail thickness of the distribution. However, according to the influence function, data points at the center of the distribution can actually increase the kurtosis estimate. They are the so-called “inliers”. This is consistent with what one expects for the stylized facts about fat-tailed distributions being more peaked in the middle as well as having fat tails. We will see this effect in later sections too.

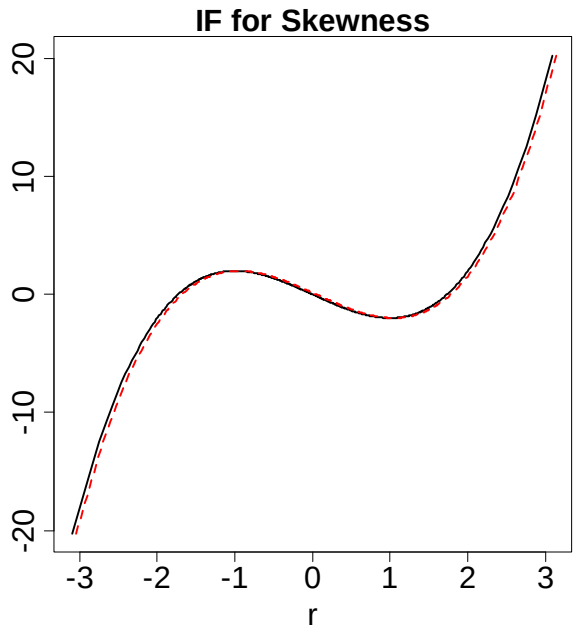

Figure 2.1 Influence Functions for Skewness: Standard Normal Distribution

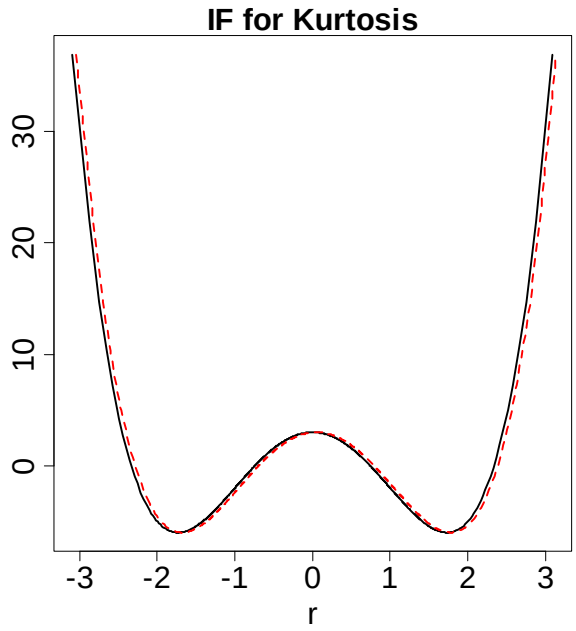

Figure 2.2 Influence Functions for Kurtosis: Standard Normal Distribution

2.2.2 Trimmed Skewness and Kurtosis

We expand the definition in trimmed moments into the following,

$$\begin{aligned}\tilde{\mu}_n &\triangleq \int \left(x-\tilde{\mu}\right)^n d\tilde{F}\left(x\right) \\ &= \frac{1}{1-2\alpha}\int_{q(\alpha)}^{q(1-\alpha)} \left(x-\tilde{\mu}\right)^n dF\left(x\right)\end{aligned} \tag{2.35}$$

and

$$\begin{aligned}\tilde{\mu} &= \int x d\tilde{F}\left(x\right) \\ &= \frac{1}{1-2\alpha}\int_{q(\alpha)}^{q(1-\alpha)} x dF\left(x\right)\end{aligned} \tag{2.36}$$

Note that the trimming sites do not necessarily have to be symmetric. The following derivations do not take advantage of this property. Every upper tail quantity expressed in terms of $1-\alpha$ can be easily replaced by an arbitrary $b > 0.5$. However, without specific knowledge of the underlying distribution, there's no incentive to trim asymmetrically on the tails. We will keep the symmetric trimming in the following representation.

We calculate the influence function for the trimmed skewness based on (2.12)

$$\begin{aligned}&\mathrm{IF}\left(r; S_{trim}, \tilde{F}\right) \\ &= \frac{\mathrm{IF}\left(r; \tilde{\mu}_3, \tilde{F}\right)\cdot \tilde{\mu}_2^{\,3/2} - \frac{3}{2}\tilde{\mu}_3 \cdot \tilde{\mu}_2^{\,1/2}\cdot \mathrm{IF}\left(r; \tilde{\mu}_2, \tilde{F}\right)}{\tilde{\mu}_2^{\,3}}\end{aligned} \tag{2.37}$$

We will need several parts to finish the above equation. First, we derive the influence function for a quantile based on the following identity (see Hampel et. al., 1986),

$$\alpha = \int_{-\infty}^{q_\gamma(\alpha)} dF_\gamma(x). \tag{2.38}$$

Taking derivatives with respect to γ on both sides of (2.38) and evaluate at $\gamma = 0$,

$$
\begin{aligned}
0 &= IF\left(r; q(\alpha), F\right)\cdot f_X\left(q(\alpha)\right) - \int_{-\infty}^{q(\alpha)} dF_X(x) + I\left(r < q(\alpha)\right) \\
IF\left(r; q(\alpha), F\right) f_X\left(q(\alpha)\right) &= \alpha - I\left(r < q(\alpha)\right) \\
IF\left(r; q(\alpha), F\right) &= \frac{\alpha - I\left(r < q(\alpha)\right)}{f_X\left(q(\alpha)\right)}
\end{aligned}
\tag{2.39}
$$

Now we compute the influence function for the first moment of the trimmed distribution based on (2.36)

$$
\begin{aligned}
&\mathrm{IF}\left(r; \tilde{\mu}, F\right) = \frac{d}{d\gamma}\frac{1}{1-2\alpha}\int_{q_\gamma(\alpha)}^{q_\gamma(1-\alpha)} x dF_\gamma(x)\Big|_{\gamma=0} \\
&= \frac{1}{1-2\alpha}\cdot IF\left(r; q(1-\alpha), F\right)\cdot q(1-\alpha)\cdot f\left(q(1-\alpha)\right) \\
&\qquad -\frac{1}{1-2\alpha}\cdot IF\left(r; q(\alpha), F\right)\cdot q(\alpha)\cdot f\left(q(\alpha)\right) \\
&\qquad +\frac{1}{1-2\alpha}\cdot\int_{q(\alpha)}^{q(1-\alpha)} x\cdot d\left[\delta_r(x) - F(x)\right] \\
&= \frac{1}{1-2\alpha}\cdot\left[\left((1-\alpha) - I_{0,1-\alpha}(r)\right)\cdot q(1-\alpha) - \left(\alpha - I_{0,\alpha}(r)\right)\cdot q(\alpha)\right] \\
&\qquad +\frac{1}{1-2\alpha}\cdot r\cdot I_{\alpha,1-\alpha}(r) - \tilde{\mu}
\end{aligned}
\tag{2.40}
$$

For higher order centered moments based on (2.30), we have

$$
\begin{aligned}
&\mathrm{IF}\left(r;\tilde{\mu}_n,\tilde{F}\right)\\
&=\frac{d}{d\gamma}\frac{1}{1-2\alpha}\int_{q_\gamma(\alpha)}^{q_\gamma(1-\alpha)}\left(x-\tilde{\mu}_\gamma\right)^n dF_\gamma\left(x\right)\Bigg|_{\gamma=0}\\
&=\frac{1}{1-2\alpha}IF\left(r;q\left(1-\alpha\right),F\right)\cdot\left(q\left(1-\alpha\right)-\tilde{\mu}\right)^n\cdot f\left(q\left(1-\alpha\right)\right)\\
&\qquad-\frac{1}{1-2\alpha}\cdot IF\left(r;q\left(\alpha\right),F\right)\cdot\left(q\left(\alpha\right)-\tilde{\mu}\right)^n\cdot f\left(q\left(\alpha\right)\right)\\
&\qquad-\frac{1}{1-2\alpha}\cdot\int_{q(\alpha)}^{q(1-\alpha)}n\left(x-\tilde{\mu}\right)^{n-1}\cdot\mathrm{IF}\left(r;\tilde{\mu},F\right)\cdot dF\left(x\right)\\
&\qquad+\frac{1}{1-2\alpha}\cdot\int_{q(\alpha)}^{q(1-\alpha)}\left(x-\tilde{\mu}\right)^n\cdot d\left[\delta_r(x)-F(x)\right]\\
&=\frac{1}{1-2\alpha}\cdot\left[\begin{array}{r}\left(\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)\right)\cdot\left(q\left(1-\alpha\right)-\tilde{\mu}\right)^n\\ -\left(\alpha-I_{0,\alpha}\left(r\right)\right)\cdot\left(q\left(\alpha\right)-\tilde{\mu}\right)^n\end{array}\right]\\
&\qquad-n\cdot\mathrm{IF}\left(r;\tilde{\mu},F\right)\tilde{\mu}_{n-1}+\frac{1}{1-2\alpha}\left(r-\tilde{\mu}\right)^n I_{\alpha,1-\alpha}\left(r\right)-\tilde{\mu}_n\\
&=\frac{1}{1-2\alpha}\cdot\left\{\left(\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)\right)\cdot\left(\left(q\left(1-\alpha\right)-\tilde{\mu}\right)^n-q\left(1-\alpha\right)\cdot n\cdot\tilde{\mu}_{n-1}\right)\right.\\
&\qquad-\left(\alpha-I_{0,\alpha}\left(r\right)\right)\cdot\left(\left(q\left(\alpha\right)-\tilde{\mu}\right)^n-q\left(\alpha\right)\cdot n\cdot\tilde{\mu}_{n-1}\right)\\
&\qquad\left.+I_{\alpha,1-\alpha}\left(r\right)\cdot\left(\left(r-\tilde{\mu}\right)^n-r\cdot n\cdot\tilde{\mu}_{n-1}\right)\right\}+n\cdot\tilde{\mu}_{n-1}\cdot\tilde{\mu}-\tilde{\mu}_n
\end{aligned}
\tag{2.41}
$$

Note the above can be rewritten in three parts that are parallel to the influence function (2.23) of the untrimmed moments

$$
\begin{aligned}
&\mathrm{IF}\left(r;\tilde{\mu}_n,\tilde{F}\right)\\
&=\frac{1}{1-2\alpha}\cdot\Big\{\Big((1-\alpha)-I_{0,1-\alpha}(r)\Big)\cdot\Big(q(1-\alpha)-\tilde{\mu}\Big)^n\\
&\qquad-\Big(\alpha-I_{0,\alpha}(r)\Big)\cdot\Big(q(\alpha)-\tilde{\mu}\Big)^n\\
&\qquad+I_{\alpha,1-\alpha}(r)\cdot\Big(r-\tilde{\mu}\Big)^n\Big\}\\
&-\frac{1}{1-2\alpha}\cdot n\cdot\tilde{\mu}_{n-1}\Big\{\Big((1-\alpha)-I_{0,1-\alpha}(r)\Big)\cdot q(1-\alpha)\\
&\qquad-\Big(\alpha-I_{0,\alpha}(r)\Big)\cdot q(\alpha)\\
&\qquad+I_{\alpha,1-\alpha}(r)\cdot r\Big\}+n\cdot\tilde{\mu}_{n-1}\cdot\tilde{\mu}\\
&-\tilde{\mu}_n
\end{aligned}
\tag{2.42}
$$

Now we proceed to the influence function for trimmed skewness

$$
\begin{aligned}
&\mathrm{IF}\left(r;S_{trim},\tilde{F}\right)\\
&=\mathrm{IF}\left(r;\tilde{\mu}_3,\tilde{F}\right)\cdot\tilde{\mu}_2{}^{-3/2}-\frac{3}{2}\tilde{\mu}_3\cdot\tilde{\mu}_2{}^{-5/2}\cdot\mathrm{IF}\left(r;\tilde{\mu}_2,\tilde{F}\right)\cdot\\
&=T_1+T_2
\end{aligned}
$$

We will write everything in parallel to the untrimmed version (2.28). The first term is

$$
\begin{aligned}
&T_1=\\
&\frac{1}{1-2\alpha}\cdot\Big\{\Big((1-\alpha)-I_{0,1-\alpha}(r)\Big)\cdot\Big(q(1-\alpha)-\tilde{\mu}\Big)^3\\
&\qquad-\Big(\alpha-I_{0,\alpha}(r)\Big)\cdot\Big(q(\alpha)-\tilde{\mu}\Big)^3\\
&\qquad+I_{\alpha,1-\alpha}(r)\cdot\Big(r-\tilde{\mu}\Big)^3\Big\}\cdot\frac{1}{\tilde{\sigma}^3}\\
&-\frac{1}{1-2\alpha}\cdot 3\cdot\frac{1}{\tilde{\sigma}}\Big\{\Big((1-\alpha)-I_{0,1-\alpha}(r)\Big)\cdot q(1-\alpha)\\
&\qquad-\Big(\alpha-I_{0,\alpha}(r)\Big)\cdot q(\alpha)\\
&\qquad+I_{\alpha,1-\alpha}(r)\cdot r\Big\}+3\cdot\tilde{\mu}/\tilde{\sigma}\\
&+\frac{1}{2}\tilde{\mu}_3\Big/\tilde{\sigma}^3
\end{aligned}
\tag{2.43}
$$

The second term is

$$
\begin{aligned}
T_2 &= -\frac{3}{2}\tilde{\mu}_3 \cdot \tilde{\mu}_2^{-5/2} \\
&= -\frac{3}{2}\tilde{\mu}_3 / \tilde{\sigma}^5 \cdot \frac{1}{1-2\alpha} \cdot \\
&\quad \left\{ \left((1-\alpha) - I_{0,1-\alpha}(r) \right) \cdot \left(q(1-\alpha) - \tilde{\mu} \right)^2 \right. \\
&\quad - \left(\alpha - I_{0,\alpha}(r) \right) \cdot \left(q(\alpha) - \tilde{\mu} \right)^2 \\
&\quad \left. + I_{\alpha,1-\alpha}(r) \cdot \left(r - \tilde{\mu} \right)^2 \right\}
\end{aligned}
\tag{2.44}
$$

If we set $\alpha = 0$, i.e. untrimmed, in the above equations, we have

$$
\begin{aligned}
&\mathrm{IF}\left(r; S_{trim=0}, F\right) \\
&= \frac{(r-\mu)^3}{\sigma^3} - 3 \cdot \frac{r-\mu}{\sigma} + \frac{1}{2}\frac{\mu_3}{\sigma^3} - \frac{3}{2}\frac{\mu_3 (r-\mu)^2}{\sigma^5} \\
&= \frac{(r-\mu)^3 - 3(r-\mu)\sigma^2 + \frac{1}{2}\mu_3}{\sigma^3} - \frac{3\mu_3}{2\sigma^5}(r-\mu)^2
\end{aligned}
\tag{2.45}
$$

which is exactly what we got in the original form in (2.28).

As a special case, we assume a symmetric and centered distribution where

$$
\begin{aligned}
&\tilde{\mu} = 0, \\
&\tilde{\mu}_3 = 0, \\
&q(\alpha) = -q(1-\alpha)
\end{aligned}
\tag{2.46}
$$

we have

$$
\begin{aligned}
&\mathrm{IF}\left(r;S_{trim},F\right)\\
&=\frac{1}{1-2\alpha}\cdot\Big\{\Big(\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)\Big)\cdot\left(-q\left(\alpha\right)\right)^{3}\\
&\qquad-\Big(\alpha-I_{0,\alpha}\left(r\right)\Big)\cdot\left(q\left(\alpha\right)\right)^{3}\\
&\qquad+I_{\alpha,1-\alpha}\left(r\right)\cdot r^{3}\Big\}\cdot\frac{1}{\tilde{\sigma}^{3}}\\
&-\frac{1}{1-2\alpha}\cdot 3\cdot\frac{1}{\tilde{\sigma}}\Big\{\Big(\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)\Big)\cdot\left(-q\left(\alpha\right)\right)\\
&\qquad-\Big(\alpha-I_{0,\alpha}\left(r\right)\Big)\cdot q\left(\alpha\right)\\
&\qquad+I_{\alpha,1-\alpha}\left(r\right)\cdot r\Big\}\\
&=\frac{1}{1-2\alpha}\cdot\frac{1}{\tilde{\sigma}^{3}}\cdot\Big\{\Big[I_{0,\alpha}\left(r\right)\cdot q\left(\alpha\right)^{3}+I_{\alpha,1-\alpha}\left(r\right)\cdot r^{3}-I_{1-\alpha,1}\left(r\right)\cdot q\left(\alpha\right)^{3}\Big]\\
&\qquad-3\cdot\tilde{\sigma}^{2}\Big[I_{0,\alpha}\left(r\right)\cdot q\left(\alpha\right)+I_{\alpha,1-\alpha}\left(r\right)\cdot r-I_{1-\alpha,1}\left(r\right)\cdot q\left(\alpha\right)\Big]\Big\}
\end{aligned}
\tag{2.47}
$$

which is close to what we get in the untrimmed case The main difference is that the original cubic form is truncated at the trimming sites and it becomes flat beyond these points. The inside of trimmed region is “compressed” (Fig. 2.3). It is essentially a set of splines with two knots at the trimming points. In Fig. 2.3, we plot the asymptotic influence function for trimmed skewness (red dashed lines) on top of the influence functions of classical skewness (black curve). The underlying distribution is standard normal.

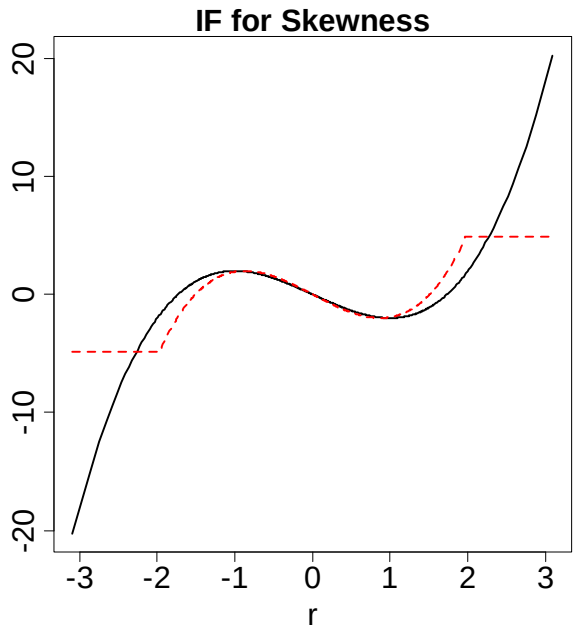

Figure 2.3 Influence Functions for Trimmed Skewness: Standard Normal Distribution, Trimming 2.5% on both Tails

Now we proceed to trimmed kurtosis. Given the parallel form above, we will directly write out the influence function for trimmed kurtosis.

$$\begin{aligned}&\mathrm{IF}\left(r;K_{trim},\tilde{F}\right)\\&=T_1+T_2\end{aligned} \tag{2.48}$$

The two terms are as follows

$$\begin{aligned}T_1=&\frac{1}{1-2\alpha}\cdot\left\{\left(\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)\right)\cdot\left(q\left(1-\alpha\right)-\tilde{\mu}\right)^4\right.\\&\left.-\left(\alpha-I_{0,\alpha}\left(r\right)\right)\cdot\left(q\left(\alpha\right)-\tilde{\mu}\right)^4+I_{\alpha,1-\alpha}\left(r\right)\cdot\left(r-\tilde{\mu}\right)^4\right\}\cdot\frac{1}{\tilde{\sigma}^4}\\&-\frac{1}{1-2\alpha}\cdot4\cdot\frac{\tilde{\mu}_3}{\tilde{\sigma}^4}\left\{\left(\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)\right)\cdot q\left(1-\alpha\right)\right.\\&\left.-\left(\alpha-I_{0,\alpha}\left(r\right)\right)\cdot q\left(\alpha\right)+I_{\alpha,1-\alpha}\left(r\right)\cdot r\right\}+4\cdot\frac{\tilde{\mu}\cdot\tilde{\mu}_3}{\tilde{\sigma}^4}\\&+\tilde{\mu}_4/\tilde{\sigma}^4\end{aligned} \tag{2.49}$$

and

$$T_2 = -2\,\tilde{\mu}_4 / \tilde{\sigma}^6 \cdot \frac{1}{1-2\alpha} \cdot \left\{ \left((1-\alpha) - I_{0,1-\alpha}(r) \right) \cdot \left(q(1-\alpha) - \tilde{\mu} \right)^2 - \left(\alpha - I_{0,\alpha}(r) \right) \cdot \left(q(\alpha) - \tilde{\mu} \right)^2 + I_{\alpha,1-\alpha}(r) \cdot \left(r - \tilde{\mu} \right)^2 \right\} \tag{2.50}$$

It is obvious we have bounded the influence function. Other than that, the influence function stays pretty much the same under standard normal. We plot (2.50) under standard normal below (Fig. 2.4). The influence function for trimmed kurtosis is in red dashed line and the influence function of classical kurtosis is in solid black curve. We will show influence functions of trimmed skewness and kurtosis under skewed and fat-tailed distributions in later sections.

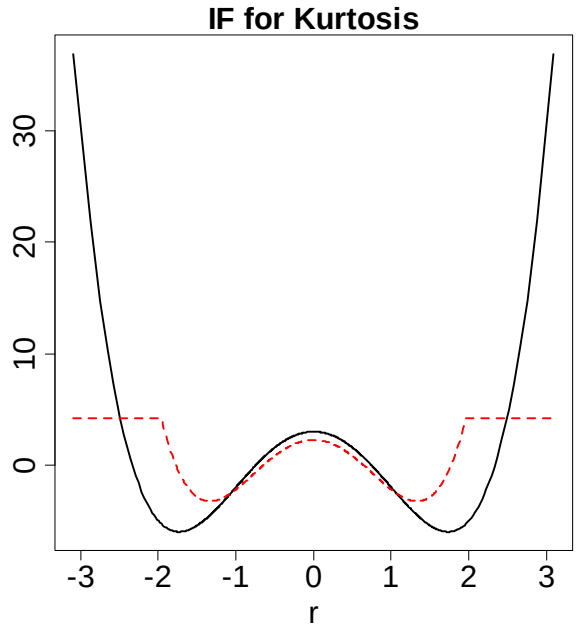

Figure 2.4 Influence Functions for Trimmed Kurtosis: Standard Normal Distribution, Trimming 2.5% on both Tails

2.2.3 *Quantile Based Skewness and Kurtosis*

We continue to show the influence functions for quantile based robust skewness and kurtosis (see also Groeneveld (1991)) . First, for quantile based robust skewness, we follow (2.16) and (2.39),

$$
\begin{aligned}
&\mathrm{IF}\left(r; S_q, F\right) \\
&= \mathrm{IF}\left(r; \frac{q(\alpha)+q(1-\alpha)-2q(0.5)}{q(1-\alpha)-q(\alpha)}, F \right) \\
&= \frac{1}{\left(q(1-\alpha)-q(\alpha)\right)^2} \cdot \\
&\quad \left\{ \left[\frac{\alpha - I_{0,\alpha}(r)}{f_X\left(q(\alpha)\right)} + \frac{(1-\alpha)-I_{0,1-\alpha}(r)}{f_X\left(q(1-\alpha)\right)} - 2\frac{0.5-I_{0,0.5}(r)}{f_X\left(q(0.5)\right)} \right] \cdot \left(q(1-\alpha)-q(\alpha)\right) \right. \\
&\quad \left. -\left[q(\alpha)+q(1-\alpha)-2q(0.5)\right] \cdot \left\{ \frac{(1-\alpha)-I_{0,1-\alpha}(r)}{f_X\left(q(1-\alpha)\right)} - \frac{\alpha - I_{0,\alpha}(r)}{f_X\left(q(\alpha)\right)} \right\} \right\}
\end{aligned}
\tag{2.51}
$$

For a symmetric distribution centered at 0, we have

$$
\begin{aligned}
&q(0.5) = 0 \\
&q(a) = -q(1-a) \\
&f\left(q(a)\right) = f\left(q(1-a)\right)
\end{aligned}
\tag{2.52}
$$

Hence (2.51) reduces to

$$
\begin{aligned}
&\mathrm{IF}\left(r;S_q,F\right)\\
&=\frac{1}{\left\{2q\left(1-\alpha\right)\right\}^2}\cdot\left[\left(\frac{\alpha-I_{0,\alpha}\left(r\right)}{f_X\left(q\left(\alpha\right)\right)}+\frac{\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)}{f_X\left(q\left(\alpha\right)\right)}-2\frac{0.5-I_{0,0.5}\left(r\right)}{f_X\left(0\right)}\right)\cdot\left(2q\left(1-\alpha\right)\right)\right]\\
&=\frac{1}{2q\left(1-\alpha\right)}\left[\frac{1-I_{0,\alpha}\left(r\right)-I_{0,1-\alpha}\left(r\right)}{f_X\left(q\left(\alpha\right)\right)}-\frac{1-2\cdot I_{0,0.5}\left(r\right)}{f_X\left(0\right)}\right]\\
&=\frac{1}{2q\left(1-\alpha\right)}\left\{-\left(\frac{1}{f_X\left(q\left(\alpha\right)\right)}-\frac{1}{f_X\left(0\right)}\right)I_{0,\alpha}\left(r\right)+\frac{I_{\alpha,0.5}\left(r\right)\text{-}I_{0.5,1-\alpha}\left(r\right)}{f_X\left(0\right)}\right.\\
&\qquad\left.+\left(\frac{1}{f_X\left(q\left(\alpha\right)\right)}-\frac{1}{f_X\left(0\right)}\right)I_{1-\alpha,1}\left(r\right)\right\}
\end{aligned}
$$

(2.53)

It is piecewise flat. The positions where the influence function in (2.51) changes value is $q\left(\alpha\right),q\left(0.5\right),q\left(1-\alpha\right)$. We showcase (2.51) under standard normal in Fig. 2.5. The black curve is the influence function for quantile based robust skewness. The red dashed line is the influence function for trimmed skewness.

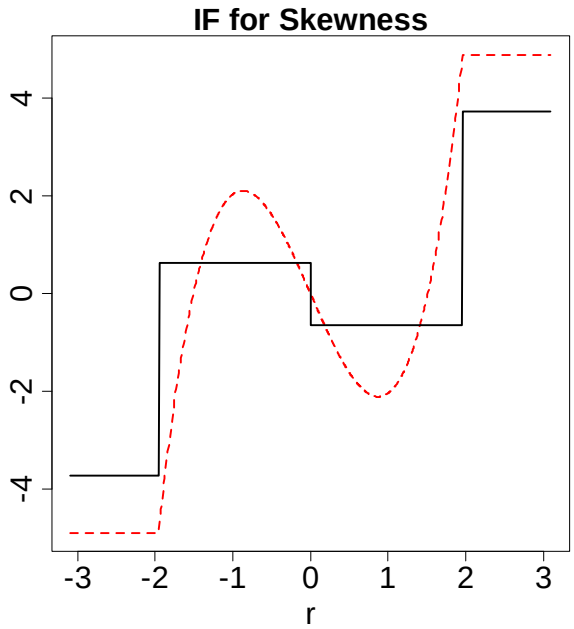

Figure 2.5 Influence Functions for Quantile Based Robust Skewness: Standard Normal Distribution $\alpha=2.5\%$

It is a “flat” version of Fig. 2.3. The quantile based robust skewness neglects the details between the three quantiles used in its definition. It has a bounded influence function. The impact by outliers (the value on the influence function) is close to that by the trimmed version under the standard normal distribution.

For quantile based robust kurtosis, we have

$$\begin{aligned}
&\mathrm{IF}\left(r;K_q,F\right)\\
&=\mathrm{IF}\left(r;\frac{q\left(1-\alpha\right)-q\left(\alpha\right)}{q\left(0.75\right)-q\left(0.25\right)},F\right)\\
&=\frac{1}{\left(q\left(0.75\right)-q\left(0.25\right)\right)^2}\cdot\left\{\left[\frac{\left(1-\alpha\right)-I_{0,1-\alpha}\left(r\right)}{f_X\left(q\left(1-\alpha\right)\right)}-\frac{\alpha-I_{0,\alpha}\left(r\right)}{f_X\left(q\left(\alpha\right)\right)}\right]\cdot\left(q\left(0.75\right)-q\left(0.25\right)\right)\right.\\
&\qquad\left.-\left(q\left(1-\alpha\right)-q\left(\alpha\right)\right)\cdot\left\{\frac{0.75-I_{0,0.75}\left(r\right)}{f_X\left(q\left(0.75\right)\right)}-\frac{0.25-I_{0,0.25}\left(r\right)}{f_X\left(q\left(0.25\right)\right)}\right\}\right\}
\end{aligned}$$

(2.49)

We plot (2.54) under standard normal distribution in Fig. 2.6. The black curve is the influence function for quantile based robust kurtosis. The red dashed line is the influence function for trimmed kurtosis.

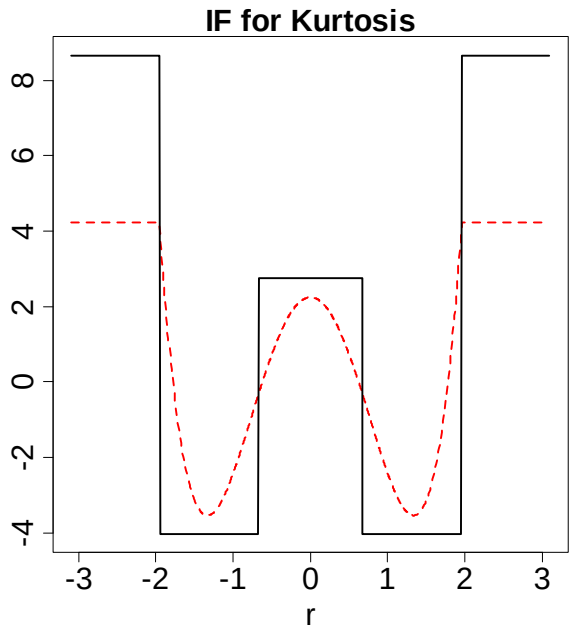

Figure 2.6 Influence Functions for Quantile Based Robust Kurtosis: Standard Normal Distribution $\alpha = 2.5\%$

It is also bounded. However, it is interesting to see that the influence function for quantile based robust kurtosis is flat "W" shape. It is also close to the trimmed version.

2.2.4 *Influence Functions of Skewness and Kurtosis under Skewed and Fat-Tailed Distributions*

In this section, we will give examples of influence function we derived in 2.2.1, 2.2.2 and 2.2.3 under skewed and fat-tailed distributions. Note in some cases, e.g. t-distribution with $df = 4$, there is no finite kurtosis under the classical definition. We replace it with a minimally truncated student-t distribution with $df = 4$. The truncation happens at 0.01% and 99.99% quantile of the distribution. Therefore the moments are finite under this setting. We include such cases to compare with the robust versions of kurtosis and show the extent of reduction of impact by outliers.

Student-t Distribution with $df = 4$

The classical skewness and kurtosis. The red dashed curve is the influence function under standard normal distribution for comparison purpose.

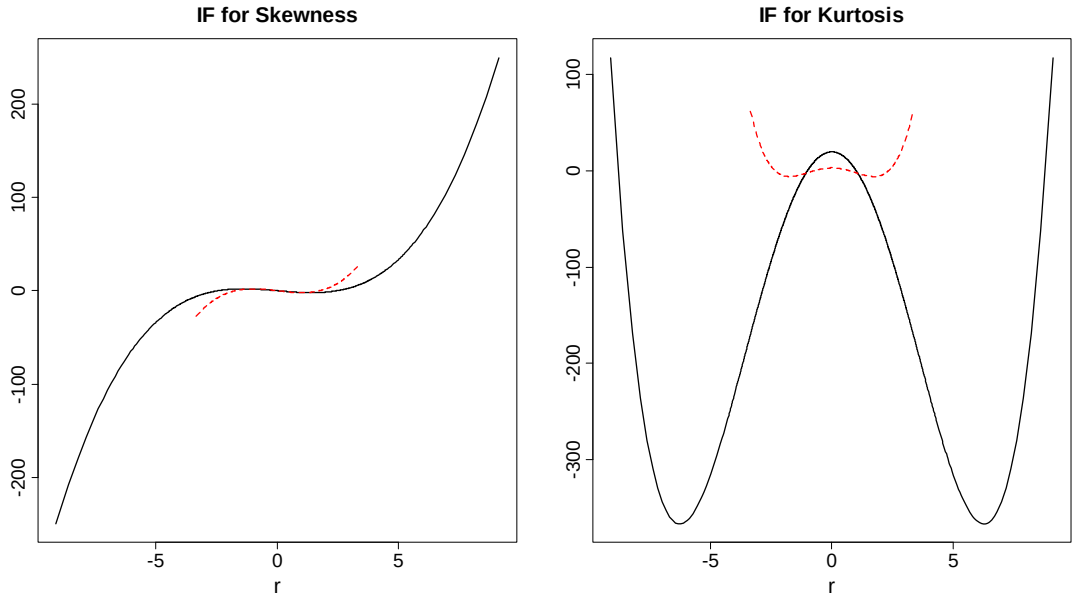

Figure 2.7 Influence Functions for Classical Skewness and Kurtosis: Student-t Distribution with $df = 4$ **(black) and Standard Normal Distribution (red dashed)**

The difference of impact between the student-t distribution and the normal distribution is very obvious.

The trimmed skewness and kurtosis ($\alpha = 2.5\%$). The black curve was the original influence function for untrimmed case. The read dashed curve is the trimmed case. The reduction of influence is dramatic.

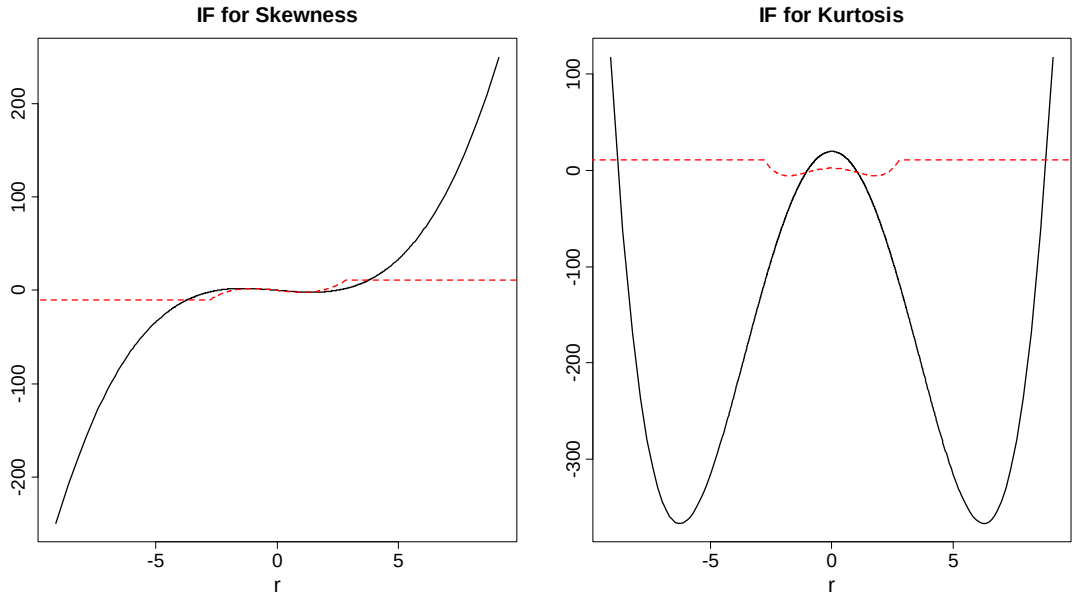

Figure 2.8 Influence Functions for Trimmed Skewness and Kurtosis: Student-t Distribution with $df = 4$

The quantile based skewness and kurtosis ($\alpha = 2.5\%$)

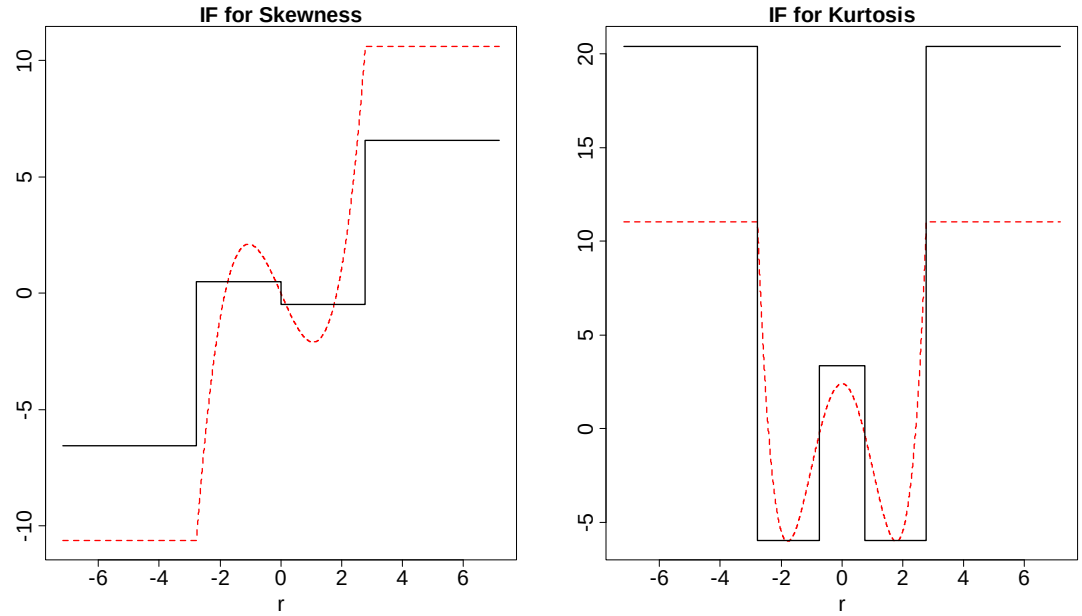

Figure 2.9 Influence Functions for Quantile Based Robust Skewness and Kurtosis: Student-t Distribution with $df = 4$

In Fig. 2.7, 2.8, 2.9 we display the classical, trimmed and quantile skewness and kurtosis influence functions for student-t distribution with $df = 4$. By sheer comparison in the scale we can see that the classical estimates can be dramatically swung by the outliers, where the two robust versions are bounded at a very small range. The difference in kurtosis influence functions are more than 100 times between the classical and the robust versions. With degrees of freedom set at 4, this is a very fat-tailed distribution. We follow this example by comparing the student-t distribution under $df = 10$

Student-t Distribution with $df = 10$

The classical skewness and kurtosis

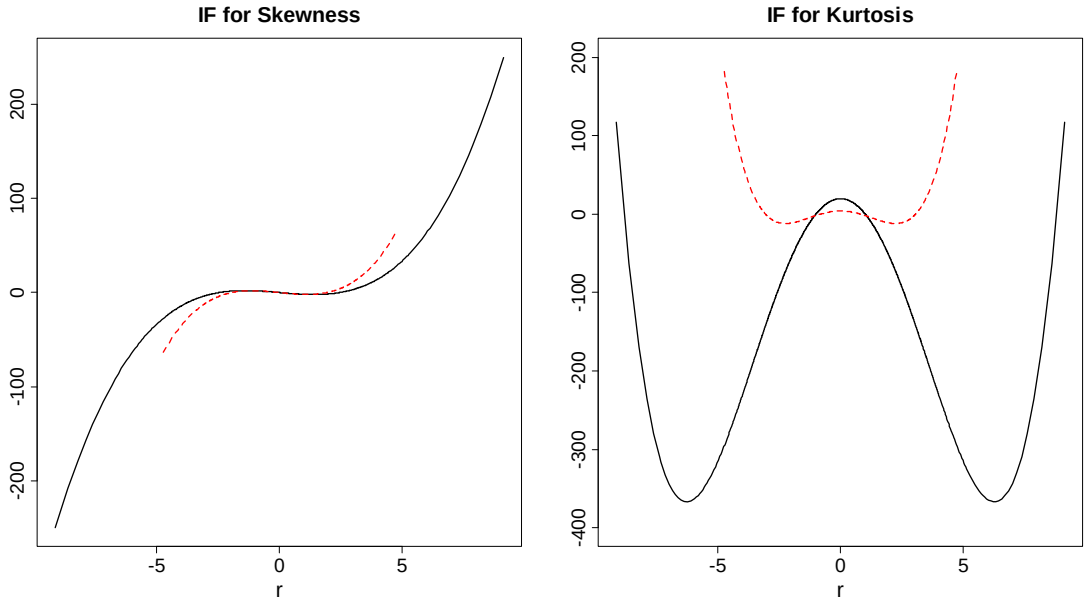

Figure 2.10 Influence Functions for Classical Skewness and Kurtosis: Student-t Distribution with $df = 10$ (black) and $df = 4$ (red dashed)

The trimmed skewness and kurtosis ($\alpha = 2.5\%$)

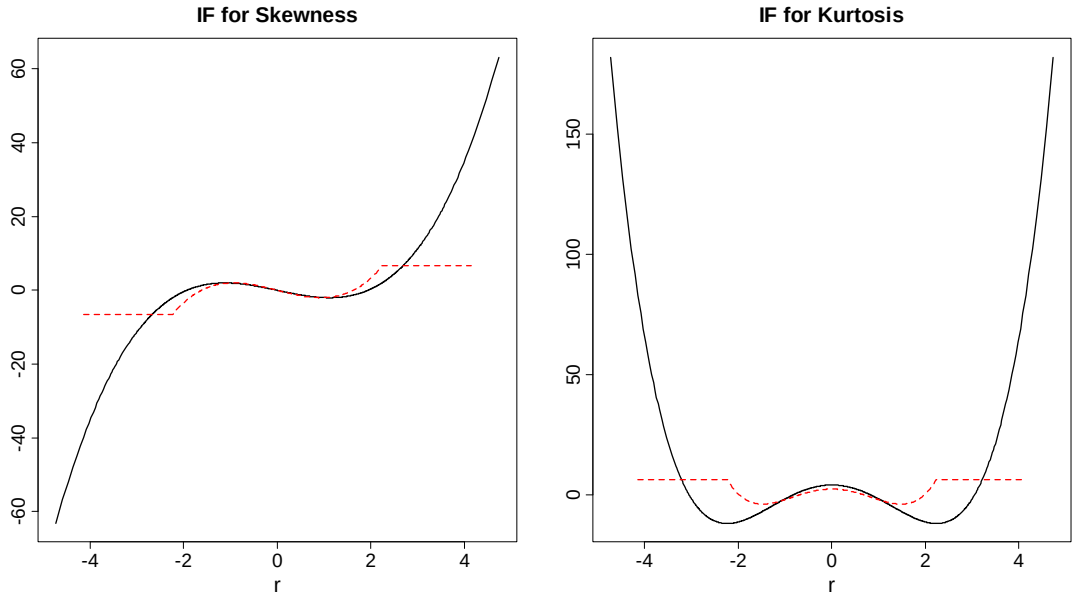

Figure 2.11 Influence Functions for Trimmed Skewness and Kurtosis: Student-t Distribution with $df = 10$

The quantile based skewness and kurtosis ($\alpha = 2.5\%$)

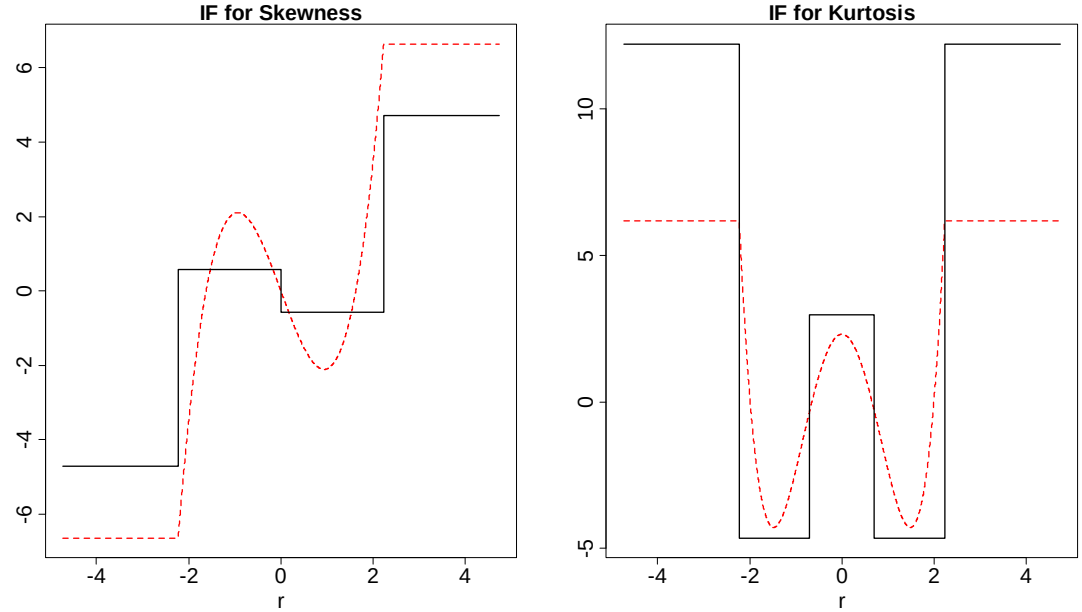

Figure 2.12 Influence Functions for Quantile Based Robust Skewness and Kurtosis: Student-t Distribution with $df = 10$

In this case (Fig. 2.10, 2.11, 2.12), we can see the difference between classical and robust versions are smaller than the previous more fat-tailed distribution. The shapes are

the same except all trimming and robust sites are closer to the center because the quantiles are closer to the center under less fat-tailed distribution. The symmetry in the original distribution is more or less reserved in all influence functions.

Skewed-t Distribution with $df = 10, \gamma = 1.25$

Besides the most commonly know student-t distribution, we'd like to show examples under skewed distributions, such as the skewed-t distribution, as well. The skewed-t distribution we use here is defined by

$$f(x;n,\gamma) = \begin{cases} \frac{2}{\gamma + 1/\gamma} t(\gamma \cdot x; n) & \text{for } x < 0 \\ \frac{2}{\gamma + 1/\gamma} t(\frac{1}{\gamma} x; n) & \text{for } x \geq 0 \end{cases} \tag{2.55}$$

where $t(\cdot; n)$ is the density of student-t distribution with *n* degrees of freedom.

Under the parameters we choose $df = 10, \gamma = 1.25$, we have the following probability density plot (Fig. 2.12). The red dotted line indicates the trimming sites (2.5% and 97.5%).

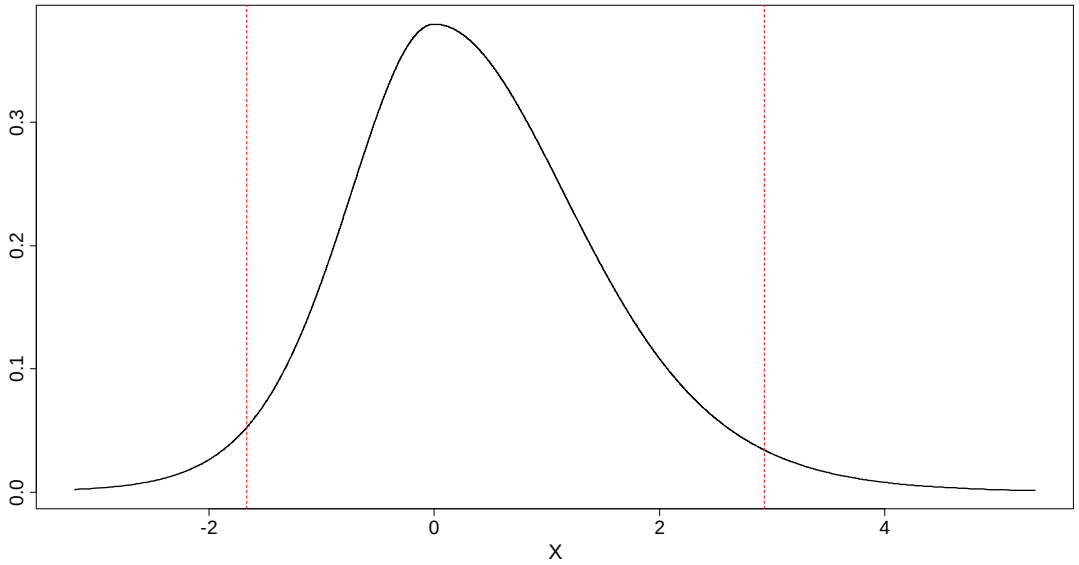

Figure 2.13 Probability Density for Skewed-t Distribution with $df = 10, \gamma = 1.25$. The red dotted-lines are the trimming sites at 2.5% and 97.5% quantiles.

The classical skewness and kurtosis

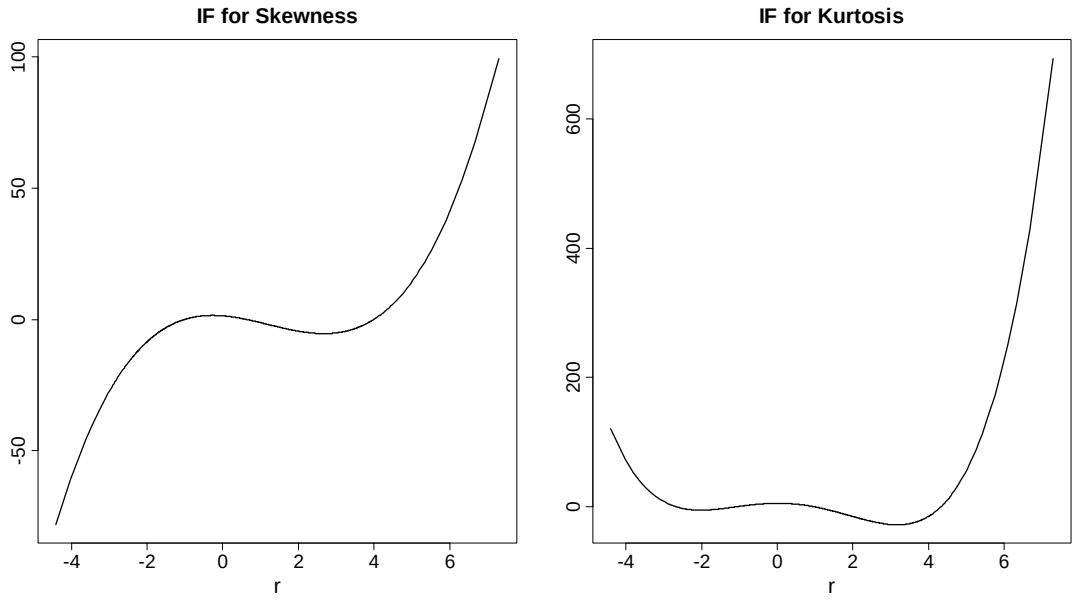

Figure 2.14 Influence Functions for Classical Skewness and Kurtosis: Skewed-t Distribution with $df = 10, \gamma = 1.25$

The trimmed and quantile based robust skewness and kurtosis ($\alpha = 2.5\%$). The black solid curve is the influence function of the quantile based robust skewness and kurtosis. The red dashed curve is the influence function for trimmed skewness and kurtosis.

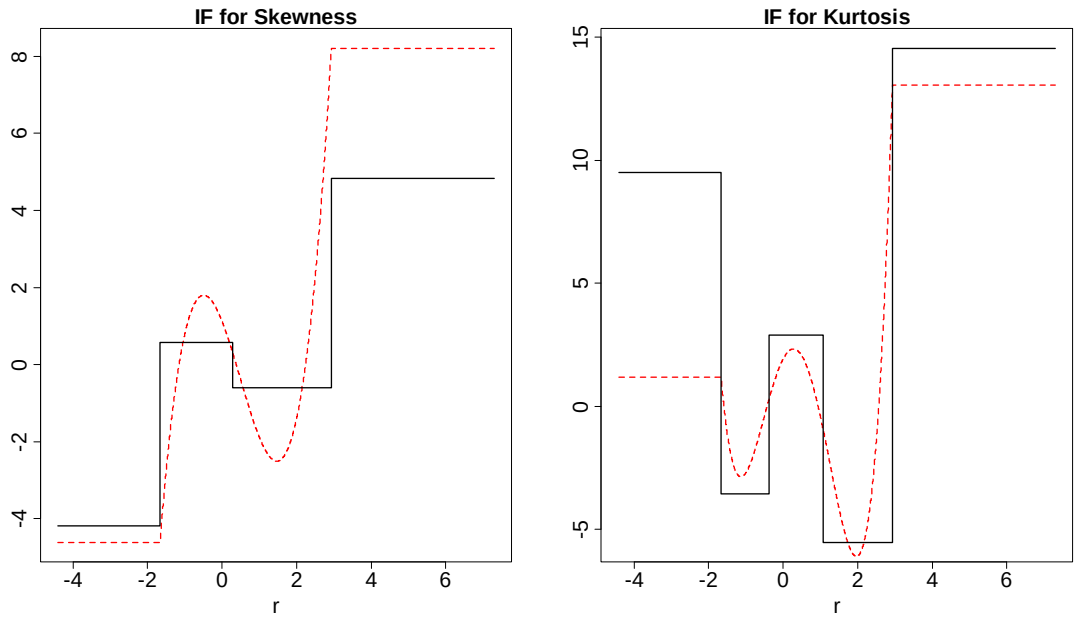

Figure 2.15 Influence Functions for Quantile Based Robust Skewness and Kurtosis: Skewed-t Distribution with $df = 10, \gamma = 1.25$

This is a right skewed-distribution with most of its density between -2 and 2. From Fig. 2.14 and 2.15, we can see both robust version of skewness and kurtosis bounded their influence functions pretty well. The influence function for skewness is no longer symmetric. It has longer up swing on the right side of the distribution. This is reasonable because an outlier on the right will further skew the already right skewed distribution. The influence function for kurtosis shows a similar effect. Because of the right skewness, the tail on the right is “fatter” than the tail on the left, giving an outlier on the right more apparent effect than one on the left. Also because of the break of symmetry, the quantile based robust kurtosis shows more structure in its influence function as shown in Fig. 2.15.

Skewed-t Distribution with $df = 10, \gamma = 0.8$

In the same spirit, we check the same but left skewed distribution. We take the new α to be the reciprocal of the case above. The density and trimming sites are shown in Fig. 2.16.

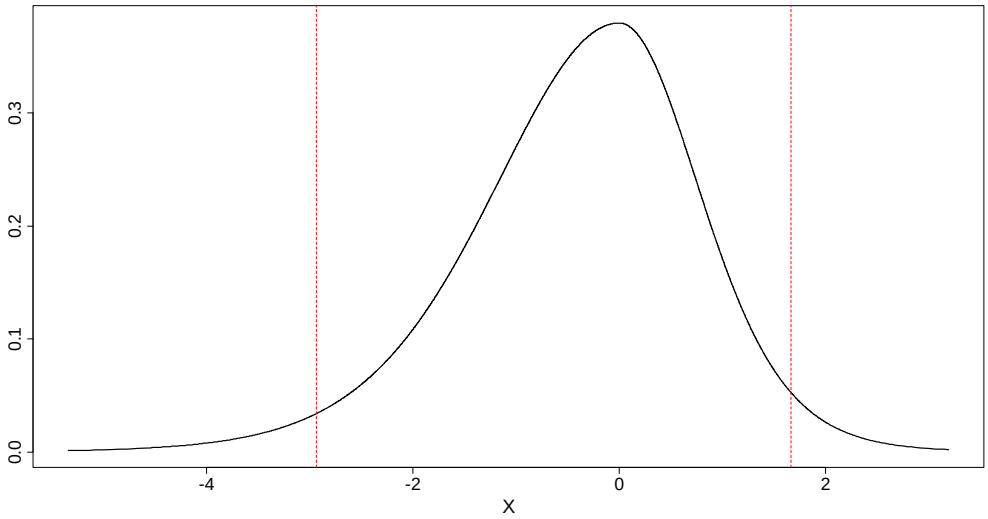

Figure 2.16 Probability Density for Skewed-t Distribution with $df = 10, \gamma = 0.8$. The red dotted-lines are the trimming sites at 2.5% and 97.5% quantiles.

The classical skewness and kurtosis

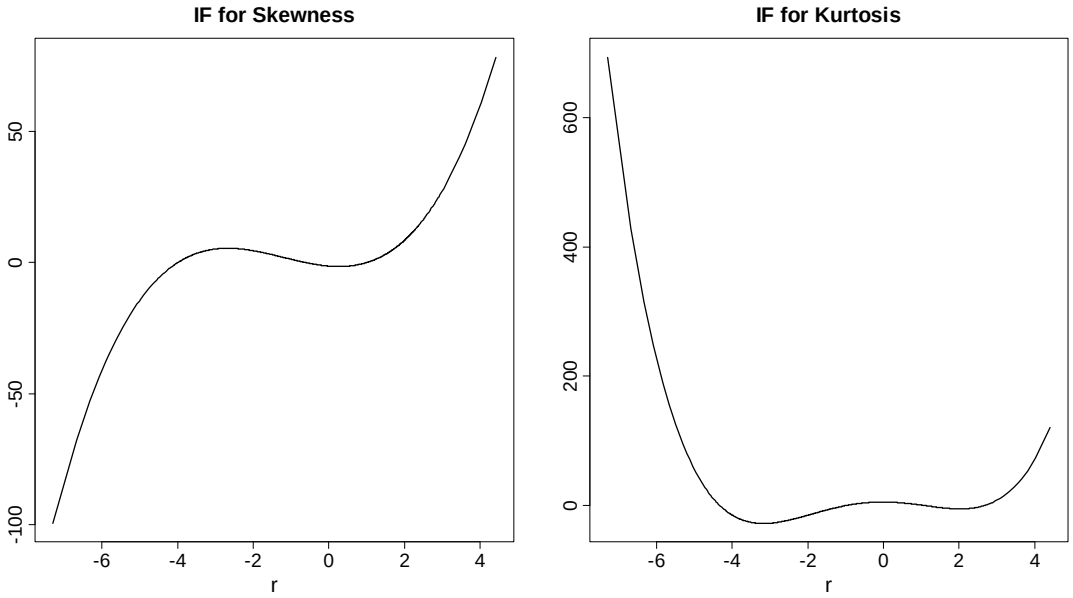

Figure 2.17 Influence Functions for Classical Skewness and Kurtosis: Skewed-t Distribution with $df = 10, \gamma = 0.8$

The trimmed skewness and kurtosis ($\alpha = 2.5\%$)

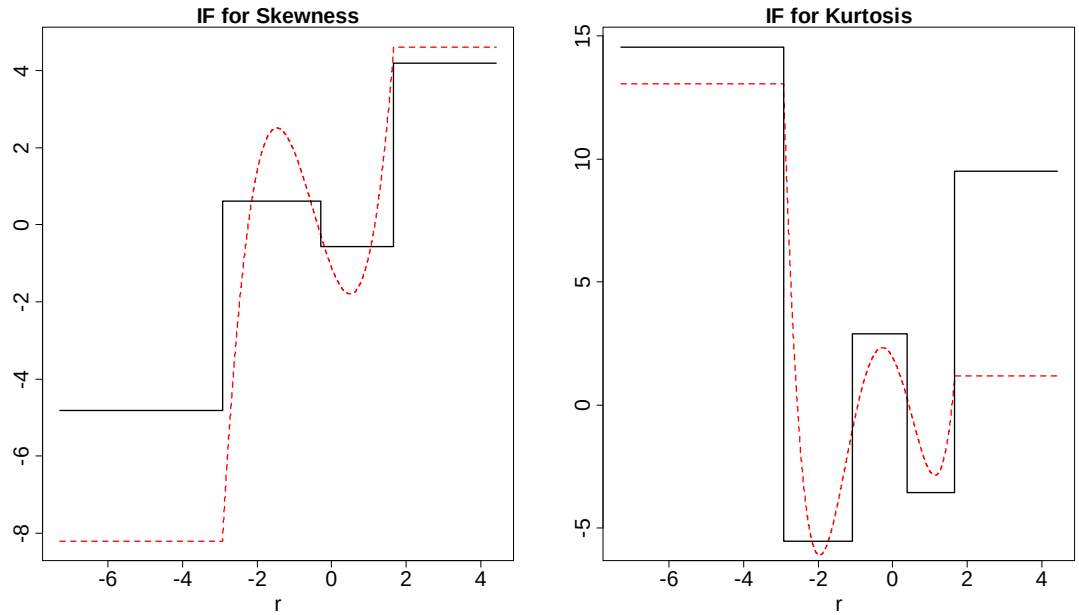

Figure 2.18 Influence Functions for Quantile Based Robust Skewness and Kurtosis: Skewed-t Distribution with $df = 10, \gamma = 0.8$

The black solid curve is the influence function of the quantile based robust skewness and kurtosis. The red dashed curve is the influence function for trimmed skewness and

kurtosis. Under this setting, we see all influence functions are mirrors of the previous section. This situation is more likely to happen in a downturn of the market. The effects are similar to what we stated previously, but in an opposite direction.

2.2.5 Asymptotic Variance

Following Vaart (1998), we derive the asymptotic variance of the sample skewness and kurtosis parameters. Suppose random variable X has finite 8th moment. The sample skewness and kurtosis can be written as $\phi_S\left(\overline{X},\overline{X^2},\overline{X^3},\overline{X^4}\right)$ and $\phi_K\left(\overline{X},\overline{X^2},\overline{X^3},\overline{X^4}\right)$ given by

$$\begin{aligned}\phi_S(a,b,c,d) &= \frac{c-3ab+2a^3}{\left(b-a^2\right)^{3/2}}\\ \phi_K(a,b,c,d) &= \frac{d-4ac+6a^2b-3a^4}{\left(b-a^2\right)^2}\end{aligned} \tag{2.56}$$

Denote $\mu = E(X)$ and $\mu_n = E(X-\mu)^n$. For convenience, we also denote $\sigma = \sqrt{\mu_2}$. Hence $S = \mu_3/\sigma^3$ and $K = \mu_4/\sigma^4$ (we didn't include the constant adjustment for kurtosis here, because it does not affect its variance calculation). For simplicity, we define $Y = (X-\mu)/\sigma$, the standardized variable. Note X and Y have the same sample skewness and kurtosis, i.e. $\phi_S\left(\overline{X},\overline{X^2},\overline{X^3},\overline{X^4}\right) = \phi_S\left(\overline{Y},\overline{Y^2},\overline{Y^3},\overline{Y^4}\right)$ and $\phi_K\left(\overline{X},\overline{X^2},\overline{X^3},\overline{X^4}\right) = \phi_K\left(\overline{Y},\overline{Y^2},\overline{Y^3},\overline{Y^4}\right)$. We can now directly work out the variance of sample skewness and kurtosis under Y. By central limit theorem, we have

$$\sqrt{n}\begin{pmatrix} \overline{Y} \\ \overline{Y^2}-1 \\ \overline{Y^3}-S \\ \overline{Y^4}-K \end{pmatrix} \to_d N\left(0, \begin{pmatrix} 1 & S & K & \mu_5/\sigma^5 \\ S & K-1 & \mu_5/\sigma^5 - S & \mu_6/\sigma^6 - K \\ K & \mu_5/\sigma^5 - S & \mu_6/\sigma^6 - S^2 & \mu_7/\sigma^7 - SK \\ \mu_5/\sigma^5 & \mu_6/\sigma^6 - K & \mu_7/\sigma^7 - SK & \mu_8/\sigma^8 - K^2 \end{pmatrix}\right). \quad (2.57)$$

$$\triangleq N(0,\Sigma)$$

We use the delta method to get the asymptotic variance for sample skewness first. Take derivatives of ϕ_S at the point $(0,1,S)$ we get $D_S = (-3,-3S/2,1,0)$. Therefore the asymptotic variance of sample skewness is $D_S \Sigma D_S^T$. Similarly, take derivatives of ϕ_K at the point $(0,1,S,K)$ we get $D_K = (-4S,-2K,0,1)$. Therefore the asymptotic variance of sample skewness is $D_K \Sigma D_K^T$.

Now we list the asymptotic variances calculated using the formula above for different distributions.

Table 2.1 Asymptotic Variance of Sample Skewness and Kurtosis under Different Distributions.

	Asymptotic Variance	
	Skewness	*Kurtosis*
Standard Normal	6	24
Student-t, $df = 4$	∞	∞
Student-t, $df = 10$	25	720
Skewed-t $df=10, \gamma=1.25$	28.8	1578.8

It's easy to find that the asymptotic variance of sample skewness and kurtosis explode under skewed and fat-tailed distributions. In particular, while the degrees of freedom are the same, the skewed-t distribution has much larger asymptotic variance for its sample kurtosis estimator.

2.2.6 Finite Sample Variance

In this section, we explore the effect of the bounded influence functions for curbing the variances on the sample estimators for skewness and kurtosis. The robust estimators should reduce the impact of the outliers and stabilize the variability in the sample estimates.

We compare the finite sample variance of the skewness and kurtosis estimates we compard in this chapter. In the following two tables we display the finite sample variances of the classical estimates, and the two robust estimates of skewness and kurtosis, under the normal distribution for various sample sizes. This is done using both the finite sample approximation based on the asymptotic variances (labeled “Asymptotic Approximation”) and Monte Carlo simulation with 10,000 replicates (labeled “Simulation”). The asymptotic variances are computed based on the expression in terms of the influence function (see for example, Hampel et. al., 1986):

$$Var_{\infty}\left(\theta\right) = E\left(IF^{2}\left(x;\theta,F\right)\right) \tag{2.59}$$

This will reach the same results as we got in the last section. We estimate the finite sample variances from (2.59) and substitute arithmetic means for the expectation, and dividing by the sample size. The numbers in parentheses are the standard errors of each estimate.

Table 2.2 Standard Deviations of Sample Skewness under Three Different Methods

Standard deviations of Sample Skewness, Standard Normal Distribution						
	Asymptotic Approximation			*Simulation*		
Sample Size	*Classical*	*Trimmed*	*Quantile*	*Classical*	*Trimmed*	*Quantile*
20	0.5477 (0.0102)	0.4183 (0.0031)	0.2296 (0.0031)	0.4397 (0.0036)	0.4397 (0.0036)	0.663 (0.0061)
30	0.4472 (0.0083)	0.3415 (0.0025)	0.1875 (0.0026)	0.3815 (0.003)	0.3815 (0.003)	0.5303 (0.0043)
60	0.3162 (0.0059)	0.2415 (0.0018)	0.1326 (0.0018)	0.2894 (0.0023)	0.2468 (0.0018)	0.3861 (0.0031)
120	0.2236 (0.0042)	0.1708 (0.0013)	0.0937 (0.0013)	0.2172 (0.0016)	0.172 (0.0012)	0.2711 (0.0021)
1000	0.0775 (0.0014)	0.0592 (4e-04)	0.0325 (4e-04)	0.0779 (6e-04)	0.0601 (4e-04)	0.0961 (7e-04)

Table 2.3 Standard Deviations of Sample Kurtosis under Three Different Methods

Standard deviations of Sample Kurtosis, Standard Normal Distribution						
	Asymptotic Approximation			*Simulation*		
Sample Size	*Classical*	*Trimmed*	*Quantile*	*Classical*	*Trimmed*	*Quantile*
20	1.0954 (0.0139)	0.4904 (0.0024)	0.8611 (0.0042)	0.6961 (0.0095)	0.6961 (0.0095)	0.8511 (0.0097)
30	0.8944 (0.0113)	0.4004 (0.0019)	0.7031 (0.0034)	0.6473 (0.0086)	0.6473 (0.0086)	0.74 (0.0085)
60	0.6325 (0.008)	0.2831 (0.0014)	0.4972 (0.0024)	0.5319 (0.0067)	0.3522 (0.0034)	0.4948 (0.0041)
120	0.4472 (0.0057)	0.2002 (0.001)	0.3515 (0.0017)	0.4121 (0.0046)	0.2165 (0.0018)	0.3491 (0.0027)
1000	0.1549 (0.002)	0.0693 (3e-04)	0.1218 (6e-04)	0.1541 (0.0012)	0.0706 (5e-04)	0.1218 (9e-04)

We note the following:

1. The bounded influence function of the robust estimators results in substantially reduced variances of robust estimators relative to the classical estimators.
2. The asymptotic approximation works well for the trimmed skewness estimates at all sample sizes, while it underestimates the simulation based estimate by an amount for sample sizes 20 and 30 and is essentially negligible at sample size 60.

This is no doubt because larger sample sizes are needed in order for the standard deviation to fully reflect the largest observations even in the case of a normal distribution, whereas these observations do not contribute much to the standard deviation of the robust estimates.

3. The asymptotic approximation underestimates the quantile based robust skewness estimates at all sample sizes. This is probably because for the asymptotic variance calculated using influence functions (2.51), the second term is completely cancelled out because of the symmetry for normal distribution. However, under real simulations, the actual distribution is far from symmetric, hence elevating variance in the estimator. The quantile based robust kurtosis estimator does not have this property and the asymptotic approximation works quite well under all circumstances.
4. The asymptotic approximation of the standard deviation of trimmed robust kurtosis estimates underestimate the actual standard deviation for sample sizes 20, 30 and 60, but the approximation is quite good at sample size 120 and higher. This is because trimming of 2.5% at each end does not work exactly for small sample sizes.
5. The asymptotic approximation for the classical kurtosis estimate is an overestimate of the actual standard deviation until at least sample size 60 and is fairly on target for sample size 120 and above. This is because that the trimmed and untrimmed kurtosis is almost the same when sample sizes are small. This means we are overestimating variance of the trimmed version and underestimating the variance of the classical version under small sample sizes.

Applying the same technique to other heavy tailed or asymmetric distributions is expected to show similar conclusions. Below is a table showing the standard deviation of

the robust estimates of skewness and kurtosis as a percentage relative to the classical estimates under the normal distribution and three non-normal distributions, two of which are fat-tailed and the last one is skewed. These percentages are based on Monte Carlo simulation with 10,000 replicates at sample size 120.

Table 2.4 Standard Deviation of Sample Skewness and Kurtosis: Trimming and Quantile methods vs. Classical Estimates. Standard deviations are presented as a percentage of the original standard deviations obtained by classical methods.

	Skewness		*Kurtosis*	
	Trimming	*Quantile*	*Trimming*	*Quantile*
Standard Normal	79.4%	124.0%	52.5%	84.4%
Student-t, $df = 4$	25.7%	47.9%	6.8%	8.8%
Student-t, $df = 10$	56.1%	93.7%	24.2%	37.0%
Skewed-t $df=10, \gamma=1.25$	53.3%	92.7%	22.5%	28.4%

All experiments show significant reduction in variance compared to the classical estimates. The more skewed (skewed-t) and the more fat-tailed (student-t compared to normal) the distributions are, the more reduction of variances of our estimators get. The reduction on the variance of sample kurtosis is significant, particularly in the case of student-t, $df = 4$. The trimming method does better in all cases.

Overall we prefer to focus on symmetric trimming for the remainder of this dissertation. The influence function for trimmed version is a smooth function (no jumps) that retains the form of influence functions of the classical estimators better within the central region. By sacrificing the information of a few points on the tail, we can obtain a much stable higher moment estimate.

2.2.7 *Efficiency of Estimators*

The classical, trimming and quantile based estimates for kurtosis we compared in the previous sections do not assume any parametric form of the underlying distribution. If the underlying follows a certain parametric distribution, kurtosis estimators obtained by maximum likelihood methods will be asymptotically efficient. In this section, we make a small example from the student-t distribution. We will calculate the asymptotic variance of the maximum likelihood estimator (MLE) and compare to the other estimators we have. We will have finite sample approximation and simulation results for comparison, too.

For simplicity we set the in our experiment. Therefore the kurtosis estimates should have finite asymptotic variance. The student-t distribution with the parameter (ν, μ, s) is

$$\begin{aligned} f_0(x) &= \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma\left((\nu+1)/2\right)}{\Gamma\left(\nu/2\right)} \left(1 + x^2/\nu\right)^{-\frac{\nu+1}{2}} \\ f(x) &= f_0\left(\frac{x-\mu}{s}\right) \Big/ s \end{aligned} \tag{2.60}$$

Its information matrix $\mathbf{I} = \mathbf{I}_{\nu,\mu,s}$ is given by (Lange et. al., 1989):

$$\begin{aligned} \mathbf{I} &= \begin{pmatrix} A & 0 & D \\ 0 & B & 0 \\ D & 0 & C \end{pmatrix}. \\ A &= \frac{1}{4}\left[\Omega'\left(\frac{\nu}{2}\right) - \Omega'\left(\frac{\nu+1}{2}\right)\right] - \frac{1}{\nu}\left[\frac{1}{\nu+1} - \frac{1}{2(\nu+3)}\right] \\ B &= \frac{1}{s^2}\left[1 - \frac{2}{\nu+3}\right] \\ C &= \frac{2}{s^2}\frac{\nu}{\nu+3} \\ D &= \frac{1}{s}\left[\frac{1}{\nu+3} - \frac{1}{\nu+1)}\right] \end{aligned} \tag{2.61}$$

The MLE for kurtosis under student-t distribution is

$$K_{MLE} = \frac{6}{\nu - 4}. \tag{2.62}$$

Take the derivative of K_{MLE} with respective to (ν, μ, s), we have

$$K'_{MLE} = \left(0, 0, -\frac{6}{(\nu - 4)^2} \right) \tag{2.63}$$

Therefore by delta method, the asymptotic variance of K_{MLE} is $K'_{MLE} I^{-1} K'_{MLE}{}^{T}$.

For student-t distribution with $\nu = 10$, the asymptotic variance of K_{MLE} is 225.4. This is significantly smaller than the asymptotic variance of the non-parametric kurtosis estimator, which is 720. The MLE is asymptotically efficient and the non-parametric kurtosis estimator has an efficiency of 0.31. According to the similar calculations through the influence functions as in (2.59), the asymptotic variance for the trimmed and quantile based estimators are 7.05 and 21.5, respectively. They are “super-efficient” but they are also biased.

We list the asymptotic variance as well as the finite sample approximation for each method below. We also provide the Monte Carlo simulation results for comparison. The Monte Carlo simulation has 1000 replicates in each experiment with respective sample size. We calculate two kinds of biases. The first bias b_1 is the expected difference between a sample estimate and the true kurtosis under each method. For trimming method, the true kurtosis is -1.5. For quantile method, true kurtosis is -.82. Therefore

$$b_1 = \begin{cases} E\left(K_{trim} + 0.5\right) & \text{trimmed} \\ E\left(K_q - 0.18\right) & \text{quantile} \end{cases} \tag{2.64}$$

The second bias b_2 is the expected difference between the sample estimate and the true kurtosis based on MLE, i.e. $\frac{6}{\nu - 4} = 1$.

Table 2.5 Finite Sample Approximation and Monte Carlo Results: student-t distribution df=10.

		Finite Sample Approximation			Monte Carlo		
	Sample Size	100	250	1000	100	250	1000
Variance	MLE	2.25	0.90	0.23	1.22E+06	3.21E+05	0.29
	Classical	7.20	2.88	0.72	1.08	0.92	0.41
	Trimmed	0.07	0.03	0.01	0.12	0.03	0.01
	Quantile	0.22	0.09	0.02	0.22	0.08	0.02
Bias 2	MLE	0			211	54.57	0.05
	Classical	0			-0.34	-0.12	-0.05
	Trimmed	-1.5			-1.33	-1.45	-1.49
	Quantile	-0.82			-0.8	-0.81	-0.81
Bias 1	Trimmed	0			0.17	0.05	0.01
	Quantile	0			0.02	0.01	0.01

From the table above we observe the following:

1. MLE is a poor estimate when the sample size is small. For sample size 100 and 250, its estimation is basically not useful. At sample size of 1000, it is close to efficient.
2. Classical non-parametric estimates are much better estimates than the MLE when the sample size is small. The small sample size has a “trimming” effect as the data points do not go the tails too far. Therefore, it has much better of variance comparing to the MLE. At higher sample size, the variance becomes

smaller, but not at a rate of $1/n$. This is because as more data points going to the tail, the “trimming effect” is less obvious.

3. The trimming and quantile based robust method both have very good control of the variance, especially the trimming method. Both methods have $1/n$ rate of reduction in their variances as the sample size increases. They are the only methods that the finite sample approximation agrees with the Monte Carlo results even for the small sample sizes.
4. All methods have diminishing biases when the sample size increases. The trimming and quantile methods are consistent in regard with their respective asymptotic results (bias1). They will have an asymptotic bias from the true kurtosis (bias 2). It is notable that bias for trimmed and quantile methods have settled down at very sample sizes. This can also be seen from the distribution of the estimates.
5. In the plot below, we plot the density of all our Monte Carlo results. The estimates from the robust methods have very normal distribution even at small sample sizes, whereas the MLE and the classic methods are not close to symmetric even at the largest sample size. This is another evidence to support using the robust methods.

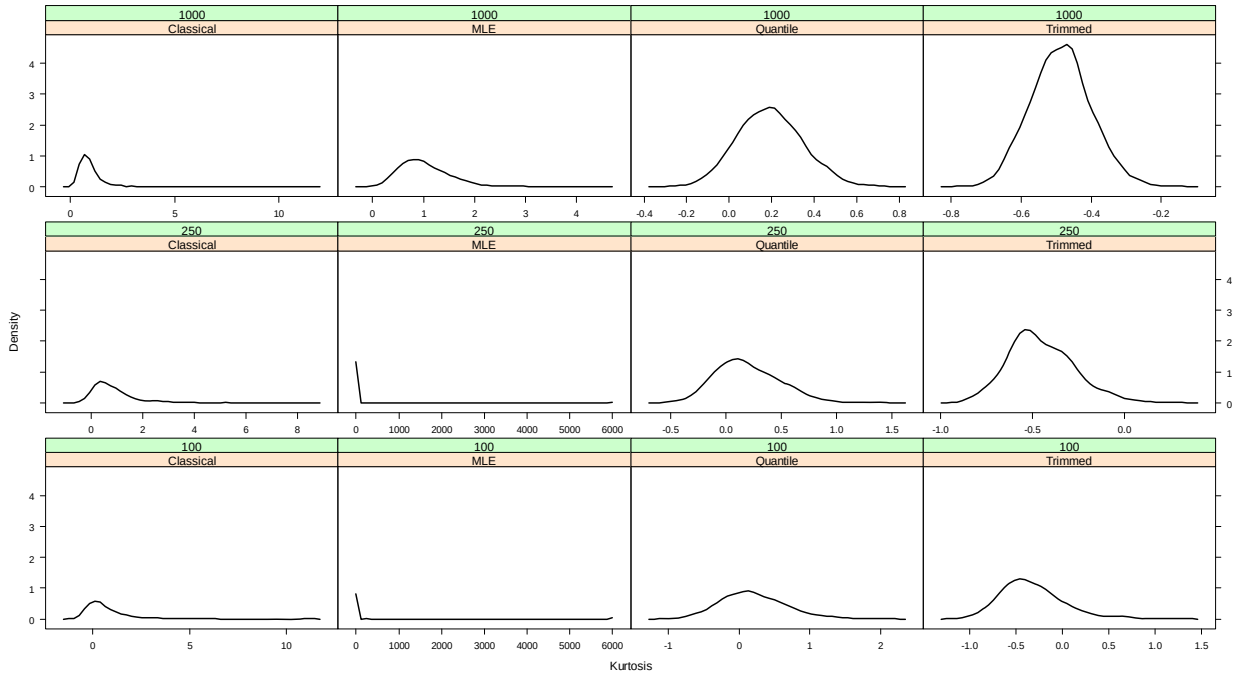

Figure 2.19 Density Plots of Simulation Results from Different Kurtosis Estimators.

2.2.8 Hedge Fund Examples

We now study the effect of robust methods for estimating skewness and kurtosis through some real-world data examples. First we apply these methods to the hedge fund data as described in section 3.4.1. The data set includes 120 monthly returns for 379 hedge funds. Applying the classical, trimming (2.5% on each tail) and quantile ($\alpha = 2.5\%$) methods to each hedge fund, we get the following statistics in table 2.6. We also show the kernel densities as follows. The range of the classical estimates is very wider than those of the robust methods. For skewness, the trimmed skewness estimates have the wider range than the quantile methods. For kurtosis, the difference is marginal.

Table 2.6 Summary Statistics for Hedge Fund Data Using Three Estimation Methods

	Skewness			Kurtosis		
	Classical	Trimmed	Quantile	Classical	Trimmed	Quantile
Min	-8.90	-3.90	-0.74	-0.24	-0.03	-0.38
25%	-1.46	-0.56	-0.19	4.73	0.98	1.16
Median	-0.22	-0.09	-0.03	8.69	1.90	2.32
Mean	-0.24	-0.09	-0.03	15.80	2.60	3.00
75%	1.01	0.40	0.12	20.02	3.36	3.89
Max	9.58	3.13	0.61	97.26	18.80	18.15

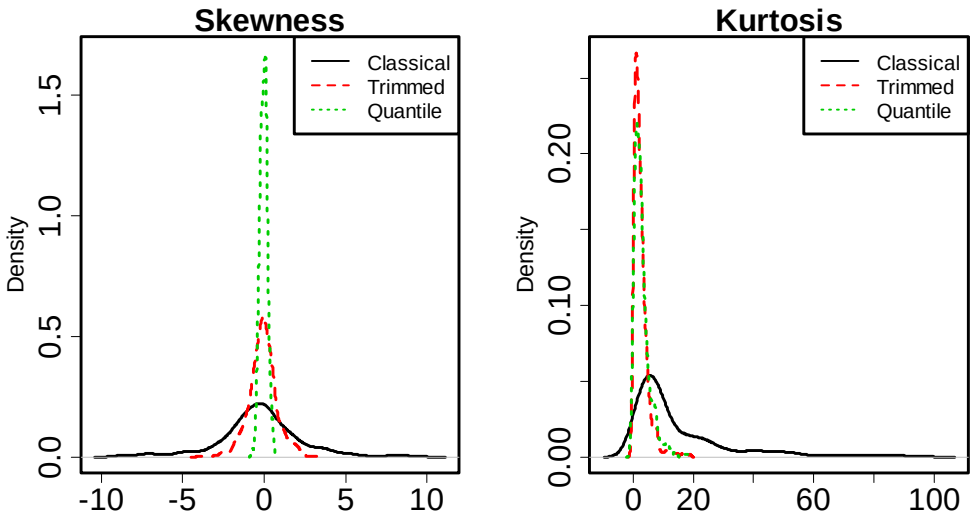

Figure 2.20 Kernel Densities for Hedge Fund Data Using Three Estimation Methods

We now study the effect of robust methods for estimating skewness and kurtosis through some real-world data examples. First we apply these methods to the hedge fund data as described in section 3.4.1. The data set includes 120 monthly returns for 379 hedge funds. Applying the classical, trimming (2.5% on each tail) and quantile ($\alpha = 2.5\%$) methods to each hedge fund, we get the following statistics in table 2.6. We also show the kernel densities as follows. The range of the classical estimates is very wider than those of the robust methods. For skewness, the trimmed skewness estimates have the wider range than the quantile methods. For kurtosis, the difference is marginal.

There is no easy guideline for determining whether a small or large change is preferred. Generally speaking, if the true nominal distribution is close to normal and only displays non-normality because of the influence of an outlier, we'd like to see robust methods give small estimates for skewness and kurtosis and small changes after removing the outlier. This is what happens in the example of S&P 500 in Kim and White (2004). The S&P 500 index shows very little non-normality after removing the 1987 crash. The robust methods tested all give very little change before and after the removing that data point. In our example, the hedge fund sample between the lower and upper quantile shows this type of behavior under the robust methods, too. Both their absolute and relative changes are smaller than the classical estimates. A different scenario is that the true nominal distribution is non-normal, i.e. even after removing a couple of points on the tail, the bulk of the data still shows strong skewness and kurtosis. In some cases, the much larger impact brought forth by the extreme tail data masked the skewness and kurtosis represented by the bulk data. Consequently, the whole sample shows very little skewness and kurtosis or the opposite direction of skewness and kurtosis of that represented by the bulk data. For such samples, removing the extreme tail data may reveal the true skewness and kurtosis and therefore a large relative change in estimates is justified. This may be the cause for large relative changes for some robust estimates.

We move on to remove three maximum and three minimum returns from each hedge fund and repeat the exercise. The table below shows the difference between the new estimate and the original estimate. The percentage change is shown in the parentheses. Besides the same conclusion we find above, we find that the relative changes are larger than those when we only removed one pair of extreme returns across the board. This tells us that removing too many points may have a material change in the nominal distribution.

Table 2.7 Summary Statistics for Hedge Fund Data Using Three Estimation Methods: After removing one pair of extreme returns.

	Skewness			Kurtosis		
	Classical	**Trimmed**	**Quantile**	**Classical**	**Trimmed**	**Quantile**
Min	-10.04 (-10070%)	-1.17 (-1860%)	-0.27 (-14050%)	-87.61 (-752%)	-9.14 (-3347%)	-4.81 (-306%)
25%	-0.39 (-87%)	-0.06 (-30%)	-0.04 (-36%)	-10.85 (-81%)	-0.64 (-35%)	-0.9 (-34%)
Median	0.07 (-41%)	0.01 (-9%)	0 (-8%)	-3.08 (-53%)	-0.27 (-21%)	-0.41 (-21%)
Mean	0.07 (-84%)	0.03 (-14%)	0 (-97%)	-10.33 (-54%)	-0.56 (-31%)	-0.65 (-27%)
75%	0.61 (-6%)	0.11 (4%)	0.05 (18%)	-0.97 (-26%)	-0.13 (-10%)	-0.17 (-10%)
Max	8.67 (584%)	1.47 (5113%)	0.28 (587%)	13.91 (92%)	2.17 (689%)	0.4 (427%)

Table 2.8 Summary Statistics for Hedge Fund Data Using Three Estimation Methods: After removing three pairs of extreme returns.

	Skewness			Kurtosis		
	Classical	**Trimmed**	**Quantile**	**Classical**	**Trimmed**	**Quantile**
Min	-9.47 (-5074%)	-2.16 (-2279%)	-0.43 (-26280%)	-90.3 (-1020%)	-13.99 (-9423%)	-10.64 (-440%)
25%	-0.75 (-105%)	-0.11 (-64%)	-0.06 (-75%)	-17.89 (-98%)	-1.54 (-71%)	-1.88 (-72%)
Median	0.14 (-77%)	0.05 (-29%)	0.02 (-22%)	-6.87 (-90%)	-0.77 (-52%)	-1.14 (-54%)
Mean	0.15 (-99%)	0.08 (6%)	0.02 (-90%)	-13.94 (-88%)	-1.31 (-75%)	-1.5 (-56%)
75%	1.01 (-33%)	0.24 (0%)	0.11 (23%)	-2.97 (-74%)	-0.36 (-33%)	-0.6 (-38%)
Max	8.71 (848%)	2.23 (13280%)	0.43 (7373%)	1.95 (265%)	0.86 (993%)	0.52 (959%)

Chapter 3
Portfolio Optimization with Modified VaR

Value-at-Risk (VaR) as a popular measure among practitioners is the negative quantile on the downside of the return distribution. It is defined as

$$VaR_r\left(1-\alpha\right) = -\inf_{r_0}\left\{P\left(r \le r_0\right) \ge \alpha\right\} \tag{3.1}$$

for confidence level $1-\alpha$ or equivalent tail probability α . The importance of VaR is well established for the purposes of risk management in the banking industry in the 1990s. It tells when the worst few scenarios happen, what is the least amount loss the investor should expect. It is a downside risk measure and does not depend on normality. Although under normality, VaR is can be easily represented by

$$VaR_r\left(1-\alpha\right) = -\left(\mu + z\left(\alpha\right)\sigma\right) \tag{3.2}$$

where $z\left(\alpha\right)$ is the α quantile on the standard normal distribution. Its sample estimate can conveniently calculated by plugging in the sample estimate $\hat{\mu}$ and $\hat{\sigma}$ into (3.2).

VaR as a quantile measure does not consider any losses above the threshold. This feature makes it not a coherent measure. Although there's no simple method to make VaR coherent, it is possible to estimate VaR using the entire distribution instead of a simple quantile. The idea is to first build on quantiles of the normal distribution like (3.2) and compensate slight departures from normality by making adjustments using higher moments. There has also been a recent proposal (Jaschke, 2001) to use skewness and kurtosis in a Cornish-Fisher (CF) expansion (1938 and 1960) to improve the approximation of VaR for use in portfolio optimization. The general format of the CF expansion is

$$
\begin{aligned}
&MVaR_r\left(1-\alpha\right) = -\left(\mu + \sigma \cdot \tilde{q}\left(\alpha\right)\right) \\
&\tilde{q}\left(\alpha\right) \triangleq z\left(\alpha\right) + \frac{1}{6}\left[z^2\left(\alpha\right)-1\right]\cdot S + \\
&\qquad \frac{1}{24}\left[z^3\left(\alpha\right)-3z\left(\alpha\right)\right]\cdot K - \frac{1}{36}\left[2z^3\left(\alpha\right)-5z\left(\alpha\right)\right]\cdot S^2
\end{aligned}
\qquad . \qquad (3.3)
$$

where S and K are the skewness and kurtosis of the underlying distribution. Its sample estimate can be conveniently calculated by plugging in the sample estimate $\hat{\mu}$, $\hat{\sigma}$, $\hat{S}$ and $\hat{K}$ into (3.3), i.e.

$$
\begin{aligned}
&\widehat{MVaR_r\left(1-\alpha\right)} = -\left(\hat{\mu} + \hat{\sigma} \cdot \hat{q}\left(\alpha\right)\right) \\
&\hat{q}\left(\alpha\right) \triangleq z\left(\alpha\right) + \frac{1}{6}\left[z^2\left(\alpha\right)-1\right]\cdot \hat{S} + \\
&\qquad \frac{1}{24}\left[z^3\left(\alpha\right)-3z\left(\alpha\right)\right]\cdot \hat{K} - \frac{1}{36}\left[2z^3\left(\alpha\right)-5z\left(\alpha\right)\right]\cdot \hat{S}^2
\end{aligned}
\qquad (3.4)
$$

One advantage this format is it is a smoother function than VaR. It also gives us a convenient tool so that we can easily estimate VaR without having to specify the exact distribution of historical returns and sufficiently characterize its non-normality. From a practical point of view the use of the first four moments is the most workable. Since higher moments are quite vulnerable to outliers, appropriate robust methods should prove to be helpful. Jaschke (2001) discussed the behavior of CF expansion under specific distribution families. Martellini, Vaissié and Ziemann (2005) tested the idea of modifying VaR and using structured comoments matrix when applying Cornish-Fisher (CF) expansion. In industry, the MVaR has been developed by AlternativeSoft (see Favre and Galeano, 2002). They found the skewness and the kurtosis effect is high if the VaR is computed at 99%.

3.1 Modified VaR Influence Functions

We now build on our results in Chapter 2 on the influence functions and variance of classical and trimmed skewness and kurtosis estimates by deriving the influence functions for MVaR based on these estimates. The main reason for choosing the trimmed robust estimates over the quantile based robust estimates is its better control of sample estimate variances and preservation of influence function structure inside the trimmed region.

For simplicity, we first focus on the part of MVaR that only involves skewness and kurtosis, namely:

$$\begin{aligned}\hat{q}(\alpha) \triangleq\ & z(\alpha)+\frac{1}{6}\left[z^2(\alpha)-1\right]\cdot\hat{S}+ \\ & \frac{1}{24}\left[z^3(\alpha)-3z(\alpha)\right]\cdot\hat{K}-\frac{1}{36}\left[2z^3(\alpha)-5z(\alpha)\right]\cdot\hat{S}^2\end{aligned} \tag{3.5}$$

For simplicity, we will omit the hat sign in the following text. It is straightforward to compute the influence function for $-\hat{q}(\alpha)$ based on the influence functions $\mathrm{IF}(r;S,F)$ and $\mathrm{IF}(r;K,F)$ of skewness and kurtosis:

$$\begin{aligned}&\mathrm{IF}(r;-\hat{q},F) \\ &=-\frac{1}{6}\left[z^2(\alpha)-1\right]\cdot\mathrm{IF}(r;S,F)-\frac{1}{24}\left[z^3(\alpha)-3z(\alpha)\right]\cdot\mathrm{IF}(r;K,F) \\ &\qquad+\frac{1}{18}\left[2z^3(\alpha)-5z(\alpha)\right]\cdot S\cdot\mathrm{IF}(r;S,F).\end{aligned} \tag{3.6}$$

Therefore the influence function of MVaR is

$$\begin{aligned}&\mathrm{IF}\left(r;MVaR_{1-\alpha},F\right)\\&=-\mathrm{IF}\left(r;\mu,F\right)+\sigma\cdot\mathrm{IF}\left(r;-q,F\right)-\mathrm{IF}\left(r;\sigma,F\right)\cdot q,\end{aligned}\tag{3.7}$$

where (See Maronna et. al. 2006).

$$\begin{aligned}\mathrm{IF}\left(r;\mu,F\right)&=r-\mu\\\mathrm{IF}\left(r;\sigma,F\right)&=\frac{1}{2\sigma}\left(\left(r-\mu\right)^2-\sigma^2\right).\end{aligned}\tag{3.8}$$

Plugging (3.8) and (3.6) into (3.7), we get the complete influence function for MVaR:

$$\begin{aligned}&\mathrm{IF}\left(r;MVaR_{1-\alpha},F\right)\\&=-\left(r-\mu\right)+\sigma\cdot\left\{-\frac{1}{6}\left[z^2\left(\alpha\right)-1\right]\cdot\mathrm{IF}\left(r;S,F\right)\right.\\&\qquad-\frac{1}{24}\left[z^3\left(\alpha\right)-3z\left(\alpha\right)\right]\cdot\mathrm{IF}\left(r;K,F\right)\\&\qquad\left.+\frac{1}{18}\left[2z^3\left(\alpha\right)-5z\left(\alpha\right)\right]\cdot S\cdot\mathrm{IF}\left(r;S,F\right)\right\}\\&\quad-\frac{1}{2\sigma}\left(\left(r-\mu\right)^2-\sigma^2\right)\cdot\tilde{q}\end{aligned}\tag{3.9}$$

We display the influence functions for classical and trimmed MVaR (TMVaR) along with the influence functions for VaR (see (3.1)). The latter was given in Chapter 2 (see also Maronna et. al., 2006):

$$\mathrm{IF}\left(r;VaR_{1-\alpha},F\right)=\frac{I_{0,\alpha}\left(r\right)-\alpha}{f_R\left(q\left(\alpha\right)\right)}.\tag{3.10}$$

In the following plots, we show trimming done at the 2.5%, 1% and 0.5% levels (at each tail) and VaR is at the 97.5% confidence level. We plot the influence functions of the VaR's under different distributions.

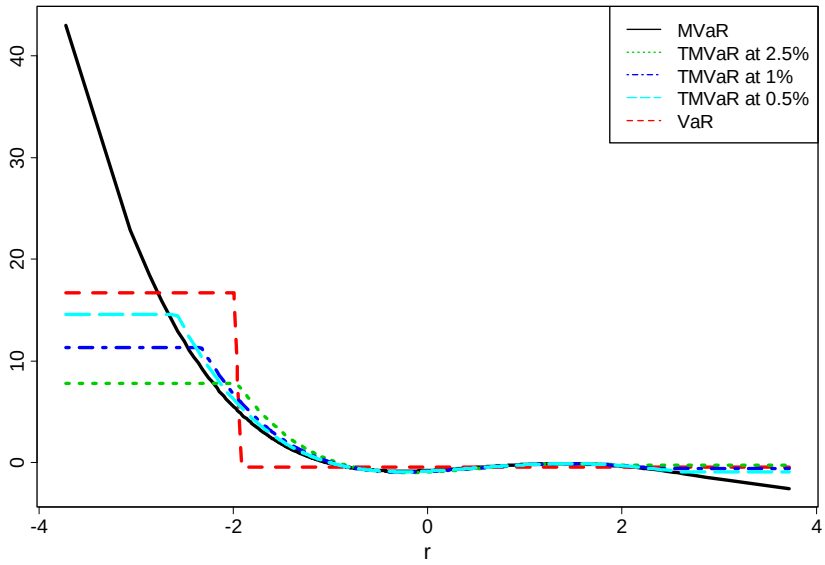

Figure 3.1 Influence Functions of MVaR and TMVaR compared to VaR (red dashed lines) under standard normal distribution.

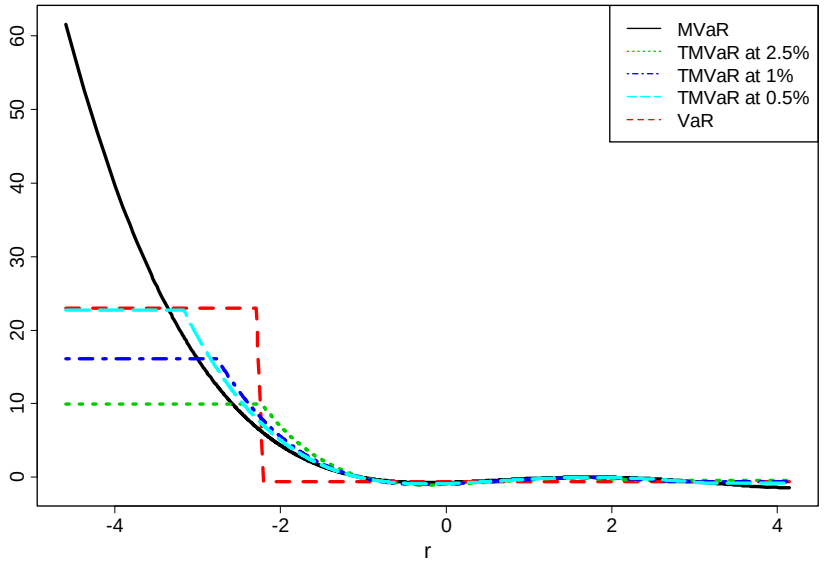

Figure 3.2 Influence Functions of MVaR and TMVaR compared to VaR (red dashed lines) under Student-t distribution with $df = 10$.

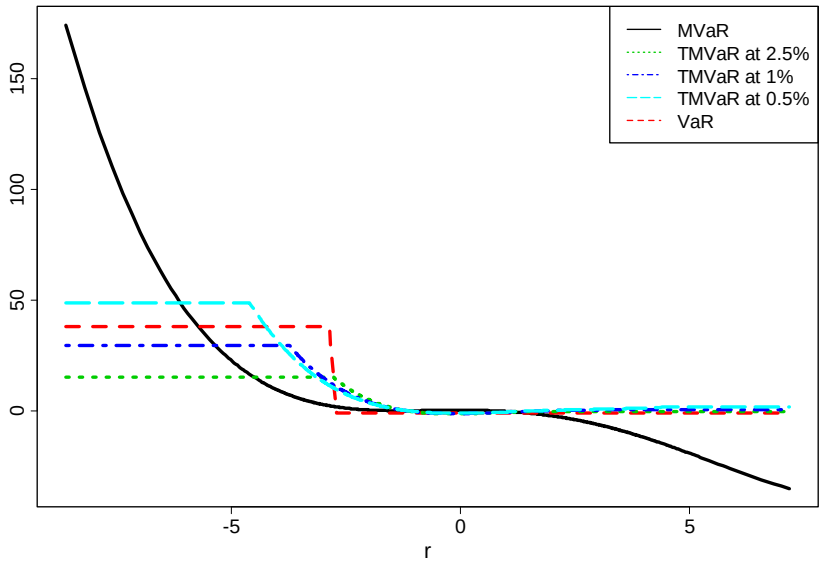

Figure 3.3 Influence Functions of MVaR and TMVaR compared to VaR (red dashed lines) under Student-t distribution with $df = 4$.

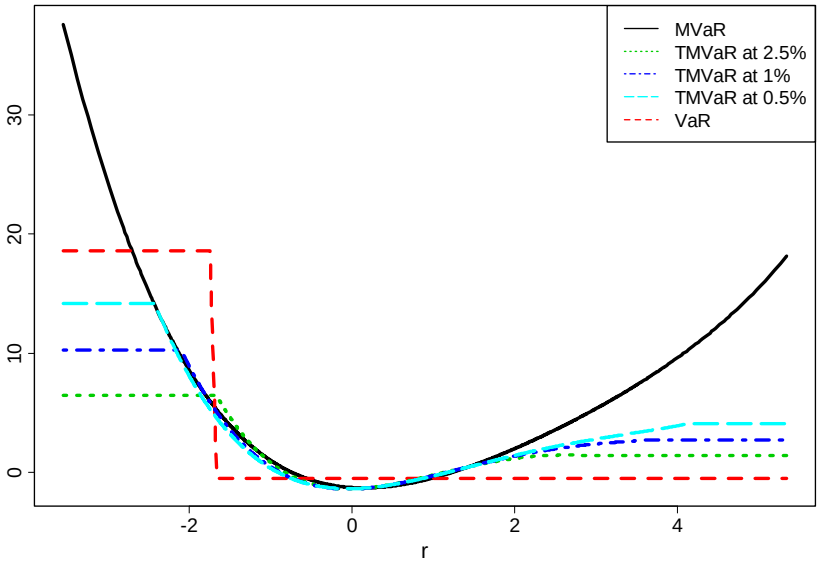

Figure 3.4 Influence Functions of MVaR and TMVaR compared to VaR (red dashed lines) under skewed-t distribution[2] with $df = 10, \gamma = 1.25$.

[2] The skewed-t distribution we use here is defined by

$$f(x;n,\gamma) = \begin{cases} 2/(\gamma + 1/\gamma)\cdot t(\gamma \cdot x;n) & \text{for } x < 0 \\ 2/(\gamma + 1/\gamma)\cdot t(1/\gamma \cdot x;n) & \text{for } x \geq 0 \end{cases}$$

where $t(\cdot;n)$ is the density of student-t distribution with *n* degrees of freedom.

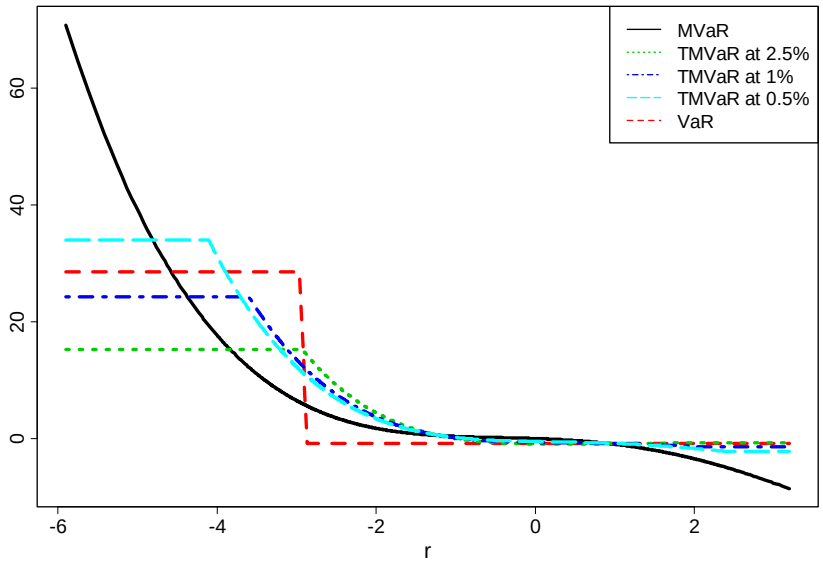

Figure 3.5 Influence Functions of MVaR and TMVaR compared to VaR (red dashed lines) under skewed-t distribution with $df = 10, \gamma = 0.8$.

Ideally increases in a return in the right tail should not affect value-at risk type measures such as VaR, MVaR and TMVaR, and with this in mind one observes the following behaviors of these three risk measures:

1. The influence function of VaR is constant for values less than the value of VaR without the influential return, and is essentially zero for larger values. Thus any return in the left tail has the same influence on VaR, no matter how large the loss, which is not a desirable property of a risk measure. On the other hand any return in the right tail has negligible impact, which is quite reasonable.
2. MVaR is a function of skewness and kurtosis, both of which are affected by a return value in a non-constant manner (see the plots of skewness and kurtosis influence functions in chapter 2). Thus the influence function of MVaR increases as the value of a return moves farther out in the left tail, while the opposite happens as a return moves farther out in the right tail.

3. TMVaR has a bounded influence function, which is considered to be a desirable property of a robust estimate. In the case of a positive skewed distribution with substantial skewness (Fig. 3.4), MVaR increases rapidly for positive returns, which is not desirable. We consider this is the case when the nominal distribution is far from normal and the outlier adds to this tendency, the CF-expansion actually breaks down. On the other hand, if the nominal distribution is left skewed (Fig. 3.5), the outlier on the left will make the MVaR much larger than it needs to be. However, this is not entirely "wrong". TMVaR has a more desirable behavior by virtue of being bounded closer to zero for all positive returns, which shows us a good reason to trim positive outliers as well as negative outliers.
4. The plots above are showing the influence functions in the range from 0.1% to 99.9% percentile of the respective distributions. Beyond the point of 99.9% percentile, there is actually more structure due to the contribution of the influence function for excess kurtosis, e.g. for student-t $df = 4$

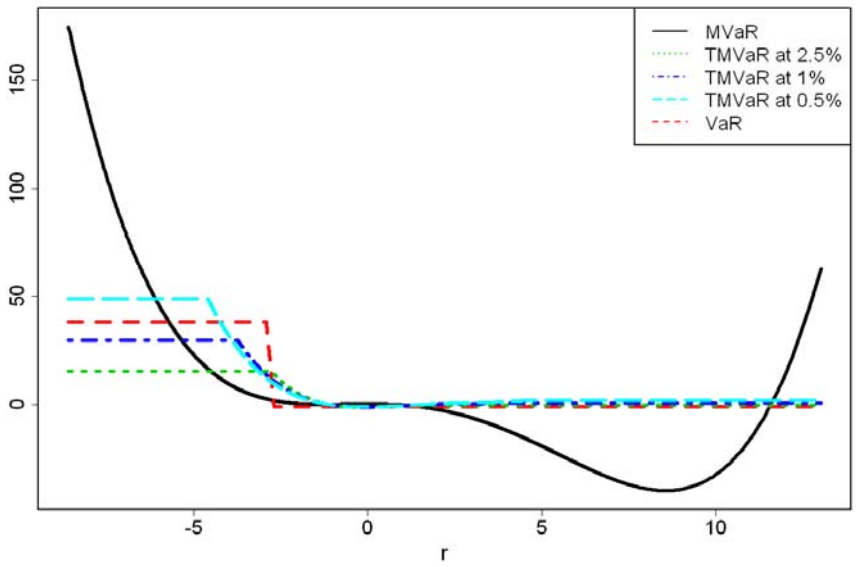

Figure 3.6 Influence Functions of MVaR , TMVaR and VaR. Interval is from 0.05% to 99.99% quantiles under Student-t distribution with $df = 4$.

It is also common in all distributions. For example, if we take the standard normal distribution, we have

$$\begin{aligned}&\mathrm{IF}\left(r; MVaR_{1-\alpha}, F\right)\\&= -r - \frac{z(\alpha)}{2} r^2 - \frac{1}{6}\left[z^2(\alpha) - 1\right]\cdot\left(r^3 - 3r\right) - \frac{1}{24}\left[z^3(\alpha) - 3z(\alpha)\right]\cdot\left(r^4 - 6r^2 + 3\right)\end{aligned}$$

which is a 4th order polynomial. Note $z^3(\alpha) - 3z(\alpha)$ is negative for small α . Therefore the influence function goes to positive infinity on both tails. This is the same for all distributions. The trimmed version completely avoided this flaw.

3.2 Finite Sample Variance

We now compare finite sample Monte Carlo estimates with asymptotic approximations to the finite sample standard deviations of VaR, MVaR and TMVaR under the standard normal distribution.

The asymptotic approximations are obtained by calculating the asymptotic standard deviations using the influence function approach $Var_\infty(\theta) = E\left(IF^2(x;\theta,F)\right)$ (see for example, Hampel et. al., 1986). The influence functions are computed using equation (3.9) and then numerically integrated using *R* routine integrate, which is based on QUADPACK (see Piessens et al. , 1983) to get the asymptotic variance. In the case of student-t distribution with $df = 4$, we limit the integration interval from 0.001% to 99.999% quantiles of the distribution, since the true asymptotic variance is infinite. We set this range to appreciate the magnitude of difference between MVaR and TMVaR. The finite sample standard deviation approximation is obtained by using $\sqrt{Var_\infty(\theta)/n}$, where n is

the sample size. The Monte Carlo simulation has 10,000 replicates (labeled “Simulation”) under each sample size.

The Table 3.1 below shows the results. Generally the asymptotic approximations agree well with the simulation results, with the main differences occurring in the small sample size cases (e.g. 20, 30 and 60). These are sample sizes (e.g. sample size of 20 and 30) where there are no data points to be trimmed at 2.5% trimming, hence making the trimmed and classical results the same.

Table 3.1 Standard Deviations of Sample MVaR, TMVaR and VaR under standard normal distribution.

Standard deviations of Sample VaR						
	Asymptotic Approximation			*Simulation*		
Sample Size	*MVaR*	*TMVaR*	*VaR*	*MVaR*	*TMVaR*	*VaR*
20	0.4645 (0.0236)	0.3975 (0.0069)	0.5972 (0.0182)	0.4251 (0.003)	0.4251 (0.003)	0.4152 (0.0029)
30	0.3792 (0.0192)	0.3246 (0.0056)	0.4876 (0.0148)	0.3597 (0.0026)	0.3597 (0.0026)	0.3699 (0.0028)
60	0.2682 (0.0136)	0.2295 (0.004)	0.3448 (0.0105)	0.2638 (0.0019)	0.2367 (0.0017)	0.2947 (0.0022)
120	0.1896 (0.0096)	0.1623 (0.0028)	0.2438 (0.0074)	0.1894 (0.0014)	0.1608 (0.0011)	0.2287 (0.0017)
250	0.1314 (0.0067)	0.1124 (0.0019)	0.1689 (0.0051)	0.133 (0.001)	0.1124 (8e-04)	0.1627 (0.0012)

The numbers in the parentheses are the standard deviations of the respective estimates. They are obtained using the following method. For asymptotic approximations, if we denote the sample based estimate of asymptotic variance $Var_{\infty}(\theta)$ by $S_{\hat{\theta}}^2$ and $M = 10,000$ is the number of simulation replicates. Denote $\mu_{IF,2} = E\left(IF^2\left(x;\theta,F\right)\right)$ and $\mu_{IF,4} = E\left(IF^4\left(x;\theta,F\right)\right)$, by central limit theorem, we have

$$\sqrt{M}\left(S_{\hat{\theta}}^2 - Var_{\infty}(\theta)\right) \xrightarrow[d]{} N\left(0, \mu_{IF,4} - {\mu_{IF,2}}^2\right)$$

By delta method, we have

$$\sqrt{M}\left(S_{\hat{\theta}} - \sqrt{Var_{\infty}(\theta)}\right) \xrightarrow[d]{} N\left(0, \frac{1}{4\mu_{IF,2}}\left(\mu_{IF,4} - {\mu_{IF,2}}^2\right)\right)$$

We have the asymptotic variance of $S_{\hat{\theta}}$ is $\frac{1}{4\mu_{IF,2}}\left(\mu_{IF,4} - {\mu_{IF,2}}^2\right)$. Therefore, we can approximate the standard deviation of $S_{\hat{\theta}}$ by $\frac{1}{\sqrt{M}}\frac{1}{2\sqrt{\hat{\mu}_{IF,2}}}\sqrt{\hat{\mu}_{IF,4} - {\hat{\mu}_{IF,2}}^2}$. Since the finite sample standard deviation is $\frac{1}{\sqrt{n}}S_{\hat{\theta}}$, where n is the respective sample size, we have the standard deviation of the estimate: $\frac{1}{\sqrt{n}}\frac{1}{\sqrt{M}}\frac{1}{2\sqrt{\hat{\mu}_{IF,2}}}\sqrt{\hat{\mu}_{IF,4} - {\hat{\mu}_{IF,2}}^2}$.

For Monte Carlo results, the similar procedure can be carried out by replacing $\mu_{IF,2}$ and $\mu_{IF,4}$ with the second and fourth moments of the respective Monte Carlo estimates.

We also reproduce Table 3.1 under the following fat-tailed distributions.

Student-t with $df = 4$

Table 3.2 Standard Deviations of Sample MVaR, TMVaR and VaR under Student-t with $df = 4$ **.**

Standard deviations of Sample VaR						
	Asymptotic Approximation			*Simulation*		
Sample Size	*MVaR*	*TMVaR*	*VaR*	*MVaR*	*TMVaR*	*VaR*
20	5.5051 (4.2906)	0.7021 (0.0144)	1.3645 (0.0415)	1.1828 (0.0216)	1.1828 (0.0216)	1.0344 (0.0182)
30	4.4949 (3.5033)	0.5733 (0.0117)	1.1141 (0.0339)	1.1606 (0.0226)	1.1606 (0.0226)	0.8792 (0.0147)
60	3.1784 (2.4772)	0.4054 (0.0083)	0.7878 (0.024)	1.3445 (0.1396)	0.5095 (0.0057)	0.6868 (0.0078)
120	2.2475 (1.7516)	0.2866 (0.0059)	0.5571 (0.017)	1.4968 (0.2205)	0.2933 (0.0023)	0.5131 (0.0046)
250	1.5571 (1.2136)	0.1986 (0.0041)	0.3859 (0.0117)	1.3576 (0.0851)	0.2031 (0.0015)	0.3701 (0.003)

Student-t with $df = 10$

Table 3.3 Standard Deviations of Sample MVaR, TMVaR and VaR under Student-t with $df = 10$

Standard deviations of Sample VaR						
	Asymptotic Approximation			*Simulation*		
Sample Size	*MVaR*	*TMVaR*	*VaR*	*MVaR*	*TMVaR*	*VaR*
20	0.8191 (0.1188)	0.4865 (0.009)	0.8235 (0.0251)	0.6035 (0.0063)	0.6035 (0.0063)	0.5782 (0.0056)
30	0.6688 (0.097)	0.3973 (0.0074)	0.6724 (0.0205)	0.5246 (0.0052)	0.5246 (0.0052)	0.5059 (0.0045)
60	0.4729 (0.0686)	0.2809 (0.0052)	0.4754 (0.0145)	0.4174 (0.0041)	0.3049 (0.0023)	0.4022 (0.0032)
120	0.3344 (0.0485)	0.1986 (0.0037)	0.3362 (0.0102)	0.3236 (0.0042)	0.1988 (0.0014)	0.3134 (0.0025)
250	0.2317 (0.0336)	0.1376 (0.0026)	0.2329 (0.0071)	0.2385 (0.0029)	0.1387 (0.001)	0.2263 (0.0017)

Skewed-t distribution with $df = 10, \gamma = 1.25$

Table 3.4 Standard Deviations of Sample MVaR, TMVaR and VaR under Skewed-t distribution with $df = 10, \gamma = 1.25$**.**

Standard deviations of Sample VaR						
	Asymptotic Approximation			*Simulation*		
Sample Size	*MVaR*	*TMVaR*	*VaR*	*MVaR*	*TMVaR*	*VaR*
20	0.6974 (0.0435)	0.3625 (0.005)	0.6655 (0.0203)	0.4998 (0.0048)	0.4998 (0.0048)	0.4727 (0.0043)
30	0.5694 (0.0355)	0.2959 (0.0041)	0.5434 (0.0165)	0.4441 (0.0045)	0.4441 (0.0045)	0.4145 (0.0036)
60	0.4026 (0.0251)	0.2093 (0.0029)	0.3842 (0.0117)	0.3448 (0.0043)	0.2497 (0.002)	0.3267 (0.0028)
120	0.2847 (0.0178)	0.148 (0.002)	0.2717 (0.0083)	0.2875 (0.0227)	0.1655 (0.0012)	0.2556 (0.0021)
250	0.1972 (0.0123)	0.1025 (0.0014)	0.1882 (0.0057)	0.1874 (0.0024)	0.1146 (8e-04)	0.181 (0.0014)

We also compared the sample standard deviations of the trimmed TMVaR and VaR to MVaR. The following table shows the results under Monte Carlo simulations with the sample size of 120. It shows that the VaR has similar standard deviation as the MVaR in general, except the case of the very heavy tailed distribution, i.e. student-t with $df = 4$, where VaR has only 1/3 of the standard deviation. The TMVaR unequivocally reduces the sample standard deviation and does the best job in the heavy tailed case.

Table 3.5 Standard Deviations of TMVaR and VaR as a Percentage of Standard Deviation of MVaR.

	TMVaR/MVaR	*VaR/MVaR*
Standard Normal	85%	121%
Student-t, $df = 4$	20%	34%
Student-t, $df = 10$	61%	97%
Skewed-t $df = 10, \gamma = 1.25$	57%	89%

Conclusions

Our influence functions analysis reveals that MVaR has a quite different behavior from that of VaR. The MVaR influence function has a highly non-linear influence unbounded behavior. Because the influence function of TMVaR is bounded (i.e., a bounded influence statistic) it is not much influenced by outliers. Consequently TMVaR avoids the break-down of the CF expansion approximation when departure from normality is substantial.

In the following sections, we will first discuss portfolio optimization experiments in general. We will discuss the overall set-up of the experiment, the empirical definitions of each risk measure and the numerical optimization methods in 3.3. In section 3.4, we will discuss the data used in our experiments. In section 3.5, we will show the optimization results of each strategy under different constraints.

3.3 Portfolio Optimization with Modified VaR

3.3.1 The Basic Portfolio Optimization Problem

Markowitz's mean-risk paradigm maximizes mean return while controlling for the risk. It is well known that mean returns are notoriously hard to estimate. The historical mean returns are not very good predictors of the future mean. The purpose of this empirical study is to compare risk measures in their original form. In order to avoid the impact brought by the mean return in portfolio optimization, we limit our experiments to the global minimum risk portfolios for each risk measure.

Minimum Risk Optimization

Consider the following one-period portfolio optimization problem: given N assets in the portfolio, fully invested and there is no risk-free asset, which method of allocating investments gives the smallest risk? Additional constraint may include long-only constraint or an upper bounded constraint that caps and floors allocation on each investment.

We formulate this problem using the following schemes

$$\begin{aligned} &\text{minimize} && Risk_p \\ &\text{s.t.} && \sum_{i=1}^{N} w_i = 1 \\ & && c \le w_i \le u \qquad \forall i \in 1 \ldots N \end{aligned} \tag{3.11}$$

where $Risk_p$ is the portfolio risk and w_i is the weight for i th asset. The risk measures we compare in our experiments include volatility, VaR, MVaR and TMVaR. The volatility measure is the only risk measure in our group that is not a tail risk measure. It serves as a comparison from the traditional Markowitz framework. In addition to these 5 global minimum risk portfolios, we add the maximum Sharpe ratio portfolio (the tangency portfolio) into the horse race as another comparison. Its optimization is slightly different as shown below

$$\begin{aligned} &\text{Minimize} && \mu_p - \frac{1}{2}\lambda\sigma_p^2 \qquad \lambda \ge 0 \\ &\text{s.t.} && \sum_{i=1}^{N} w_i = 1 \\ & && c \le w_i \le u \qquad \forall i \in 1 \ldots N \end{aligned} \tag{3.12}$$

3.3.2 Empirical Definitions of Risk Measures

In the following definitions, T is the total number of scenarios and r_i is the portfolio return in the i th period.

Volatility

The empirical estimate of variance is defined as follows

$$\widehat{\sigma} \triangleq \sqrt{\frac{1}{T-1}\sum_{i=1}^{T}\left(r_i-\hat{\mu}\right)^2} \tag{3.13}$$

where $\hat{\mu}=\frac{1}{T}\sum_{i=1}^{T} r_i$.

VaR (Value at Risk)

The empirical estimate of VaR at the α confidence level (In our experiments, $\alpha=0.95$) is defined as follows

$$\widehat{VaR_{1-\alpha}(r)} \triangleq -q_{1-\alpha}(r), \tag{3.14}$$

where $q_\alpha(\cdot)$ is the quantile function. We use the following working definition that uses the order statistics to replace the quantile function. Let $r_{(1)},\ldots r_{(T)}$ be the order statistics (from smallest to largest) of $r_1,\ldots r_T$, we define

$$q_\alpha(r) \triangleq (1-\lambda) r_{(j)} + \lambda r_{(j+1)} \tag{3.15}$$

where j satisfies $\frac{j-0.5}{T} \le \alpha < \frac{j+0.5}{T}$ and λ equals to the fractional part of $T\alpha-(j-0.5)$ (see Hyndman, R. J. and Fan, Y. (1996)).

MVaR

The empirical estimate of CF-Modified VaR at the α confidence level is defined using (3.4) and plug in the following estimates for sample skewness and kurtosis

$$\hat{S} = \frac{\sqrt{T}\cdot\sum_{i=1}^{T}\left(r_i-\hat{\mu}\right)^3}{\left(\sum_{i=1}^{T}\left(r_i-\hat{\mu}\right)^2\right)^{3/2}} \qquad \hat{K} = \frac{T\cdot\sum_{i=1}^{T}\left(r_i-\hat{\mu}\right)^4}{\left(\sum_{i=1}^{T}\left(r_i-\hat{\mu}\right)^2\right)^{2}} - 3 \tag{3.16}$$

Also in (3.4) we use $z_\alpha = \Phi^{-1}(\alpha)$, where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. In our experiments, $\alpha = 0.95$.

TMVaR

The empirical estimate of the TMVaR uses the following trimmed sample skewness and kurtosis,

$$\widehat{S_{trim}} = \frac{\sqrt{\lfloor T\cdot(1-2\alpha)\rfloor}\cdot\sum_{i=\lceil T\cdot\alpha\rceil}^{\lfloor T\cdot(1-\alpha)\rfloor}\left(r_{(i)}-\hat{\mu}\right)^3}{\left(\sum_{i=\lceil T\cdot\alpha\rceil}^{\lfloor T\cdot(1-\alpha)\rfloor}\left(r_{(i)}-\hat{\mu}\right)^2\right)^{3/2}} \tag{3.17}$$

and

$$\widehat{K_{trim}} = \frac{\lfloor T\cdot(1-2\alpha)\rfloor\cdot\sum_{i=\lceil T\cdot\alpha\rceil}^{\lfloor T\cdot(1-\alpha)\rfloor}\left(r_{(i)}-\hat{\mu}\right)^4}{\left(\sum_{i=\lceil T\cdot\alpha\rceil}^{\lfloor T\cdot(1-\alpha)\rfloor}\left(r_{(i)}-\hat{\mu}\right)^2\right)^{2}} \tag{3.18}$$

where $\lfloor\cdot\rfloor$ is the flooring function and $\lceil\cdot\rceil$ is the ceiling function. In this definition we trim α on both tails.

3.3.3 Optimization Methods

The computational challenge of VaR type portfolio optimization is the non-convexity and non-smoothness of the objective function. Since VaR is a quantile based risk measure, it often has numerous local minima as a function of the portfolio weight vector.

The following plot is an example of VaR from a very simple portfolio. This portfolio consist weekly returns of two stocks “Microsoft” and “Boeing”. As fraction of “Microsoft” varies from 0 to 1 in the portfolio, we can clearly see how the empirical 1% VaR (solid line) goes on a bumpy trip with two local minima. The CF expansion used in our optimization is not only a parametric estimation method for VaR, it also serves as an objective function smoother. As we see in Fig. 3.7, the CF expansion provides a very smooth fit to the empirical VaR.

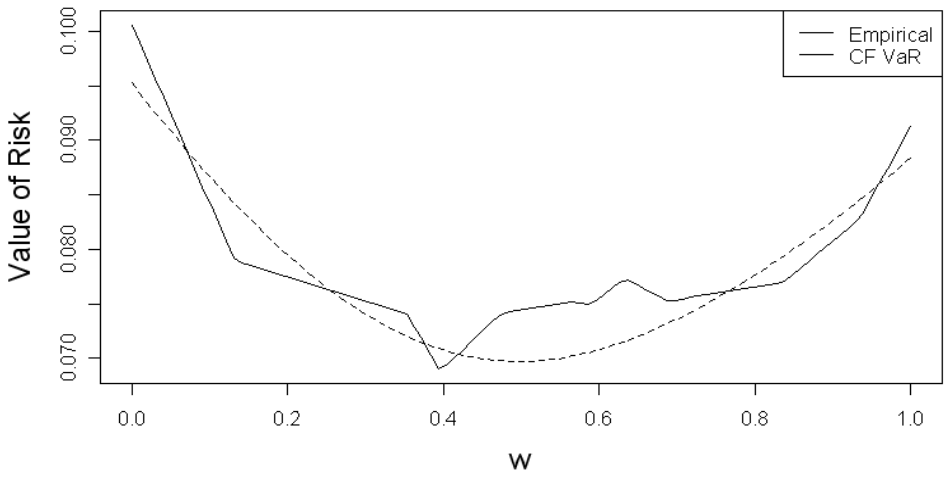

Figure 3.7 VaR of a Two-Stock Portfolio

However, having a single global minimum for the MVaR representation is rare in more general situations with three or more stocks in the portfolio. We construct an example below to show how this is possible. The following is a simulation of three standard normal variables with the correlation matrix

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & -.3 & .4 \\ -.3 & 1 & .2 \\ .4 & .2 & 1 \end{pmatrix}.$$

We then contaminate the data with 5% of correlated negative outliers, as shown in the scatter plots upper panel of Fig 3.8. The MVaR, expressed as a function of two portfolio weights (the sum of the three weights is 1), is shown in the contour plot on the lower panel of Fig. 3.8. It clearly shows two local minimum.

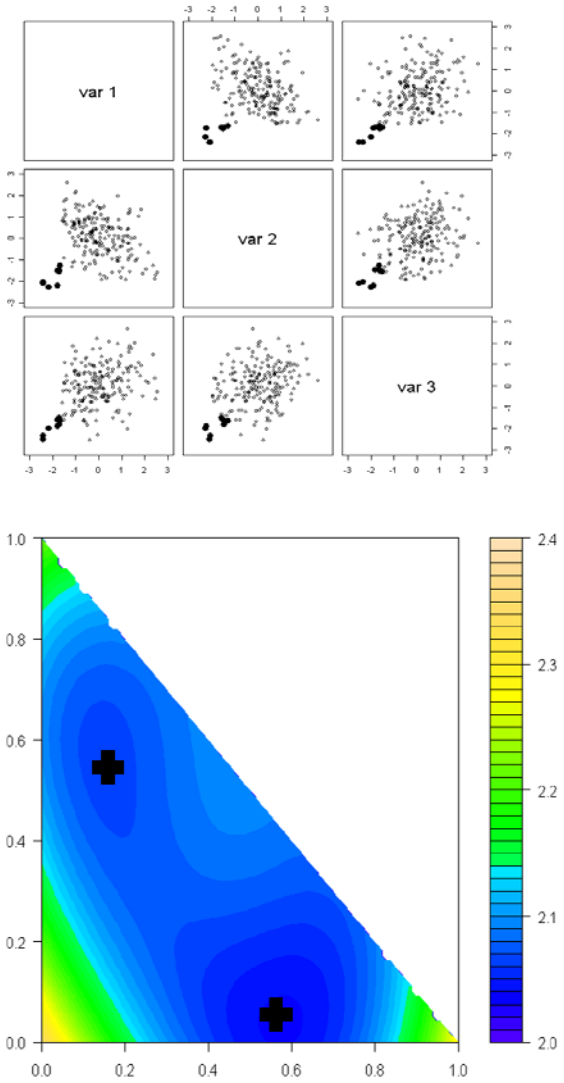

Figure 3.8 Scatter Plots of Simulated Variables (highly correlated negative values in solid cycle) and Contour Maps of MVaR

The above shows that even for a very small portfolio, the CF expansion can still have multiple local optima. For larger portfolios the number of local optima can grow in a combinatorial fashion.

We use the non-linear programming method developed by P. Spellucci (1997) to solve this problem. His donlp2 package, which is available in *R*, uses the *Sequential Quadratic Programming* (SQR) method (see Appendix A) for optimization. In order to have a higher probability of finding the global optima, we use multiple starting points. We had a previous experience comparing genetic algorithms and a different multiple starting point gradient descent algorithm (BFGS). We find both algorithms are insufficient in finding the global optimum. Comparing the quality of solutions and computational time spent, the genetic algorithm is faster in finding a good solution when the portfolio size is large, e.g. above 100 assets. For small sized portfolios, e.g. 10 or 20, there is BFGS algorithm is faster and has a higher probability for finding a good solution. Therefore we stay with the multiple starting point algorithm for our experiments.

The starting points at least include the solutions generated by other risk measures like minimum volatility, VaR or MVaR. If VaR is truly a good risk measure and having lower VaR will bring out better performance, we may not necessarily need to have the exact global minimum VaR to see this effect. Suppose we start from the minimum volatility portfolio and find a local minimum close to the starting point, it may not be the globally minimum VaR portfolio but at least it should beat the minimum volatility portfolio if VaR is truly better than volatility. We also include in our starting points the corner solutions that contain only one asset only in the portfolio, i.e. having unit weight on one asset and zero weights for the rest. In general, we start with at least 20 starting points. First all corner solutions consist 20 starting points. Once minimum volatility and maximum SR portfolios

are calculated, the solutions are included to the original 20 starting point pool for the calculation of VaR portfolio. Once VaR is done, its solution is again included for the MVaR portfolio optimization. At last MVaR solution is included for TMVaR portfolio optimization. We compare all the optimal solutions from each starting point and keep the best as our answer.

TMVaR is a little more complicated in the sense that they are defined on order statistics. Each time, the portfolio returns have to be sorted before trimmed. Otherwise they are treated the same as the non-trimmed version in terms of optimization methods.

3.4 Data and Experiment Settings

3.4.1 Data

The ideal place to experiment portfolio optimization with tail risk measures are data sets with lots of skewness and fat-tails. Hedged fund returns are a good place to start. Historically, hedge funds are known to have skewed and fat-tailed distribution for their returns. Different strategies applied by hedge funds make their return distributions even more diversified. We are very thankful to HedgeFund.net for kindly providing us the hedge fund database[3].

We focused our experiments with hedge fund returns for the 10 year period between 2000 and 2009 (inclusive). We believe longer history may not be beneficial to our experiments. Hedge funds may change their strategy or change their portfolio managers during longer periods of time. Such uncertainty may change the persistence in the

[3] Hedge fund data has been provided by channel Capital Group Inc. and its affiliates, and the aforementioned firms and each of their respective shareholders, employees, directors and agents have not independently verified the hedge fund data, do not represent it as accurate, true or complete, make no warranty, express or implied regarding it and shall not be liable for any losses, damages or expenses relating to its adequacy, accuracy, truth or completeness.

characteristics of the underlying return distribution, making it difficult for portfolios optimized for one period to work for the next period. We use the following steps to select our initial pool of candidates:

1. We got all hedge funds that have records in this period between 12/1999 and 11/2009, which totals 120 months.

2. We removed identical and highly similar funds (with a Pearson correlation greater than 0.95) from the candidates. This is the iterative process that we remove one fund in each iteration until no highly correlated funds are available.

3. We removed all hedge funds that have prolonged stretch of missing periods (more than 3 consecutive months of 0 returns on record). We believe having these funds may greatly distort our results. The missing returns sometimes are aggregated in the next available period, making the performance of any portfolio that includes this fund highly unstable.

In the end we have collected 379 funds for the entire 120-month period. The (log) return of the i th fund for the j the period is defined as the following,

$$r_{i,j} = \log \frac{P_{i,j}}{P_{i,j-1}}, \qquad i = 1 \ldots 379,\ j = 1 \ldots 120$$

where $P_{i,j}$ is the price for the i th fund for the j the period.

After decided on our pool of candidates, we form 100 portfolios, each with 20 hedge funds randomly selected from this hedge fund pool. We will compare the performance of every risk measure across the entire 100 portfolios. We believe the results will show us

the efficiency as well as the variability of each method. The random selection process uses a simple uniform distribution and each fund has equal opportunity to be selected in each portfolio.

For each portfolio, we use the first 5 years (60 months, from 2000-2005) historical returns as the training dataset. We perform optimization methods in this period and obtain the weights for the global minimum risk portfolio (except the maximum Sharpe ratio portfolio). We do not rebalance until the next month. The out-of-sample return $r_{p,i,j}$ of the i th portfolio for j the period is calculated as

$$r_{p,i,j} = \log\left\{\sum_{k=1}^{20} w_{k,j} \exp\left(r_{k,j}\right)\right\}$$

where $w_{k,j}$ is the optimal weights for k th fund in j th period. Note we have transformed the log returns to the arithmetic returns first and then transformed back. The reason for doing so is because log returns are not multiplicative when we scale the assets by weight $w_{k,j}$. However, we maintain the linear combination $r_{p,i,j} = \sum_{k=1}^{20} w_{k,j} r_{k,j}$ in our optimization procedures for simplicity and preserve convexity.

Once this process is finished, we roll forward for one month and keep our training period at 60 month (removing the first month in the last training set). We reiterate the optimization process and the performance calculation. This rolling window is moved forward 59 times to create 60 out-of-sample returns. These returns as well as the rebalancing weights for each period are analyzed using the various performance measures we selected.

3.4.2 Empirical Statistics on Hedge Fund Data

We first check the basic statistics of the returns of the 379 hedge funds we selected. The following Table 3.6 shows quantiles of the mean, volatility, skewness, kurtosis and the p-values from the Jarque Bera normality test of all funds.

Table 3.6 Empirical Statistics of Hedge Funds: Mean, Volatility, Skewness, Kurtosis and p-Values of Jarque Bera normality test

	Mean	Volatility	Skewness	Kurtosis	J.B. p-Val
Min.	-0.309	0.068	-8.895	-0.236	0.000
25%	0.014	0.245	-1.460	4.727	0.000
50%	0.138	0.362	-0.221	8.692	0.000
75%	0.267	0.544	1.013	20.020	0.000
Max.	0.977	7.785	9.582	97.260	0.945
Mean	0.145	0.448	-0.238	15.800	0.020

Among all these statistics, mean and volatility are annualized and based on arithmetic returns. Skewness and kurtosis are monthly based and on log returns. We find:

1. The average annualized return is about 14.4%, which is pretty high. The worst fund lost 30.1% on average annually. The first quartile (1.4%) and the third quartile (26.7%) are almost symmetric about the median (13.8%).

2. Annualized volatility averages 44.8%. Volatility grows much faster after passing the median line. However, the higher volatility fund has better returns compared to their volatility. The 25% line pairs 1.4% return with 24.5% volatility. The 50% line shows 13.8% vs. 36.2%. The 75% line matches 26.7% return with 54.4%.

3. On average all funds are slightly negatively skewed. The distribution of sample skewness is almost symmetric. The amount of skewness is prevalent in these funds. At least 50% of the funds have an absolute skewness greater than 1.

4. Kurtosis is very large for most of the funds. Collectively their average kurtosis is 15.8. The quantiles statistics also show most of the funds are fat-tailed.

5. Jarque-bera test also shows that most of the funds fail the normality test. At least 75% of the funds are rejected for the null hypothesis of normality at the level of 0.001.

In company of the statistics shown, we also present the histograms of sample skewness, kurtosis and p-values of J.B. test in Fig. 3.9.

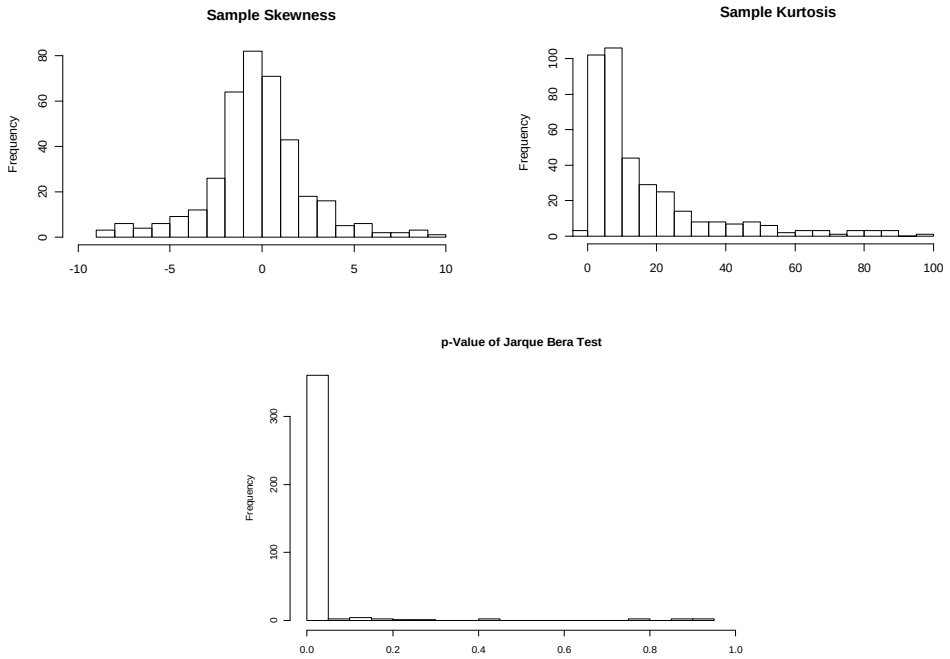

Figure 3.9 Histograms of Sample Skewness, Kurtosis and p-values of J.B. test.

As we can see from the histograms, about half of the funds are symmetric and most of them are fat-tailed.

We plot the normal quantile-quantile plots for selected funds in the Fig. 3.10 and 3.11. The funds are selected based on their sample kurtosis. We select 9 funds whose sample kurtosis ranks from 10% to 90% in all sample kurtosis.

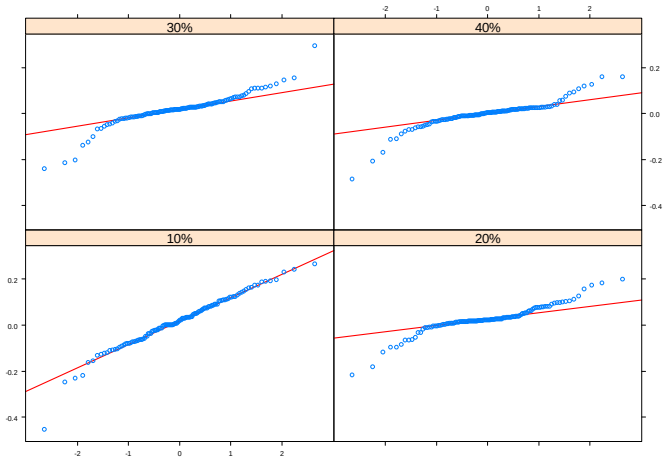

Figure 3.10 Quantile-Quantile plots for Hedge funds: First 4 plots. Each panel is a selected fund that ranks at the respective percentile in terms of sample kurtosis.

The fund ranks at 10% of all sample kurtosis has only one particular large negative return that deviates from the otherwise almost normal return distribution. At 20%, this departure from normality is more apparent. At 30% and 40% levels, even more so. However, these departures are relatively on small scale compared to the higher levels, particularly the 90% level, as in Fig. 3.11. That particular fund has both huge down-swings and up-swings.

The analysis above only shows the returns as if they are all independent. We'd also like to view these returns in perspective with time. Instead of drawing all return series, we use the following method to give a picture of how all funds behaved.

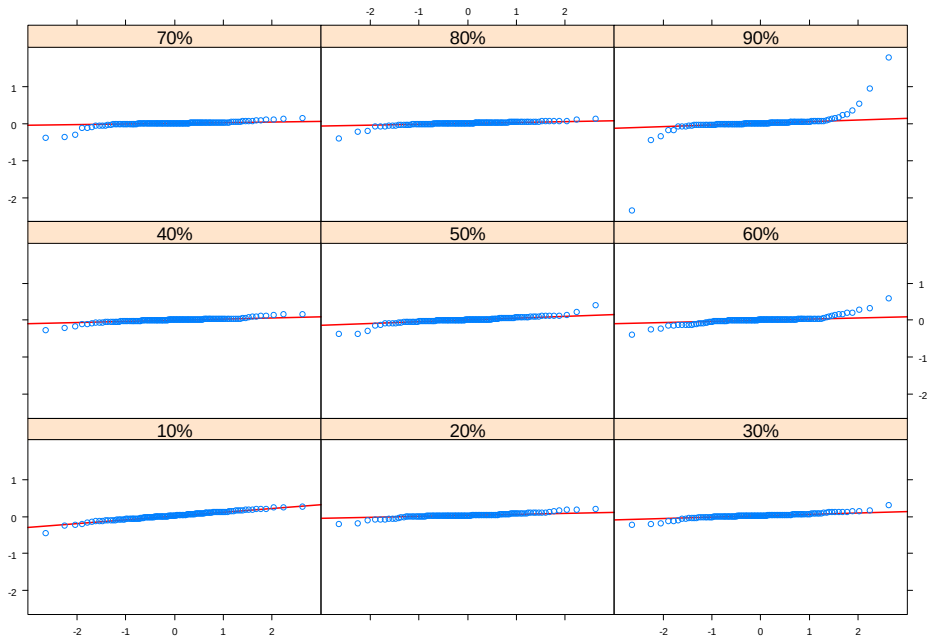

Figure 3.11 Quantile-Quantile plots for Hedge funds. Each panel is a selected fund that ranks at the respective percentile in terms of sample kurtosis.

For a particular percentile, we draw the cumulative return that ranks at this percentile at the cross section of each time period in Fig. 3.12 and call this the cumulative quantile plot. For example, we track the 25% quantile of all cumulative returns from Dec., 1999 to Nov., 2009 and connect them together to form the “25% line”. In this way, we can be assured that in any arbitrary period, there are 25% of the funds whose cumulative returns are below this line. These lines do not track any particular fund, but collectively they show how the majority of the funds evolve in time. We also add the S&P 500 index for the same period in the graph for reference. All curves are on log returns.

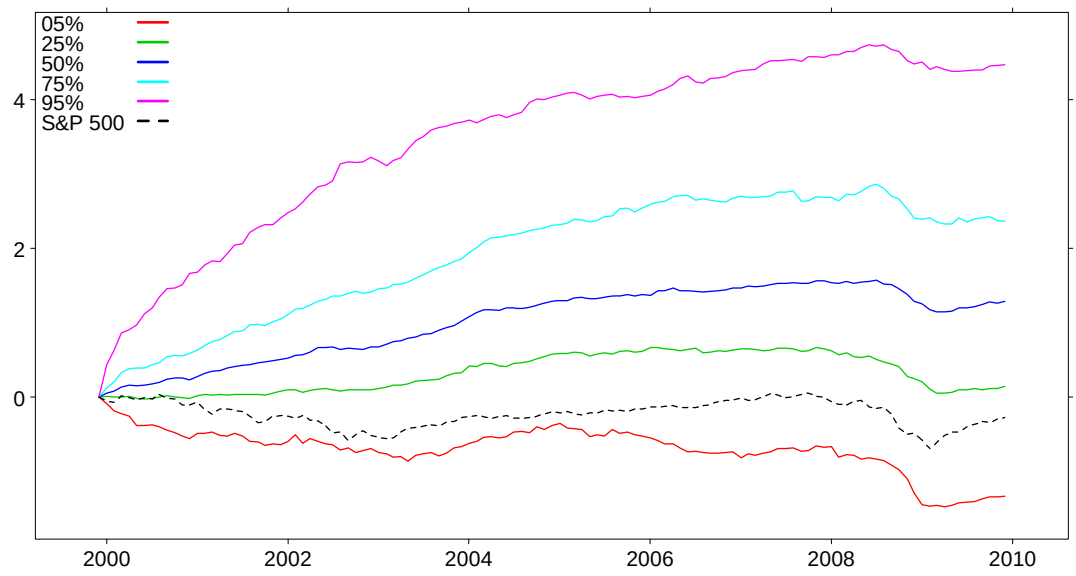

Figure 3.12 Quantile Cumulative Returns Curves for Hedge Funds.

We observe the following facts:

1 The S&P 500 (black dashed) is still in loss after 10 years. There are two particular downturns. One is the tech bubble from 2000 to 2003. The other is the financial crisis from 2008 to 2009. The first period in our initial training period. The second is in our out-of-sample period. The second decline is much sharper than the first one. The period from 2003 to 2008 is a gradual ascending period. After Mar., 2009, S&P 500 climbed sharply out of the bottom, too.

2 The majority of the hedge funds we selected outperformed S&P 500. Among the curves we draw, only the 5% line is below S&P 500. The 25% line is about even after 10 years. Therefore more than 75% of the funds made money after this decade.

3 The hedge funds returns are highly skewed towards the positive side. While the interquartile range is almost symmetric, the best funds made much more than the worst funds had lost. The 95% line shows that there are 5% of the funds ended with more than 54 ($e^4 \approx 54$) times of their original capital.

4 Most of the funds were affected by the 2008 financial crisis. Most of them did not climb back well as compared to the S&P 500. This is in sharp contrast to their performance in the tech bubble. At least 75% of the funds did not suffer from the bubble burst. Only the worst 5% went down with the index. The best performing 5% even gained a lot during that time.

In conclusion with our empirical studies of the hedge fund returns, we find the distributions of these funds are far from normal and they have high differentiability. Both characteristics are ideal for optimization with tail risk measures. We expect the appropriate risk measure that controls the downside risk will be more successful than the others in our experiments.

We formed 100 portfolios which each consist of 20 randomly selected hedge funds. The following Pareto chart shows that each of the 379 funds (the horizontal axis) is roughly uniformly represented in our portfolios.

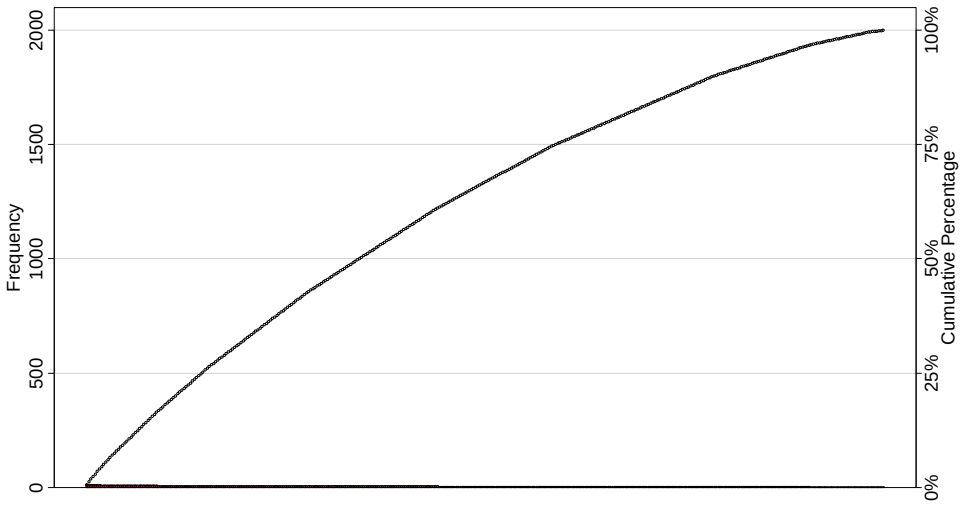

Figure 3.13 Pareto Chart of Hedge Funds

3.5 Results

3.5.1 Performance and Risk Measures

Before we get into the optimization results, we first discuss the performance measures we select to compare the performance of each method in three areas: return, risk and cost. These measures are selected based on the quantity they measure and their popularity in the real world.

1. Annualized Mean Return

$$\hat{\mu} = \frac{12}{T}\sum_{i=1}^{T} r_i$$

2. Annualized Volatility

$$\hat{\sigma} = \sqrt{\frac{12}{T-1}\sum_{i=1}^{T}\left(r_i - \hat{\mu}\right)^2}$$

3. Annualized Sharpe Ratio

$$SR = \frac{\hat{\mu} - r_f}{\hat{\sigma}},$$

where r_f is the 90 day T-bill rate for the same period.

4. ETL and VaR

ETL as a performance measure is based on the original definition of Conditional Value at Risk (CVaR) as follows,

$$\widehat{CVaR_{1-\alpha}(r)} \triangleq -\frac{1}{T(1-\alpha)} \sum_{i=1}^{\lceil T(1-\alpha) \rceil} r_{(i)}$$

where $r_{(i)}$ is the ith order statistic. VaR is the same as defined in (3.14)

5. Maximum Drawdown

We define drawdown as the arithmetic return of from the highest cumulative return in available history to the current point of time. The maximum drawdown is the maximum among all drawdown from each period as follows,

$$\text{maxDrawdown} = \left[\exp\left\{ \max_{1 \le i \le T} \left(\max_{1 \le j \le i} \sum_{k=1}^{j} r_k - \sum_{k=1}^{i} r_k \right) \right\} - 1 \right] \times 100\% .$$

6. Diversification

We use the following working definition of average diversification,

$$\text{diversification} = \frac{1}{T} \sum_{i=1}^{T} \left(1 - \sum_{i=1}^{N} w_{i,j}^2 \right).$$

For each period, the diversification measure $d_j = 1 - \sum_{i=1}^{N} w_{i,j}^2$ is a measure of how spread the weight allocation is. For example, for 60 assets, if a particular asset takes all capital, $d = 0$. If weights are equally divided among 60 assets

$$d = 1 - 60 \times \left(\frac{1}{60} \right)^2 = \frac{59}{60}.$$

7. Turnover Rate

We use the following working definition of average turnover.

$$\text{Turnover}=\frac{1}{T}\sum_{i=2}^{T}\left(\sum_{j=1}^{N}\left|w_{i,j}-w_{i-1,j}\right|\right)$$

It measures the total change in weights for every period and averages them from all periods. The 0 period (one period before the first out-of-sample period) weights are all 0.

8. STARR Ratio

We use the following working definition of (centered) STARR ratio,

$$\text{Starr}=\frac{\hat{\mu}-r_f}{\hat{\mu}+\widehat{CVaR}_{1-\alpha}(r)}$$

In this way, we avoid the confusion when $\widehat{CVaR}_{1-\alpha}(r)$ is negative.

9. Semi-Standard Deviation (SSD)

We use the following working definition of SSD.

$$\text{SSD}=\sqrt{\sum_{r_i<\hat{\mu}}\left(\hat{\mu}-r_i\right)^2\Big/\sum_{r_i<\hat{\mu}}1}$$

It measures the deviation of the lower tail.

3.5.2 Experiments with Long-Only Constraints

In this experiment, we enforce the long only constraint. We have 6 methods altogether in this horse race: Minimum volatility (MinVol), Maximum Sharpe Ratio (SR), VaR (VaR), MVaR (MVaR), and Trimmed TMVaR (TMVaR). The names in the parenthesis are the code names we use in the plots and tables.

In Table 3.7 and 3.8, we display the summary statistics on respective out-of-sample performance measures for each portfolio optimization method across all 100 portfolios we formed. In Fig. 3.14, Fig. 3.15, we display the kernel densities of these statistics. In Fig. 3.16, we display the box-plots of these statistics by comparing the difference on the same portfolio by different methods. For better presentation, some of the outliers are not shown (outside the plotting region).

Table 3.7 Summary Statistics of Performance Measures of Long Only Experiments

	Annualized Mean Return					Diversification				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	-0.149	-0.137	-0.076	-1.13	-0.158	0.599	0.409	0.668	0.014	0.359
25%	0.022	-0.008	0.046	-0.005	0.025	0.741	0.76	0.771	0.289	0.586
Median	0.067	0.033	0.084	0.094	0.09	0.767	0.814	0.803	0.442	0.666
Mean	0.061	0.031	0.094	0.116	0.101	0.765	0.789	0.797	0.428	0.655
75%	0.1	0.066	0.146	0.178	0.17	0.808	0.848	0.833	0.566	0.741
Maximum	0.178	0.225	0.307	1.06	0.533	0.847	0.893	0.859	0.808	0.838
	Sharpe Ratio					Starr Ratio				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	-0.927	-1.22	-0.501	-1.14	-0.711	-0.0874	-0.142	-0.0495	-0.111	-0.0706
25%	-0.053	-0.408	0.165	-0.176	-0.028	-0.0057	-0.0408	0.0177	-0.0176	-0.0029
Median	0.288	0.044	0.499	0.381	0.359	0.0279	0.007	0.0524	0.0419	0.042
Mean	0.344	-0.005	0.546	0.409	0.43	0.0389	0.0079	0.0668	0.0652	0.0612
75%	0.617	0.34	0.818	0.815	0.872	0.062	0.0476	0.0978	0.104	0.121
Maximum	2.02	1.15	2.34	2.43	1.94	0.264	0.284	0.282	0.526	0.27

Table 3.8 Summary Statistics of Risk Measures of Long Only Experiments

	Annualized Volatility					Semi. Std. Dev.				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	0.071	0.057	0.069	0.079	0.072	0.0719	0.0636	0.0724	0.0846	0.074
25%	0.104	0.088	0.103	0.154	0.121	0.124	0.0979	0.115	0.169	0.133
Median	0.126	0.104	0.121	0.193	0.148	0.152	0.113	0.144	0.216	0.169
Mean	0.135	0.118	0.135	0.304	0.18	0.176	0.118	0.151	0.457	0.189
75%	0.151	0.128	0.139	0.302	0.201	0.201	0.135	0.169	0.292	0.213
Maximum	0.326	0.359	0.531	1.66	0.622	0.529	0.206	0.437	4.93	0.865
	Value at Risk					Expected Tail Loss				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	0.019	0.02	0.015	0.022	0.016	0.033	0.038	0.031	0.05	0.032
25%	0.044	0.04	0.04	0.047	0.045	0.073	0.059	0.071	0.097	0.079
Median	0.058	0.049	0.051	0.066	0.06	0.094	0.069	0.086	0.135	0.101
Mean	0.062	0.05	0.051	0.072	0.062	0.107	0.073	0.091	0.21	0.117
75%	0.075	0.058	0.061	0.086	0.074	0.125	0.088	0.11	0.195	0.14
Maximum	0.128	0.098	0.105	0.215	0.141	0.308	0.139	0.233	1.88	0.414
	Maximum Drawdown					Turnover				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	0.056	0.098	0.084	0.073	0.079	0.063	0.051	0.163	0.048	0.12
25%	0.254	0.217	0.23	0.287	0.252	0.081	0.073	0.231	0.139	0.162
Median	0.334	0.271	0.301	0.378	0.327	0.089	0.08	0.252	0.183	0.217
Mean	0.341	0.28	0.292	0.416	0.333	0.09	0.079	0.251	0.19	0.215
75%	0.429	0.336	0.357	0.507	0.394	0.096	0.086	0.27	0.239	0.254
Maximum	0.63	0.542	0.578	0.997	0.668	0.126	0.102	0.344	0.42	0.48

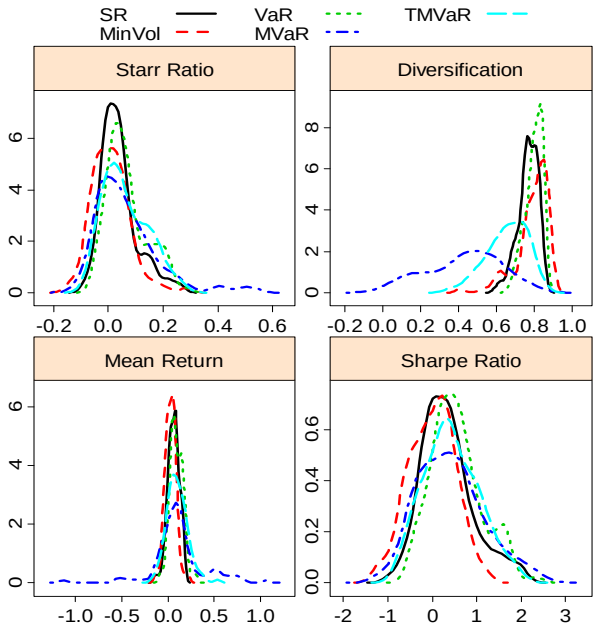

Figure 3.14 Kernel Densities of Performance Measures of Long-Only Experiments

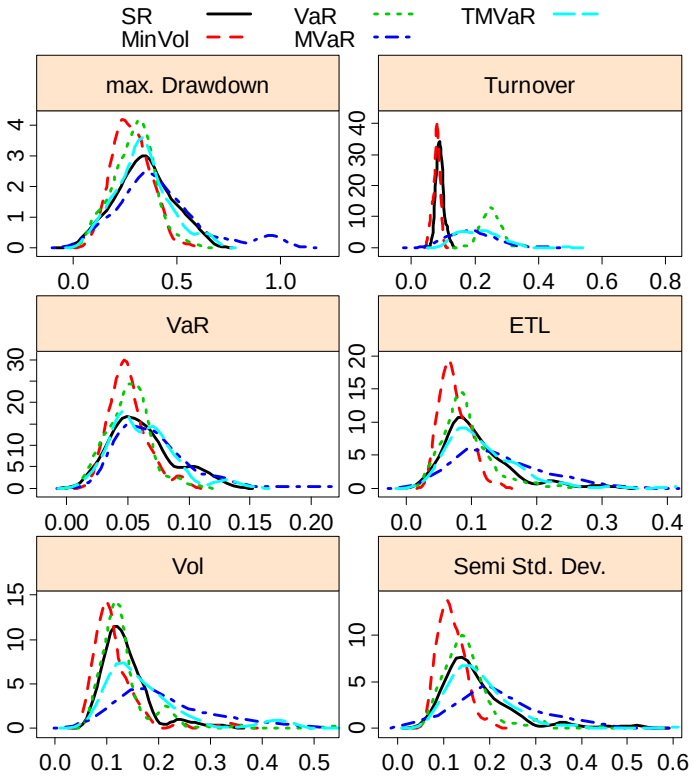

Figure 3.15 Kernel Densities of Risk Measures of Long Only Experiments

Combining results from these analyses, we observe the following facts.

1. The VaR, MVaR and TMVaR methods have better out-of-sample performance. While the MinVol method is almost symmetric in terms of Sharpe ratio and STARR ratio, the VaR family has more better-performing portfolios.

2. MVaR method is extremely volatile as seen in its out-of-sample mean returns. It also has very concentrated portfolios.

3. In terms of out-of-sample diversification, MVaR is particularly bad in this measure. It on average is equivalent to equally weighted investment in 2 assets. $(1-2\left(\frac{1}{2}\right)^2=0.5)$. The trimmed version is much better. The other methods are on par at 0.8 in average, which is equivalent to equally weighted investment in 5 assets $(1-5\left(\frac{1}{5}\right)^2=0.8)$.

4. The out-of-sample risk profiles show VaR has the best risk control compared to MVaR and TMVaR, except for the turnover rate. TMVaR is still better than MVaR. MinVol shows good control on the risk, too. MVaR shows some extreme out-of-sample risks.

5. VaR has the worst turnover rate.

In summary, the portfolios using VaR family measures behave very differently from those using the volatility measure (see Fig. 3.16, 3.17, 3.18). Minimum volatility portfolio not necessarily underperforms in the time of crisis. It shows good control of out-of-sample downside risk in general. However, it apparently underperforms when market is trending upwards (2005-2008), because it treats the upside volatility as risk, too. This is reflected from its poor cumulative returns.

VaR family measures generally perform well in those times. MVaR, by its incorporation of higher moments, can be heavily influenced by extreme returns, as proved in previous sections. This results in its extreme behaviors on both the best and worst portfolios. Overall the winners over-compensate the losers (see in Fig. 3.18). However, the huge break down in the cumulative quantile plot in Fig. 3.19 makes it a dangerous risk measure. We find 0.15% of all out-of-sample period returns have absolute values that are greater than 1 on log scale, i.e. 172% or -61% return in one month. The VaR method is not affected by such erratic returns, because VaR is set to the 95% confidence level. MVaR is heavily impacted. For a training period that contains a very large positive outlier, MVaR is grossly underestimated. MVaR method mistakenly concentrates all its investment in one asset and subsequently is hammered by a large negative return. This is exactly what happened in the period of 04/2007 in Fig. 3.20. By trimming one point on each tail, we also can avoid such big impacts (see Fig. 3.19 in absolute scale and Fig. 3.20 in log scale). Consequently, it does not have the wild super performance on the upside. However, overall it still beats the plain VaR method (see Fig. 3.21). The plain VaR method shows more moderate results compared the other two VaR measures. All VaR methods have high turnover rates.

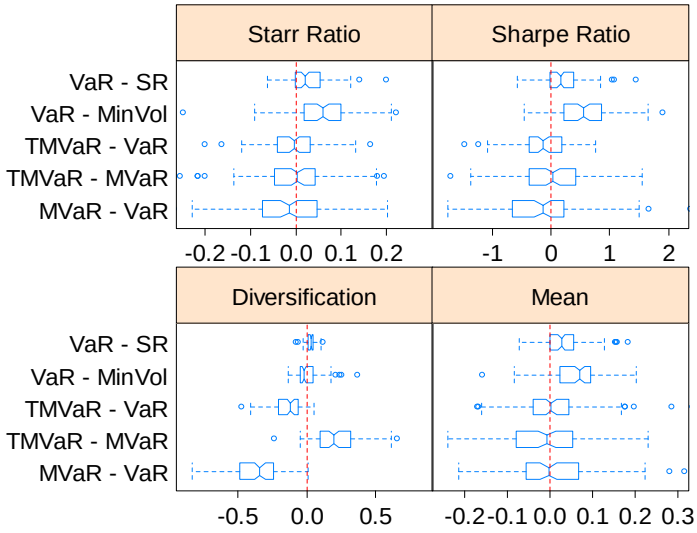

Figure 3.16 Notched Box-plots of Paired Difference for Performance Measures: Long Only Experiment

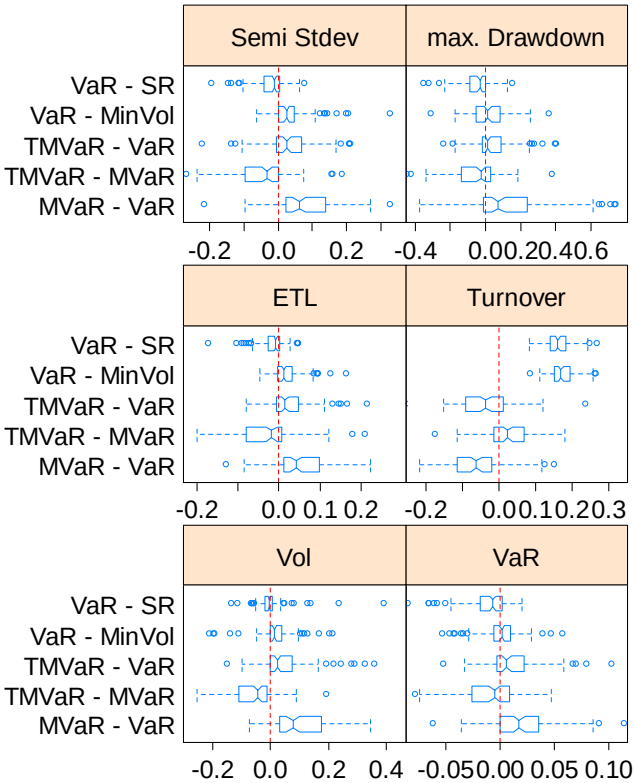

Figure 3.17 Notched Box-plots of Paired Difference for Risk Measures: Long Only Experiment

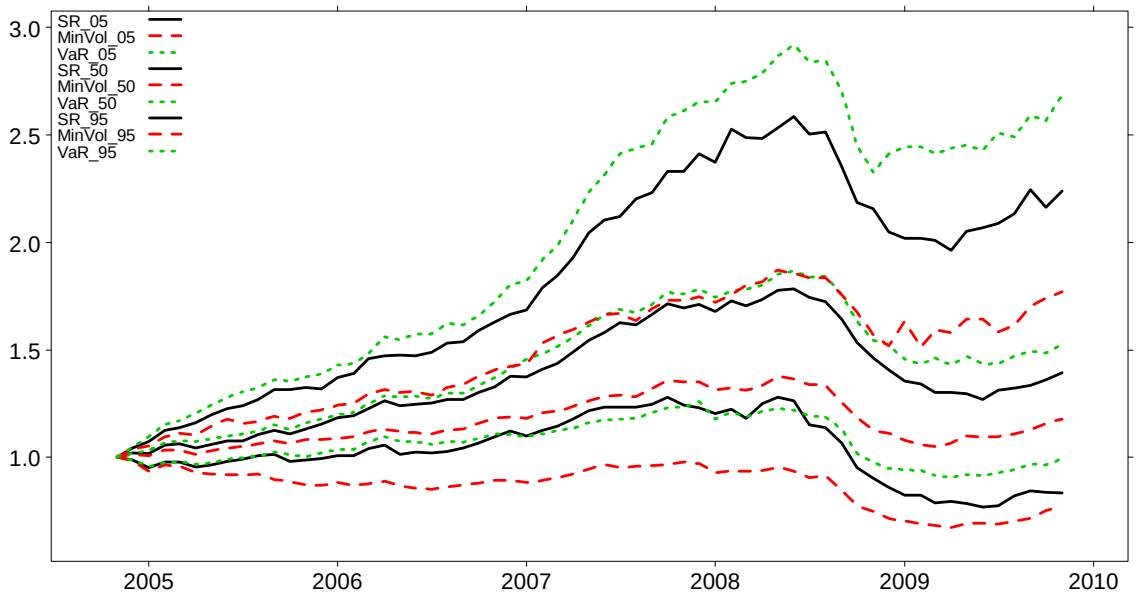

Figure 3.18 Quantile Cumulative Return Plots for the minimum Volatility, VaR and maximum Sharpe Ratio portfolio. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. VaR is set at 95% confidence level.

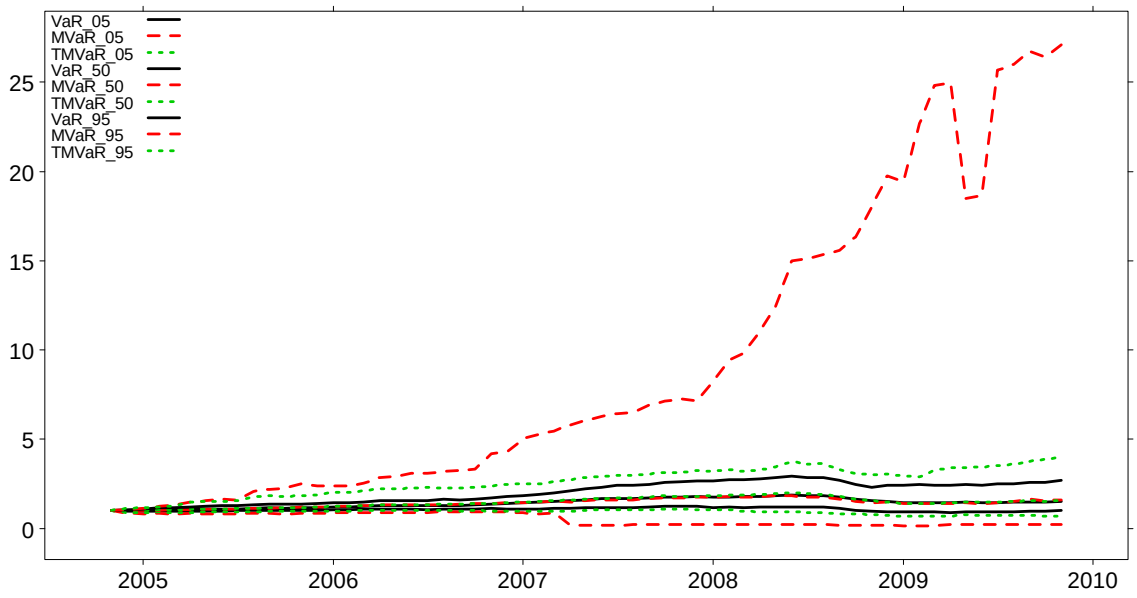

Figure 3.19 Quantile Cumulative Return Plots for the VaR family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. VaR are all set at 95% confidence level and trimming is for 2.5% of either tail.

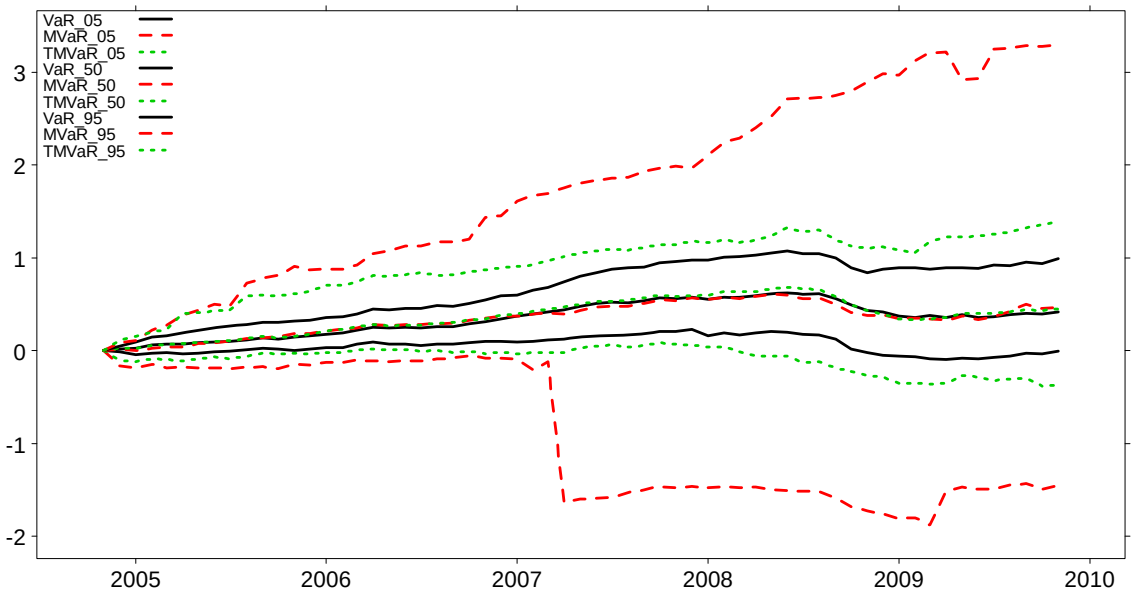

Figure 3.20 Quantile Cumulative Return Plots (Log Scale) for the VaR family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. VaR are all set at 95% confidence level and trimming is for 2.5% of either tail.

Figure 3.21 Mean Cumulative Return Plots for the VaR family. Each curve represents the average of the cumulative returns across all 100 portfolios at the different time. VaR are all set at 95% confidence level and trimming is for 2.5% of either tail.

75% VaR

The experiments above were set at 95% confidence level VaR. To explore a different confidence level and trimming, we did the same portfolio optimization with VaR confidence level set to 75% and trim 5%, i.e. three data points on either end. The results are shown in Fig. 3.22 and Fig. 3.23. Because we are taking the VaR confidence level to be so closer to the center of the distribution, non-normality becomes a much lesser issue. This is reflected in the closer performance all three VaR methods. MVaR still shows very large uncertainty in its performance. Overall it is the only one that bounced back quickly after the 2008 crisis. However, its instability still makes a dangerous risk measure, as shown by its worst 5% portfolios. TMVaR completely avoided the dramatic downside move in the worst cases for MVaR. We also find turnover rate for TMVaR is only half (17%) of the VaR method (33%) in this case. The original MVaR has almost the same turnover-rate as VaR method.

In summary, if we were to hold all these 100 portfolios, MVaR would be a good method to achieve superior average performance, especially in time of crisis. The more extreme the confidence level is set, the more extreme behavior we will get from MVaR. However, if we were only to hold one portfolio, MVaR is a very dangerous measure to use. It is a very weak measure to defend against erratic returns, especially in the high confidence level. TMVaR is much more stable in such cases. It helps to stabilize the results. In our experiments, it also slightly outperforms VaR.

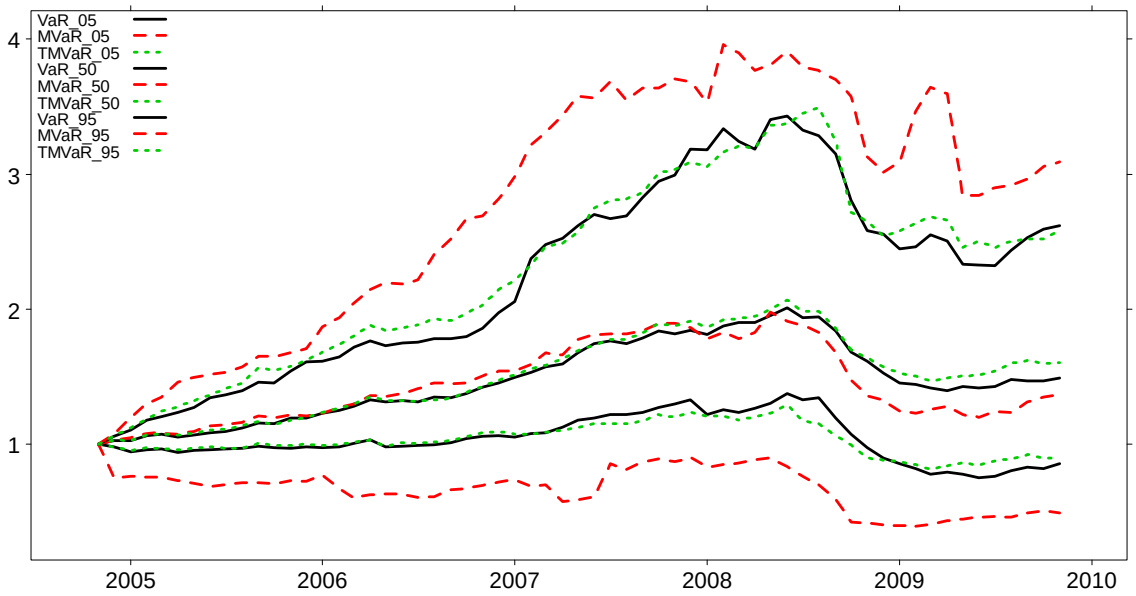

Figure 3.22 Quantile Cumulative Return Plots for the VaR family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. VaR are all set at 75% confidence level and trimming is for 5% of either tail.

Figure 3.23 Mean Cumulative Return Plots for the VaR family. Each curve represents the average of the cumulative returns across all 100 portfolios at the different time. VaR are all set at 75% confidence level and trimming is for 5% of either tail.

3.5.3 Experiments with Upper Bounded Constraints

In this section we cap our investment uniformly at 20% maximum for any single hedge fund. In the last section, we find out that some methods are already diversified (table 3.7) equivalently to this level on average. We'd like to explore the effect of this upper bounded constraint on those methods that weren't quite diversified under the long-only constraint.

We summarize our results in Table 3.9 and 3.10, Fig. 3.24 to Fig. 3.27. We find the well-diversified portfolios under long-only constraint are not changed much, e.g. the minimum volatility portfolios.

MVaR and TMVaR portfolios were not well-diversified previously. After enforcing the upper bounded constraint, it behaves much better. Fig. 3.28 shows they are almost the same as the VaR method in terms of performances prior to the crisis. They still have better performance on the best 5% portfolios during the crisis period. The leading amount is more than enough to compensate the underperformance for the worst 5% portfolios. One important difference is that MVaR has a much lower turnover rate than the VaR method. We believe this is due to the smoother forms by its definition. TMVaR retained all these characteristics. In this case, TMVaR does not hold any advantage over MVaR, except it has slightly smaller out-of-sample VaR and ETL.

Table 3.9 Summary Statistics of Performance Measures of Upper Bounded Experiments

	Annualized Mean Return					Diversification				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	-0.102	-0.113	-0.033	-0.045	-0.051	0.828	0.845	0.854	0.835	0.839
25%	0.033	-0.004	0.032	0.02	0.028	0.849	0.866	0.868	0.848	0.851
Median	0.071	0.032	0.078	0.071	0.072	0.855	0.872	0.873	0.855	0.857
Mean	0.068	0.035	0.084	0.082	0.078	0.855	0.872	0.872	0.854	0.857
75%	0.104	0.072	0.123	0.123	0.117	0.861	0.879	0.876	0.859	0.863
Maximum	0.18	0.211	0.446	0.519	0.231	0.876	0.896	0.887	0.872	0.882
	Sharpe Ratio					Starr Ratio				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	-0.957	-1.2	-0.429	-0.575	-0.673	-0.102	-0.126	-0.0577	-0.0711	-0.0815
25%	0.053	-0.352	0.034	-0.076	0	0.0049	-0.0376	0.0042	-0.0078	-0.0001
Median	0.37	0.038	0.479	0.352	0.399	0.0425	0.0038	0.0513	0.0386	0.0438
Mean	0.356	0.009	0.44	0.405	0.421	0.041	0.0101	0.0582	0.0549	0.0539
75%	0.625	0.326	0.764	0.76	0.837	0.0748	0.0486	0.1	0.0977	0.102
Maximum	1.53	1.15	1.77	1.9	1.63	0.185	0.288	0.313	0.306	0.232

Table 3.10 Summary Statistics of Risk Measures of Upper Bounded Experiments

	Annualized Volatility					Semi. Std. Dev.				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	0.076	0.068	0.076	0.074	0.072	0.0767	0.0709	0.0779	0.0817	0.083
25%	0.106	0.09	0.102	0.109	0.101	0.116	0.0932	0.115	0.116	0.111
Median	0.122	0.103	0.112	0.124	0.117	0.145	0.109	0.132	0.133	0.129
Mean	0.122	0.12	0.132	0.133	0.12	0.145	0.113	0.134	0.136	0.13
75%	0.136	0.126	0.128	0.138	0.128	0.163	0.13	0.151	0.153	0.147
Maximum	0.2	0.375	0.813	0.978	0.365	0.311	0.182	0.26	0.226	0.21
	Value at Risk					Expected Tail Loss				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	0.022	0.021	0.025	0.025	0.019	0.035	0.035	0.037	0.041	0.037
25%	0.044	0.039	0.04	0.042	0.041	0.065	0.056	0.063	0.065	0.059
Median	0.054	0.047	0.049	0.051	0.049	0.087	0.067	0.077	0.078	0.077
Mean	0.058	0.049	0.05	0.053	0.05	0.088	0.069	0.08	0.081	0.077
75%	0.072	0.056	0.06	0.06	0.061	0.104	0.082	0.091	0.09	0.089
Maximum	0.111	0.091	0.086	0.096	0.081	0.183	0.121	0.176	0.148	0.135
	Maximum Drawdown					Turnover				
	SR	MinVol	VaR	MVaR	TMVaR	SR	MinVol	VaR	MVaR	TMVaR
Minimum	0.124	0.093	0.12	0.114	0.105	0.045	0.05	0.141	0.059	0.055
25%	0.254	0.227	0.234	0.226	0.226	0.07	0.064	0.206	0.098	0.116
Median	0.316	0.273	0.296	0.293	0.287	0.077	0.07	0.228	0.117	0.135
Mean	0.319	0.28	0.292	0.289	0.284	0.077	0.07	0.225	0.126	0.139
75%	0.375	0.334	0.354	0.353	0.334	0.084	0.077	0.241	0.151	0.161
Maximum	0.584	0.463	0.502	0.445	0.461	0.108	0.091	0.307	0.27	0.261

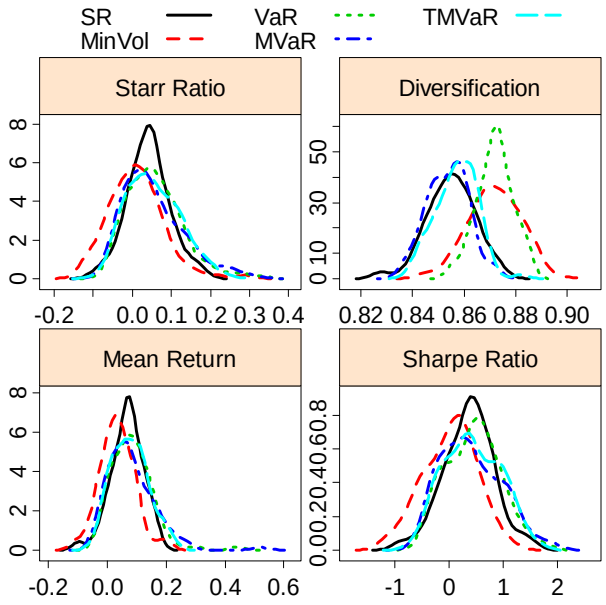

Figure 3.24 Kernel Densities of Performance Measures of Upper Bounded Experiments

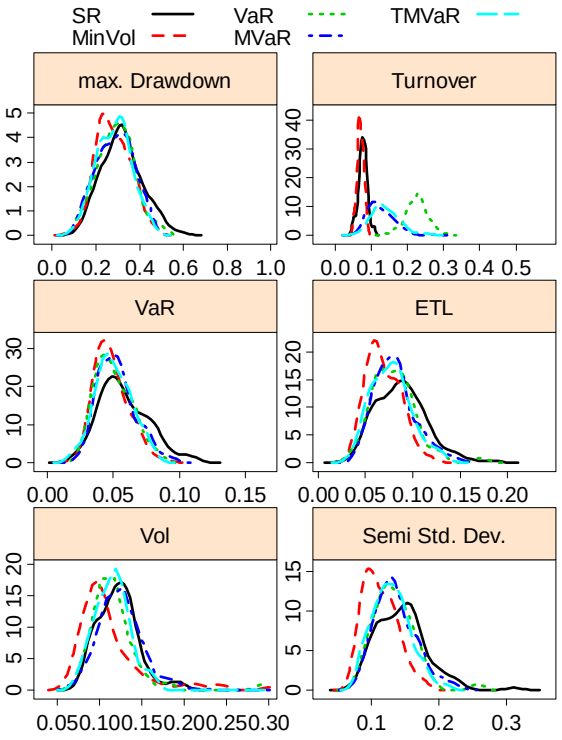

Figure 3.25 Kernel Densities of Risk Measures of Upper Bounded Experiments

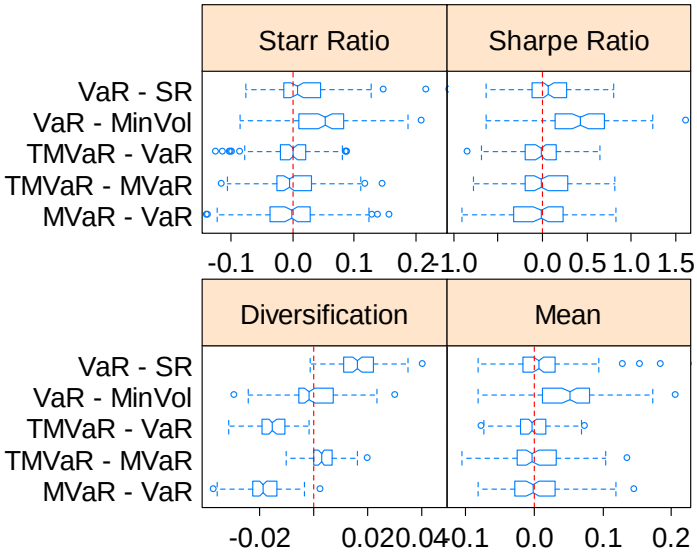

Figure 3.26 Notched Box-plots of Paired Difference for Performance Measures: Upper Bounded Experiment

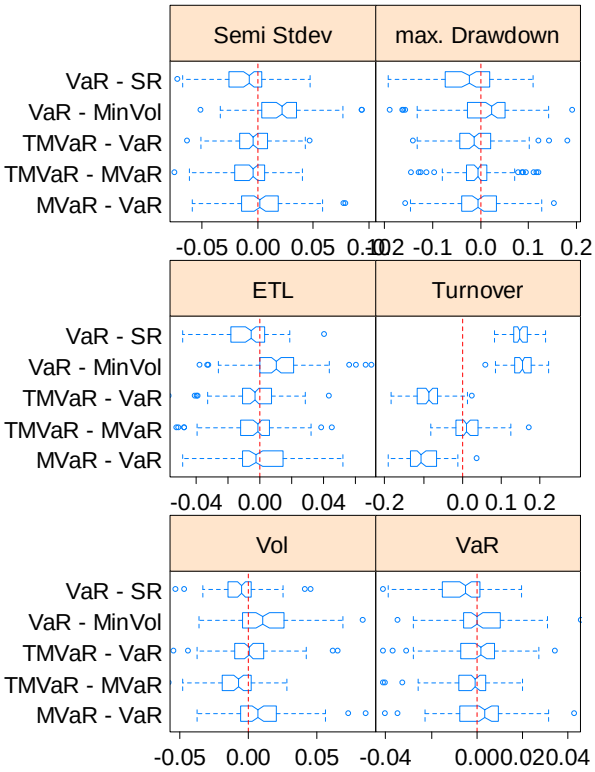

Figure 3.27 Notched Box-plots of Paired Difference for Risk Measures: Upper Bounded Experiment

Figure 3.28 Quantile Cumulative Return Plots for the VaR family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. VaR are all set at 95% confidence level and trimming is for 2% of either tail.

3.5.4 Conclusions

In this chapter we have shown MVaR is quite different from the conventional VaR. Through the study of its influence functions we find that MVaR is impacted by both the downside and upside of the distribution. Under some extreme circumstances when extreme outliers present, MVaR breaks down and grossly underestimates the risk. MVaR is generally smoother than the original non-parametric VaR.

TMVaR greatly stabilizes the variability in the estimates of MVaR. The empirical experiments with hedge funds data show that MVaR in its original form could result in disastrous portfolios that are heavily concentrated and bet on the wrong investments. Trimmed version greatly prevents this fault. If we ever need to use MVaR as a risk measure, we need to either use the trimmed version or enforce an upper bounded-constraint on our portfolio. Both would stabilize the results and improve performance.

On broader comparison, we also showed that tail risk measures are much better than variance, especially in a financial downturn environment. VaR or MVaR is competitive enough to beat minimum volatility in almost every measure.

Chapter 4
Portfolio Optimization with Modified ETL

Tail risk measures are not all equally effective. For example while VaR and CF-modified VaR (MVaR) are downside risks that take into account skewed and fat-tailed non-normality of the returns distribution, they are quite non-convex measures, leading to difficult portfolio optimization problem. On the other hand expected tail loss (ETL), a.k.a Conditional Value-at-Risk (CVaR), defined as

$$ETL_{1-\alpha}(r) = -E(r \mid r \le VaR_{1-\alpha}) \tag{4.1}$$

is a more effective risk measure in that it measures the average loss beyond VaR, thus providing much more complete loss information than VaR.

Furthermore it is well known that ETL is a convex risk measure (see Rockafellar and Uryasev, 2002, and Pflug, 2000) whose minimization subject to convex portfolio weights constraints is easily solved via linear programming, even for portfolios with quite large numbers of assets, e.g., 1500-2000 assets. Convexity also insures sub-additive that yields a highly desirable diversification property when combining assets or portfolios. Thus ETL is a risk measure that is much preferred over VaR.

Finally we note that Artzner et. al. (1999) introduced the concept of a coherent risk measure as one that satisfies a set of axioms with which a good risk measure should satisfy. Subsequently Acerbi (2001) proved that ETL is the only risk measure among volatility, VaR and ETL that is coherent.

Boudt, Peterson and Croux (2008) applied the Cornish Fisher expansion to ETL and obtained a modified ETL (METL) risk measure. The resulting modified ETL is defined as

$$\begin{aligned}\mathrm{METL}(1-\alpha) &\triangleq -\mu_p - \sigma_p \cdot E_G \\ &= -\mu_p + \frac{1}{\alpha}\sigma_p \cdot \phi(g_\alpha) \\ &+ \frac{1}{\alpha}\sigma_p \cdot \left(\begin{array}{c} \frac{1}{24}\left[I^4 - 6I^2 + 3\phi(g_\alpha) \right] \cdot K_p + \frac{1}{6}\left[I^3 - 3I^1 \right] \cdot S_p \\ + \frac{1}{72}\left[I^6 - 15I^4 + 45I^2 - 15\phi(g_\alpha) \right] \cdot S_p^2 \end{array} \right)\end{aligned} \tag{4.2}$$

where the $I = I(\cdot)$ is a function g_α is defined by

$$I^q = \begin{cases} \sum_{i=1}^{q/2} \left(\frac{\prod_{j=1}^{q/2} 2j}{\prod_{j=1}^{i} 2j} \right) g_\alpha^{2i} \phi(g_\alpha) + \left(\prod_{j=1}^{q/2} 2j \right) \phi(g_\alpha) & \text{for } q \text{ even} \\ \sum_{i=0}^{(q-1)/2} \left(\frac{\prod_{j=0}^{(q-1)/2} (2j+1)}{\prod_{j=0}^{i} (2j+1)} \right) g_\alpha^{2i} \phi(g_\alpha) - \left(\prod_{j=0}^{(q-1)/2} (2j+1) \right) \phi(g_\alpha) & \text{for } q \text{ odd} \end{cases} \tag{4.3}$$

and g_α is the α quantile on the underlying distribution $G(\bullet)$. Just as ETL may be represented as the integral of VaR(α') below the VaR tail probability α, METL is defined by integration with the Edgeworth expansion below the MVaR tail probability α. The first two terms in (4.2) is the ETL under normal distribution. The rest are revisions based on skewenss and kurtosis. Boudt et. al. (2008) found that the above METL can sometimes be very misleading in that it can sometimes have a value smaller than that of MVaR. As a simple fix for this problem they imposed a constraint that METL be at least as large as MVaR.

In section 4.1 we will show the empirical influence functions for METL. We also perform a Monte Carlo simulation study to validate our results in section 4.2. We show the results of finite sample variance for ETL, METL and trimmed modified ETL (TMETL). We will follow that by portfolio optimization experiments in section 4.3.

4.1 Modified ETL Empirical Influence Functions

In this section, we study the behavior of the MVaR through its empirical influence functions. The METL defined in (4.2) – (4.3) is very complicated and consequently is quite difficult derive an analytic expression for its (asymptotic) influence function. So instead we calculate the empirical influence functions

$$EIF(z;\hat{\theta}_n,\mathbf{r}) = (n+1)\left(\hat{\theta}_n(z,\mathbf{r}) - \hat{\theta}_n(\mathbf{r})\right) \tag{4.4}$$

where $\hat{\theta}_n = \hat{\theta}_n(\mathbf{r})$ is the univariate estimator based on an appropriate sample $\mathbf{r} = (r_1, r_2, ..., r_n)$ from a target distribution function F_0 and z is the contamination value.

We use the following method of calculating empirical influence functions. We let the sample $\mathbf{r}$ be 10,000 quantiles of F_0 from 0.01% to 99.99% in steps of .01% and use these to compute $\hat{\theta}_n(\mathbf{r})$. Then a contamination return with value z is added to the original 10,000 points to compute $\hat{\theta}_n(z,\mathbf{r})$, and the empirical influence function is calculated according to (4.4).

In the following we plot empirical influence functions for METL and trimmed METL (TMETL) at the 95% confidence level for the following distributions F_0: normal, symmetric t-distribution and skewed t-distribution. The skewed-t distribution (see Fernandez and Steel, 1998) we use here is defined by

$$f(x;n,\gamma) = \begin{cases} 2/(\gamma+1/\gamma)\cdot t(\gamma\cdot x;n) & \text{for } x<0 \\ 2/(\gamma+1/\gamma)\cdot t(1/\gamma\cdot x;n) & \text{for } x\geq 0 \end{cases} \tag{4.5}$$

where $t(\cdot;n)$ is the density of student-t distribution with n degrees of freedom. There are many other forms of skewed-t distributions, e.g. see Aas and Haff, 2006. We chose this particular form for simplicity in this application. It has the simple property that $f(\cdot;n,\gamma)$ and $f(\cdot;n,1/\gamma)$ have the same magnitude of skewness but opposite signs.

For METL we use symmetric trimming percents of 2.5%, 1% and 0.5%. The red dashed line shows the influence functions for plain ETL.

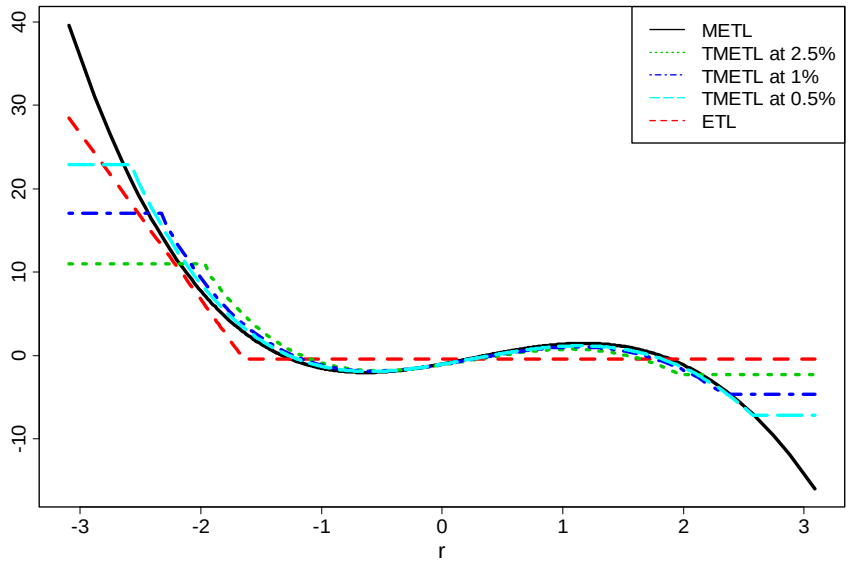

Figure 4.1Influence Functions of METL and TMETL Compared to ETL (red dashed lines) under standard normal distribution.

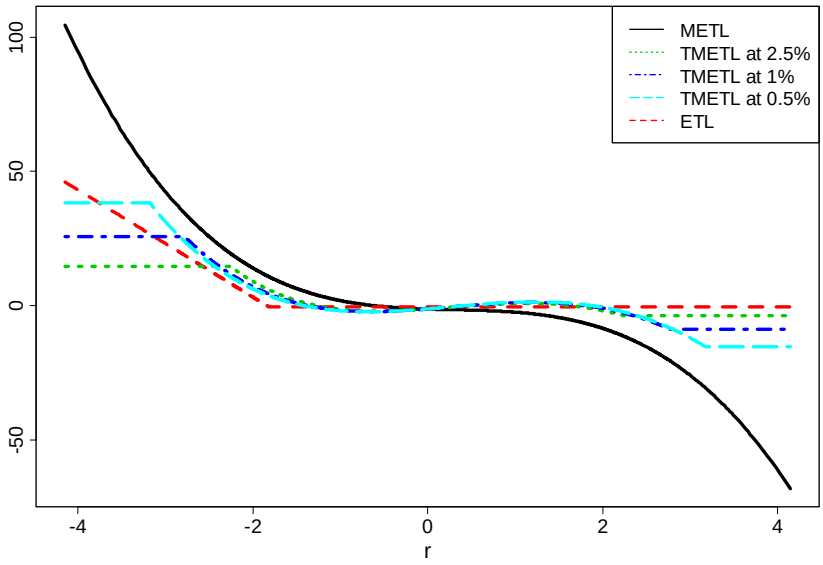

Figure 4.2 Influence Functions of METL and TMETL Compared to ETL (red dashed lines) under Student-t distribution with $df = 10$.

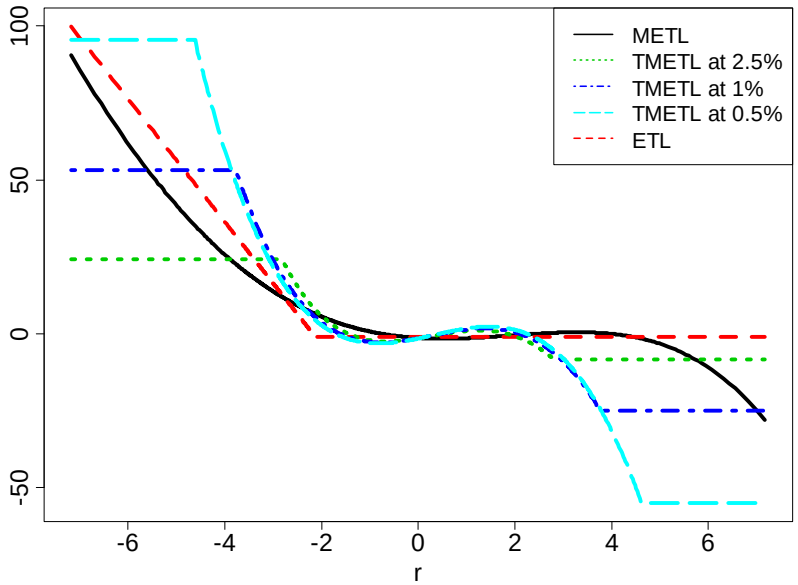

Figure 4.3 Influence Functions of METL and TMETL Compared to ETL (red dashed lines) under Student-t distribution with $df = 4$.

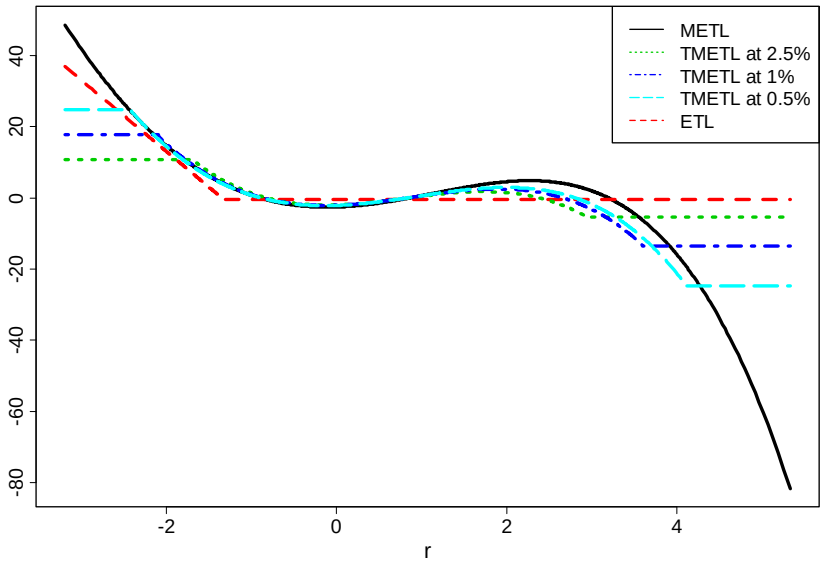

Figure 4.4 Influence Functions of METL and TMETL Compared to ETL (red dashed lines) under skewed-t distribution with $df = 10, \gamma = 1.25$.

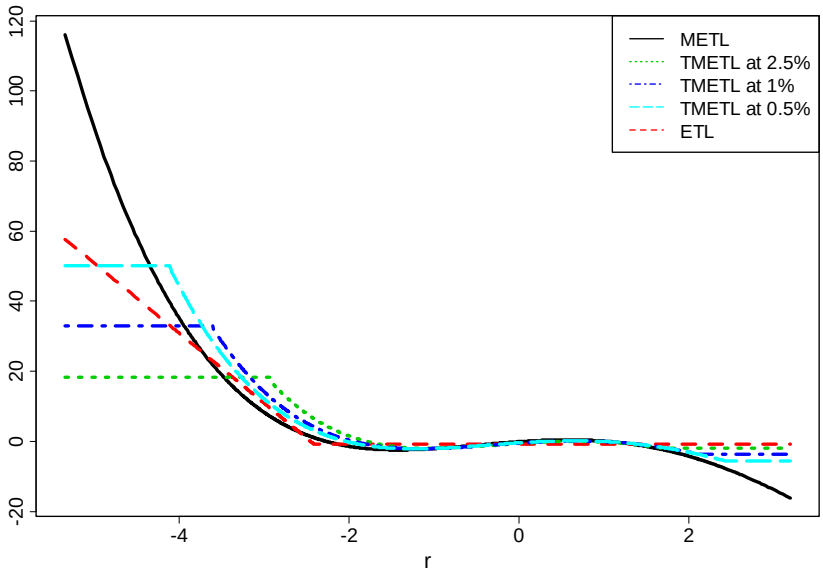

Figure 4.5 Influence Functions of METL and TMETL Compared to ETL (red dashed lines) under skewed-t distribution with $df = 10, \gamma = 0.8$.

We note the following relative behaviors of ETL, METL and TMETL:

1. The influence function of ETL is linearly increasing for values above the corresponding VaR and is essentially zero otherwise. The returns on the left tail

have increasing influence on ETL as they go deeper into the tail. On the other hand, the returns on the right tail have negligible impact, which is quite reasonable.

2. METL is a function of skewness and kurtosis. These are both highly influenced by large positive as well as large negative returns, as is revealed in the skewness and kurtosis influence functions in chapter 2. Thus the influence function of METL increases as the value of a return moves farther out in the left tail, while the opposite happens as a return moves farther out in the right tail. This is the results from the interaction of skewness and kurtosis terms in the METL definition.
3. The influence function of TMETL is similar in shape or more or less in value for returns that are not too large, but is bounded. For symmetric distributions the bound is considerable larger for negative outlier returns than for positive outlier returns, and relatively close to zero for positive outlier returns.
4. For student t-distribution with df=4, there is no finite variance for either skewness and kurtosis. Therefore the influence function for METL does not exist. However, the influence function TMETL still exists.
5. For a positively skewed distribution (Fig. 4.4), the influence for METL is much larger (more negative) for the positive outliers. So is the bound for the influence function for TMETL. For a negatively skewed distribution (Fig. 4.5), the opposite is true.

We also note the following behaviors of ETL, METL and TMETL relative to VaR, MVaR and TMVaR, respectively.

1. The influence function for VaR is a jump function with jump at the cutting-off point. The influence function of ETL is more continuous. They vanish at the same point,

but the influence for ETL to the left of this point is linearly growing whereas that for VaR is a constant. In this regard, both MVaR and METL are closer to ETL than VaR. Their influence functions resemble the one for ETL on the left tail. The difference is their influence grows non-linearly and faster in the tail.

2. Both MVaR and METL values will be reduced by a large positive return. This is a feature that ETL does not have. This can be a harmful feature for risk measures. Ideally a change in the positive side should not affect the downside risks.
3. TMVaR and TMETL behave similarly, too. They both grow in the left tail and are bounded on both tails.

4.2 Finite Sample Variance

We now compare the approximate sample standard deviations of ETL, METL and TMETL under different distributions at sample sizes 20, 30, 60, 120 and 250. Sample size 20 and 30 are used to assess how good the approximate short histories of fund-of-hedge fund returns for monthly data, while sample size 60 and 120 would be for five and ten years of monthly returns respectively, and sample size 250 would represent a year of daily returns. The tables 4.1 to 4.4 below show the results. The Monte Carlo simulation has 10,000 replicates for each sample size. For each replicate, a sample of the specific sample size is randomly generated using the particular underlying distribution. The sample moments are calculated and used in METL and TMETL calculation. The asymptotic approximation is derived using the influence function approach (see for example, Hampel et. al., 1986)

$$Var_{\infty}\left(\theta\right) = E\left(IF^{2}\left(x;\theta,F\right)\right). \tag{4.5}$$

The influence functions are computed using equation (4.4) and then numerically integrated using the *R* routine *integrate*, which is based on QUADPACK (see Piessens

et al., 1983) to estimate the asymptotic variance expression in (4.5) above. The standard errors are calculated using the same method as in section 3.2.

The simulation results approximately match the asymptotic results at the larger sample sizes (120 or 250). The results are closer when the nominal distribution is closer to a normal distribution. The asymptotic results for METL and TMETL overestimate the simulation results across board, but underestimate the simulation results for ETL. This is probably because METL and TMETL sample estimates are further from normally distributed than ETL sample estimates for the same sample size. The only exception is the student t-distribution with $df = 4$. This is when we do not have finite variances for skewness and kurtosis. The numerically integrated asymptotic results are unreliable. We can still see the simulation results show a big jump compared to the results for $df = 10$. The variance of sample METL estimate does not decrease as the sample size increases. whereas TMETL and ETL estimates do.

Below the sample size 120, the discrepancies between in asymptotic results and the simulation results are even larger. In addition to the reasons previously stated, for small sample sizes, we are effectively “trimming” on both tails. We note that for sample sizes 20 and 30 where there are no data points to be trimmed at 2.5% trimming, METL and TMETL are identical estimates.

Table 4.1 Standard Deviations of Sample ETL, METL and TMETL under standard normal distribution

Standard deviations of Sample METL						
	Asymptotic Approximation			*Simulation*		
Sample Size	*METL*	*TMETL*	*ETL*	*METL*	*TMETL*	*ETL*
20	0.6712 (0.0219)	0.5395 (0.0094)	0.4947 (0.0194)	0.4268 (0.0033)	0.4268 (0.0033)	0.5219 (0.004)
30	0.5481 (0.0179)	0.4405 (0.0077)	0.4039 (0.0159)	0.3721 (0.003)	0.3721 (0.003)	0.3978 (0.003)
60	0.3875 (0.0127)	0.3115 (0.0054)	0.2856 (0.0112)	0.2843 (0.0023)	0.2363 (0.0017)	0.3156 (0.0023)
120	0.274 (0.0089)	0.2203 (0.0038)	0.2019 (0.0079)	0.2049 (0.0016)	0.1593 (0.0011)	0.2205 (0.0015)
250	0.1899 (0.0062)	0.1526 (0.0027)	0.1399 (0.0055)	0.1458 (0.0011)	0.1115 (8e-04)	0.1527 (0.0011)

Table 4.2 Standard Deviations of Sample ETL, METL and TMETL under Student-t distribution with $df = 10$ **.**

Standard deviations of Sample METL						
	Asymptotic Approximation			*Simulation*		
Sample Size	*METL*	*TMETL*	*ETL*	*METL*	*TMETL*	*ETL*
20	*1.796 (0.0535)*	*0.6953 (0.0129)*	*0.7147 (0.0305)*	*0.6688 (0.0193)*	*0.6688 (0.0193)*	*0.7542 (0.0084)*
30	*1.4664 (0.0437)*	*0.5677 (0.0105)*	*0.5836 (0.0249)*	*0.6102 (0.0116)*	*0.6102 (0.0116)*	*0.5656 (0.0052)*
60	*1.0369 (0.0309)*	*0.4014 (0.0075)*	*0.4126 (0.0176)*	*0.5335 (0.0086)*	*0.3174 (0.0027)*	*0.4717 (0.004)*
120	*0.7332 (0.0218)*	*0.2838 (0.0053)*	*0.2918 (0.0124)*	*0.4483 (0.0076)*	*0.2055 (0.0015)*	*0.3372 (0.0026)*
250	*0.508 (0.0151)*	*0.1966 (0.0037)*	*0.2022 (0.0086)*	*0.3683 (0.0083)*	*0.1419 (0.001)*	*0.2299 (0.0017)*

Table 4.3 Standard Deviations of Sample ETL, METL and TMETL under Student-t distribution with $df = 4$.

Standard deviations of Sample METL						
	Asymptotic Approximation			*Simulation*		
Sample Size	*METL*	*TMETL*	*ETL*	*METL*	*TMETL*	*ETL*
20	*2.1684 (0.1074)*	*0.7027 (0.013)*	*0.8082 (0.0407)*	*2.0869 (0.1396)*	*2.0869 (0.1396)*	*1.5664 (0.0414)*
30	*1.7705 (0.0877)*	*0.5737 (0.0106)*	*0.6599 (0.0332)*	*2.0992 (0.1101)*	*2.0992 (0.1101)*	*1.1245 (0.0245)*
60	*1.252 (0.062)*	*0.4057 (0.0075)*	*0.4666 (0.0235)*	*1.7305 (0.0306)*	*0.6071 (0.0104)*	*0.9868 (0.0379)*
120	*0.8853 (0.0438)*	*0.2869 (0.0053)*	*0.3299 (0.0166)*	*2.3753 (0.2352)*	*0.3259 (0.0029)*	*0.6944 (0.0091)*
250	*0.6133 (0.0304)*	*0.1987 (0.0037)*	*0.2286 (0.0115)*	*2.5707 (0.1964)*	*0.2226 (0.0018)*	*0.4804 (0.0053)*

Table 4.4 Standard Deviations of Sample ETL, METL and TMETL under skewed-t distribution with $df = 10, \gamma = 1.25$.

Standard deviations of Sample METL						
	Asymptotic Approximation			*Simulation*		
Sample Size	*METL*	*TMETL*	*ETL*	*METL*	*TMETL*	*ETL*
20	*1.0516 (0.039)*	*0.5914 (0.0083)*	*0.5758 (0.0244)*	*0.696 (0.0282)*	*0.696 (0.0282)*	*0.6183 (0.0069)*
30	*0.8586 (0.0318)*	*0.4829 (0.0068)*	*0.4701 (0.0199)*	*0.5895 (0.021)*	*0.5895 (0.021)*	*0.4582 (0.0045)*
60	*0.6071 (0.0225)*	*0.3414 (0.0048)*	*0.3324 (0.0141)*	*0.4864 (0.0203)*	*0.2447 (0.0024)*	*0.3718 (0.0031)*
120	*0.4293 (0.0159)*	*0.2414 (0.0034)*	*0.2351 (0.01)*	*0.3127 (0.0091)*	*0.1596 (0.0012)*	*0.2687 (0.0021)*
250	*0.2974 (0.011)*	*0.1673 (0.0024)*	*0.1629 (0.0069)*	*0.2387 (0.0067)*	*0.112 (8e-04)*	*0.185 (0.0014)*

We also compared the sample standard deviations of the TMETL and ETL to the METL under different distributions. The following table shows the results under Monte Carlo simulations with the sample size of 120. It shows that ETL has smaller standard deviation than the METL in general. The TMETL unequivocally reduces the sample standard deviation and does the best job in the heavy tailed case.

Table 4.5 Standard Deviations of TMETL and ETL as a Percentage of Standard Deviation of METL.

	TMETL/METL	ETL/METL
Standard Normal	78%	108%
Student-t, $df = 4$	14%	29%
Student-t, $df = 10$	46%	75%
Skewed-t $df = 10, \gamma = 1.25$	51%	86%

Conclusions

METL can be unstable and very misleading when outliers are present or the nominal distribution is far from being normal. Trimmed version is a big help in such situations. TMETL helps to stabilize the variability of its estimates just like TMVaR did for TMETL. Our analysis of influence functions shows that METL behaves very similar to MVaR. They no longer have the sharp contrast between the influence functions for ETL and VaR.

4.3 Portfolio Optimization with Modified ETL

In this section, we repeat the experiments in Chapter 3 with risk measures from the ETL family.

4.3.1 Empirical Definitions of Risk Measures

The empirical estimate of METL at the α confidence level is defined as follows

$$\begin{aligned}\widehat{\mathrm{mETL}}(1-\alpha) &\triangleq -\hat{\mu} - \hat{\sigma}\cdot E_G \\ &= -\hat{\mu} + \frac{1}{\alpha}\hat{\sigma}\cdot\phi\left(\delta_{1-\alpha}\right) \\ &+ \frac{1}{\alpha}\hat{\sigma}\cdot\left(\begin{array}{c}\frac{1}{24}\left[I^4 - 6I^2 + 3\phi\left(\delta_{1-\alpha}\right)\right]\cdot\hat{K} + \frac{1}{6}\left[I^3 - 3I^1\right]\cdot\hat{S} \\ + \frac{1}{72}\left[I^6 - 15I^4 + 45I^2 - 15\phi\left(\delta_{1-\alpha}\right)\right]\cdot\hat{S}^2\end{array}\right)\end{aligned} \tag{4.5}$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution and $I(\cdot)$ is defined as in (4.3).

The TMETL uses the same empirical definition in (4.5) and replace the skewness and kurtosis parameters using their trimmed versions as in 4.3.2

4.3.2 Optimization Methods

The portfolio ETL is a convex measure and can be solved by linear programming. It can be solved via

$$\begin{aligned} \text{minimize} \quad & \eta + \frac{1}{T(1-\alpha)} \sum_{i=1}^{T} s_i \\ \text{s.t.} \quad & s_i \geq 0 \\ & s_i \geq -\eta - \mathbf{w}^T \mathbf{r}_i \\ & \mathbf{w}^T \mathbf{1} = 1, w_j \geq 0,\ j = 1..n \end{aligned} \tag{4.4}$$

η is a free parameter but equals VaR of $1-\alpha$ confidence level in the solution of (4.4). In our experiments, $\alpha = 0.95$. Since both the objective and the constraints are all linear functions, this is a linear programming problem. See Rockafellar and Uryasev (2002) for details.

Modified ETL and its trimmed version lose convexity. They belong to nonlinear programming. We follow the same routine we discussed in section 3.3.3. We use the non-linear programming method developed by P. Spellucci (1997) to solve this problem. It uses the Sequential Quadratic Programming (SQR) method for optimization. In order to have a higher probability of finding the global optima, we use multiple starting points, which include the ETL solution from (4.4), minimum volatility portfolio solution as well as

single asset solution, i.e. unit weight on one asset and zero weight on all other assets. The total number of starting points is 22. The average computation time for a single period optimization for ETL, METL and TMETL are 8 seconds, 3 minutes and 5 minutes, respectively on a 1.8GHz Intel CPU.

4.4 Results

4.4.1 Data

See section 3.4.1 for details.

4.4.2 Experiments with Long-Only Constraints

In this experiment, we enforce the long only and full investment constraints. We have compared Minimum volatility, Maximum Sharpe Ratio, VaR and MVaR in chapter 3. We found VaR led by a large margin in various measures. In this experiment we will focus on the following 3 methods ETL (ETL), Modified ETL (METL) and Trimmed METL (TMETL). The names in the parenthesis are the code names we use in the plots and tables.

In Table 4.6 and 4.7, we display the summary statistics on respective out-of-sample performance and risk measures for each portfolio optimization method across all 100 portfolios we formed. In Fig. 4.6 and 4.7, we display the kernel densities of these statistics. Some outliers are removed for better presentation (see Table 4.7). In Fig. 4.8 and 4.9, we display the box-plots of these statistics of the $\binom{3}{2}=3$ paired differences between ETL, METL and TMETL on each portfolio. For better presentation, some of the outliers outside the plotting region are not shown.

Table 4.6 Summary Statistics of Performance Measures of Long Only Experiments

Annualized Mean Return				Diversification		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	-0.061	-0.730	-0.033	0.615	0.480	0.644
25%	0.053	0.025	0.028	0.748	0.736	0.761
Median	0.107	0.066	0.069	0.782	0.780	0.810
Mean	0.108	0.057	0.075	0.772	0.766	0.796
75%	0.155	0.140	0.117	0.807	0.807	0.832
Maximum	0.415	0.585	0.257	0.856	0.857	0.869
Sharpe Ratio				Starr Ratio		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	-0.594	-0.893	-0.527	-0.067	-0.094	-0.057
25%	0.211	-0.031	-0.002	0.026	-0.004	0.000
Median	0.558	0.309	0.355	0.079	0.033	0.042
Mean	0.647	0.375	0.413	0.084	0.055	0.052
75%	1.010	0.826	0.698	0.135	0.104	0.085
Maximum	2.420	1.980	1.860	0.337	0.400	0.229

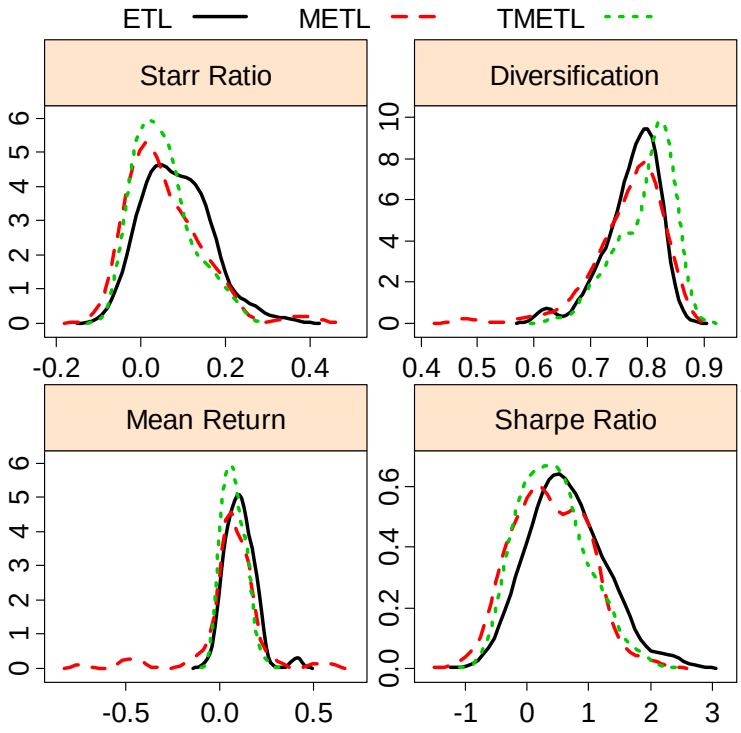

Figure 4.6 Kernel Densities of Performance Measures of Long Only Experiments

Table 4.7 Summary Statistics of Risk Measures of Long Only Experiments

Annualized Volatility				Semi. Std. Dev.		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	0.066	0.066	0.069	0.072	0.078	0.078
25%	0.098	0.107	0.096	0.104	0.114	0.108
Median	0.113	0.132	0.112	0.126	0.146	0.129
Mean	0.133	0.214	0.119	0.127	0.405	0.133
75%	0.137	0.162	0.127	0.144	0.177	0.154
Maximum	0.723	1.370	0.364	0.228	5.670	0.282
Value at Risk				Expected Tail Loss		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	0.011	0.015	0.017	0.035	0.039	0.039
25%	0.034	0.040	0.038	0.058	0.066	0.061
Median	0.046	0.052	0.046	0.073	0.091	0.081
Mean	0.045	0.054	0.048	0.076	0.147	0.080
75%	0.055	0.066	0.058	0.089	0.113	0.095
Maximum	0.096	0.107	0.102	0.147	1.340	0.167
Maximum Drawdown				Turnover		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	0.075	0.069	0.076	0.064	0.109	0.092
25%	0.169	0.224	0.212	0.095	0.167	0.125
Median	0.247	0.313	0.284	0.105	0.195	0.148
Mean	0.240	0.338	0.278	0.108	0.206	0.152
75%	0.307	0.384	0.337	0.120	0.231	0.174
Maximum	0.435	0.981	0.517	0.155	0.383	0.237

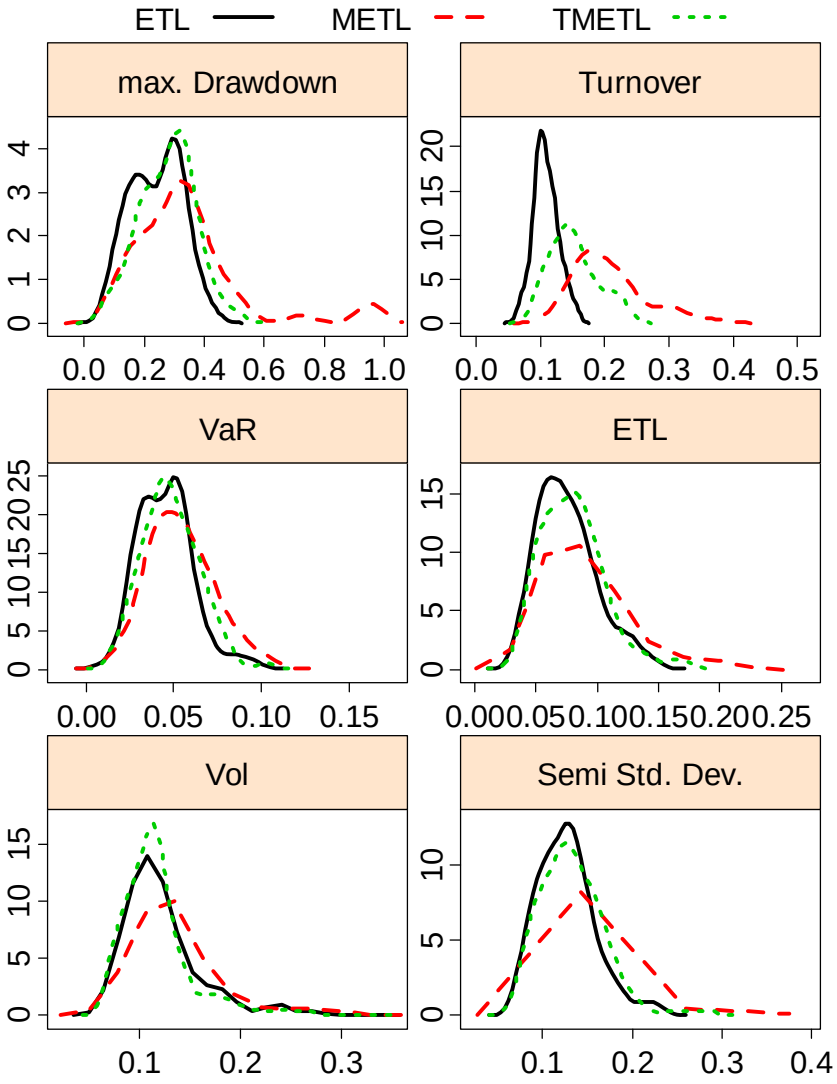

Figure 4.7 Kernel Densities of Risk Measures of Long Only Experiments

Combining results from these analyses, we observe the following facts:

1. ETL outperforms the others in all performance measures. METL has both extreme bad and good portfolios, but on average it is the worst performer in the group. TMETL is in between.
2. METL shows an unusual amount of out-of-sample volatility, especially on the high volatility side. This is caused by one huge negative return in several portfolios. This reflects badly in almost all out-of-sample risk measures, except out-of-sample VaR. Therefore such breakdowns only occurred with very low probability. Nevertheless this dramatically changed METL's overall performance.
3. TMETL shows slightly bigger out-of-sample risk than ETL, but still is much better than METL for all risk measures.

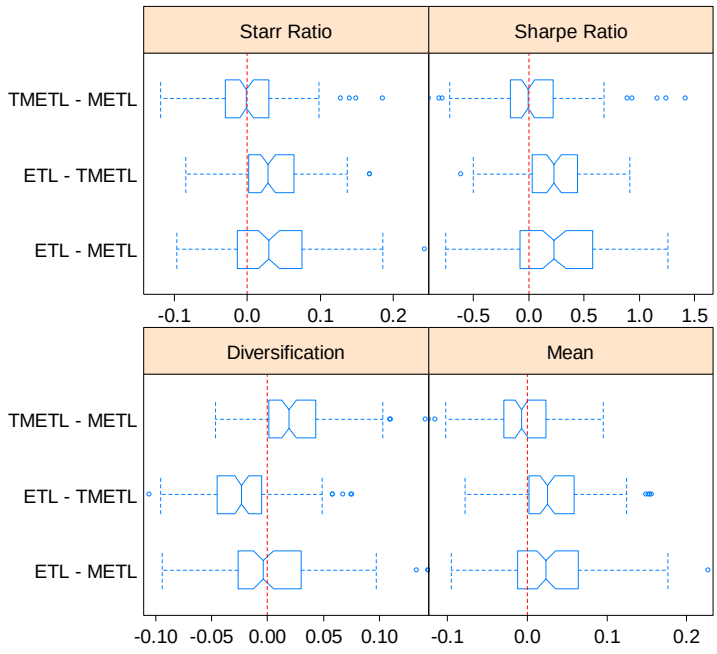

Figure 4.8 Notched Box-plots of Paired Difference for Performance Measures: ETL family

In summary, the ETL method shows better control of risk on most aspects. It is less volatile on both upside and downside. The METL method suffers from bigger volatility than the ETL. This is the same as we have observed within the MVaR method in Chapter 3. TMETL helps to stabilize the METL method.

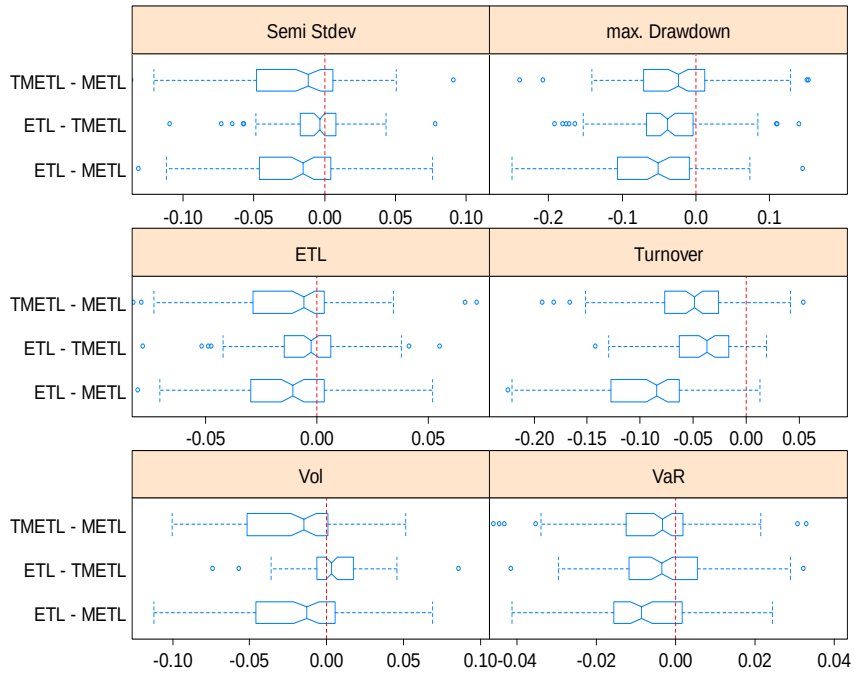

Figure 4.9 Notched Box-plots of Paired Difference for Risk Measures: ETL family

Comparing with results in Chapter 3, METL is still better than MVaR in terms of risk control. We display the paired difference of performance and risk measure box-plots in Fig. 4.10 and 4.11 for the ETL and VaR family.

The performance measures do not show significant difference for Sharpe and Starr ratios. ETL obtains better return than VaR. TMETL and METL both show better diversification than TMVaR and MVaR respectively. In terms of risk measures, ETL family clearly does better than the VaR family in every aspect.

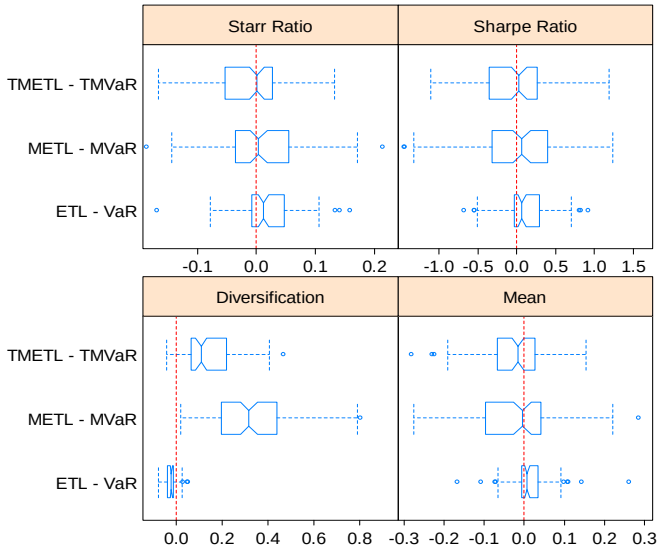

Figure 4.10 Notched Box-plots of Paired Difference for Performance Measures: ETL family vs. VaR family

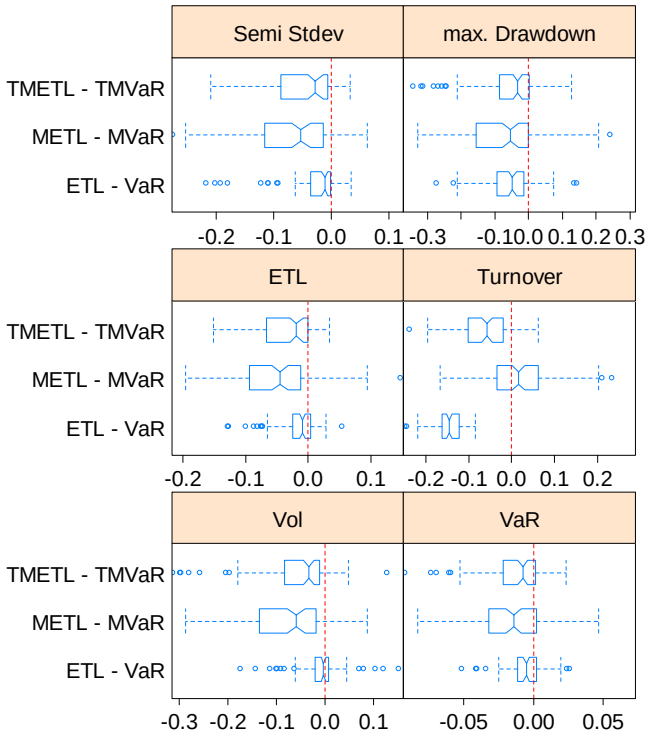

Figure 4.11 Notched Box-plots of Paired Difference for Risk Measures: ETL family vs. VaR family

Below we show the quantile cumulative return plot Fig. 4.12, which displays the cross-sectional cumulative returns of the given quantile for a particular method (see explanation also in chapter 3). It shows that there are some erratic returns happened for the METL. We find 0.15% of all out-of-sample period returns have absolute values that are greater than 1 on log scale, i.e. 172% or -61% return in one month. METL is heavily impacted by such returns. For a training period that contains a very large positive outlier, METL is grossly underestimated (see section 4.4.4). METL method mistakenly concentrates all its investment in one asset and subsequently is hammered by a large negative return. This is exactly what happened in the period of 04/2007. By trimming one point on each tail, we also can avoid such big. As a matter of fact, ETL and TMETL methods almost completely avoided this asset for this period.

The mean cumulative return of the ETL, METL and TMETL for all 100 portfolios is shown in Fig. 4.13. METL and TMETL both had more dramatic losses in the financial crisis than the original ETL. However, METL makes returns gains faster than TMETL in the recovery period. This shows that the upside trimming may have removed the early signal of recovery for TMETL. It is a dilemma that trimming can be useful to avoid dramatic breakdowns but also blocks the useful signal sometimes.

The TMETL under our setting is only trimming one point on either end of the distribution. In the order to see trimming at different levels, we did the same portfolio optimization with a smaller confidence level of 75%, i.e., larger tail probability of 0.25 and used symmetric 5% trimming, i.e. trim the three largest and the three smallest portfolio returns. The results are shown in Fig. 4.14. Because we are taking a larger tail probability, we are no longer measuring extreme tail risks. The trimming method is less helpful in this case. This is reflected in the closer performance between ETL and METL. The TMETL is

less underperforming in the better performing group of portfolios, especially before the crisis.

Figure 4.12 Quantile Cumulative Return Plots for the ETL family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. ETL are all set at 95% confidence level and trimming is for 2.5% of either tail

Figure 4.13 Mean Cumulative Return Plots for the ETL family. Each curve represents the mean of the cumulative returns across all 100 portfolios at the different time. ETL are all set at 95% confidence level and trimming is for 2.5% of either tail.

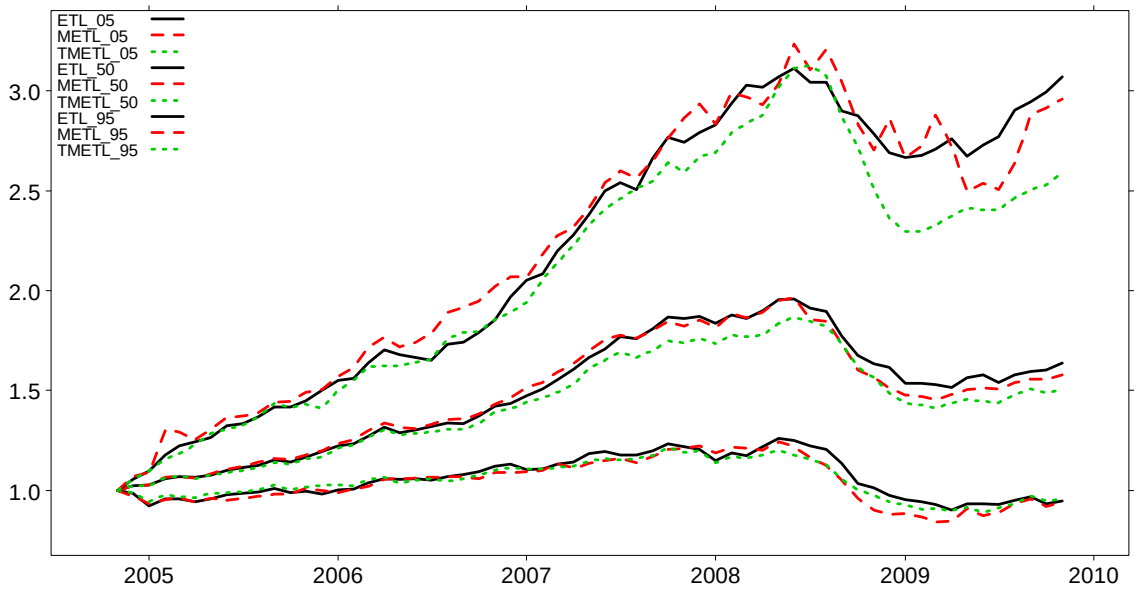

Figure 4.14 Quantile Cumulative Return Plots for the ETL family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. ETL are all set at 75% confidence level and trimming is for 5% of either tail

4.4.3 Experiments with Upper Bounded Constraints on Weights

In this section we cap our investment uniformly at 20% maximum for any single hedge fund. In the last section, we find out that some methods are already diversified (Table 4.8 and 4.9) equivalently to this level on average. We'd like to explore the effect of this upper bound constraint on those methods that weren't quite diversified under the long-only constraint.

We summarize our results in the following tables and figures using the same pattern as in the previous section. METL portfolios were not well-diversified previously. After enforcing the upper bound constraint, it behaves much better. Although both METL and TMETL are still slightly less diversified than ETL, the difference is negligible. They also have slightly larger turnover rate. The rest of risk and performance measures are more or less the same across all three methods. Fig. 4.15 to 4.19 show they are almost the

same as ETL in terms of performances. They only underperform in those better-performing portfolios.

Table 4.8 Summary Statistics of Performance Measures of Upper Bounded Experiments

Annualized Mean Return				Diversification		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	-0.045	-0.041	-0.030	0.841	0.839	0.838
25%	0.045	0.027	0.024	0.852	0.857	0.858
Median	0.090	0.069	0.068	0.857	0.866	0.866
Mean	0.094	0.075	0.071	0.857	0.865	0.865
75%	0.129	0.107	0.103	0.862	0.873	0.872
Maximum	0.500	0.471	0.251	0.876	0.886	0.889
Sharpe Ratio				Starr Ratio		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	-0.658	-0.705	-0.548	-0.066	-0.087	-0.056
25%	0.201	-0.012	-0.041	0.022	-0.002	-0.004
Median	0.511	0.390	0.399	0.071	0.045	0.049
Mean	0.527	0.384	0.370	0.073	0.051	0.048
75%	0.853	0.745	0.704	0.117	0.099	0.089
Maximum	2.090	1.700	1.390	0.339	0.361	0.220

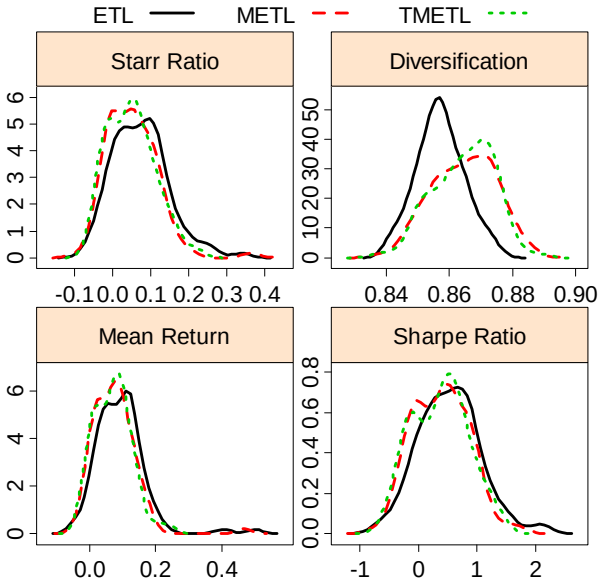

Figure 4.15 Kernel Densities of Performance Measures of Upper Bounded Experiments

Table 4.9 Summary Statistics of Risk Measures of Upper Bounded Experiments

Annualized Volatility				Semi. Std. Dev.		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	0.073	0.076	0.070	0.074	0.078	0.074
25%	0.094	0.099	0.098	0.104	0.108	0.104
Median	0.108	0.112	0.110	0.115	0.123	0.126
Mean	0.131	0.125	0.117	0.121	0.129	0.126
75%	0.126	0.130	0.122	0.133	0.147	0.144
Maximum	0.977	0.821	0.400	0.197	0.251	0.217
Value at Risk				Expected Tail Loss		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	0.019	0.019	0.020	0.038	0.038	0.037
25%	0.038	0.039	0.039	0.056	0.061	0.060
Median	0.045	0.048	0.049	0.068	0.072	0.073
Mean	0.046	0.050	0.049	0.070	0.077	0.075
75%	0.054	0.060	0.059	0.078	0.091	0.089
Maximum	0.082	0.087	0.081	0.134	0.151	0.144
Maximum Drawdown				Turnover		
	ETL	METL	TMETL	ETL	METL	TMETL
Minimum	0.086	0.109	0.127	0.064	0.109	0.092
25%	0.200	0.223	0.225	0.095	0.167	0.125
Median	0.260	0.289	0.283	0.105	0.195	0.148
Mean	0.252	0.284	0.280	0.108	0.206	0.152
75%	0.304	0.341	0.329	0.120	0.231	0.174
Maximum	0.459	0.449	0.444	0.155	0.383	0.237

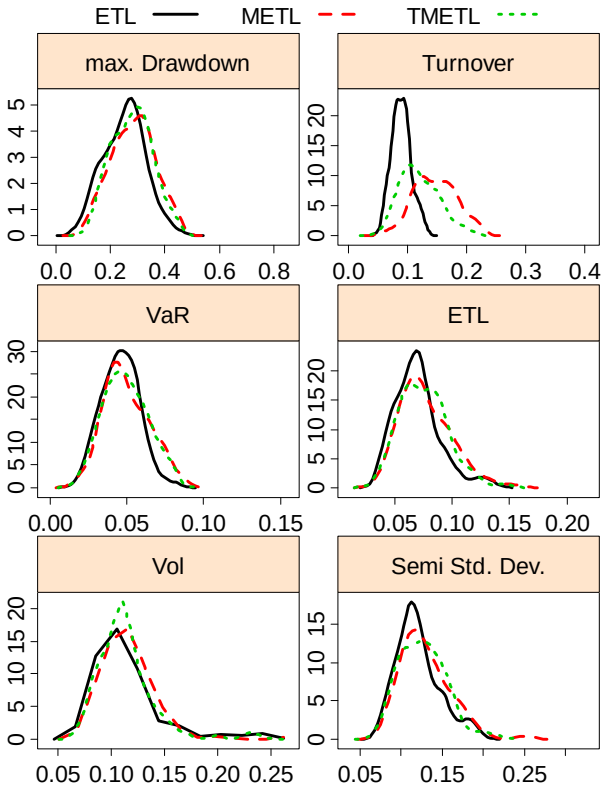

Figure 4.16 Kernel Densities of Risk Measures of Upper Bounded Experiments

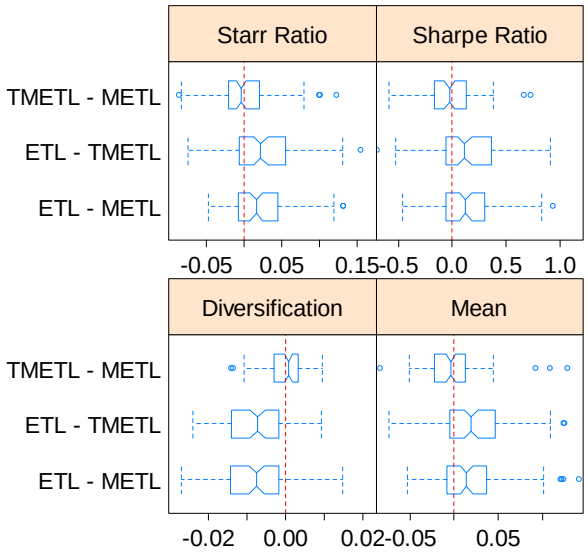

Figure 4.17 Notched Box-plots of Paired Difference for Performance Measures: ETL family, Upper bounded experiment

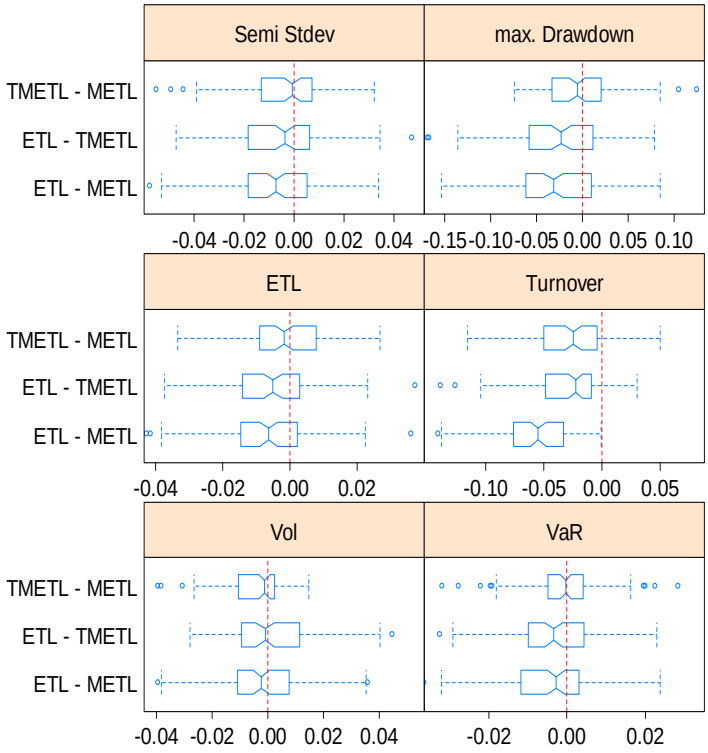

Figure 4.18 Notched Box-plots of Paired Difference for Risk Measures: ETL family, Upper bounded experiment

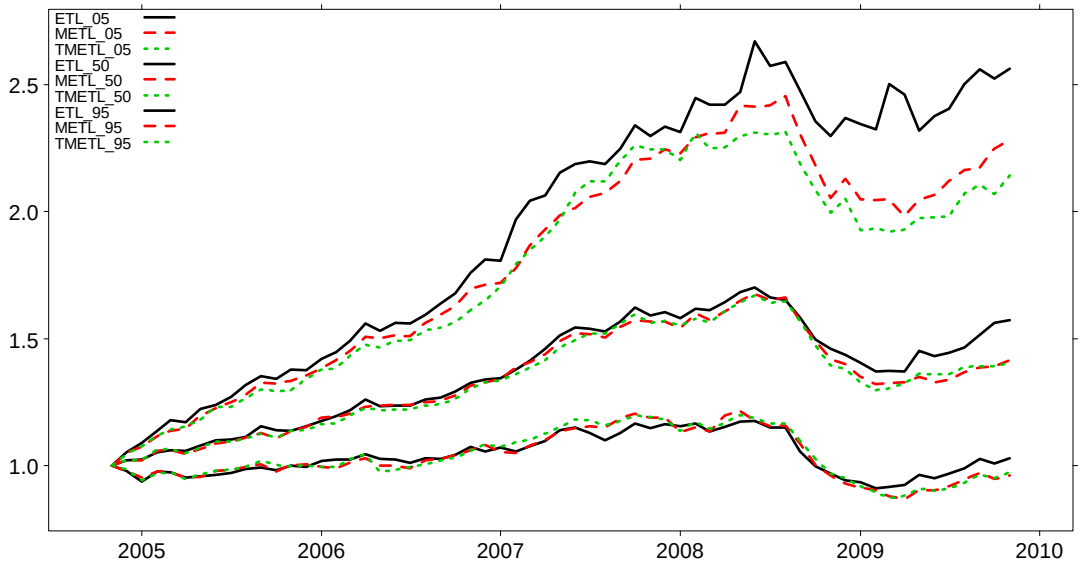

Figure 4.19 Quantile Cumulative Return Plots for the ETL family. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. ETL are all set at 95% confidence level and trimming is for 2.5% of either tail.

4.4.4 Alternative Robust Methods for METL

As seen previously, for a training period that contains a very large positive outlier, METL is grossly underestimated. The plot below shows the values of the negatives of the estimated ETL (-0.26), METL (0.837) and TMETL (-0.586, only trimmed for skewness and kurtosis estimates) superimposed on a kernel density estimate plot of 60 monthly returns in the training period for Hedge fund A (HF-A).

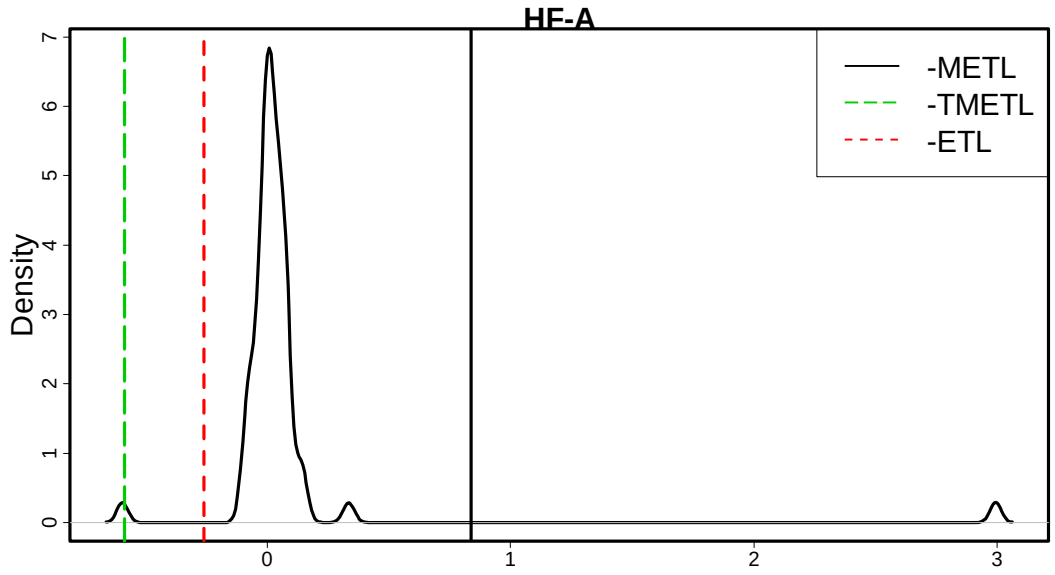

Figure 4.20 Negatives of ETL, METL and TMEL for HF-A.

It's apparent that METL is completely misleading. We note that for a risk measure $\rho(\cdot)$ that measures the tail loss, a basic requirement (see Rachev, Stoyanov, Fabozzi, 2008) is that it should be no larger than the mean return, i.e. $\rho(r) \le E[r]$. METL in this case fails to have this property (mean is 0.06). The large positive outlier has created very large kurtosis and positive skewness such that the negative influence on METL is large enough to make its value negative and hence the negative of the METL is positive as shown in the above figure. It is debatable whether ETL or TMETL is the more accurate

estimate of the average tail loss. TMETL in this case is curiously the more conservative risk estimate larger than ETL and hence is a more conservative estimate of risk than ETL. This is caused by the outlier's impact on untrimmed standard deviation. We will explain in more details in the next section.

The next monthly return right after this 60-month training period is a huge negative return (-94%). Because METL has such an optimistic view of the risk, the minimum METL portfolio based on the preceding training period allocates 100% weight to this asset and is hammered by the negative return during the next month. On the other hand, ETL and TMETL only assigned 0% and 0.2%, respectively, to this asset.

In the first section, we showed the influence functions for trimmed METL by trimming mean, standard deviation, skewness and kurtosis. However, the TMETL method we have been using so far for empirical study is based on trimming only for skewness and kurtosis (TMETL). The following is a plot comparing the influence functions for different trimming combinations. All the methods trim 2.5% on each tail for the standard normal distribution.

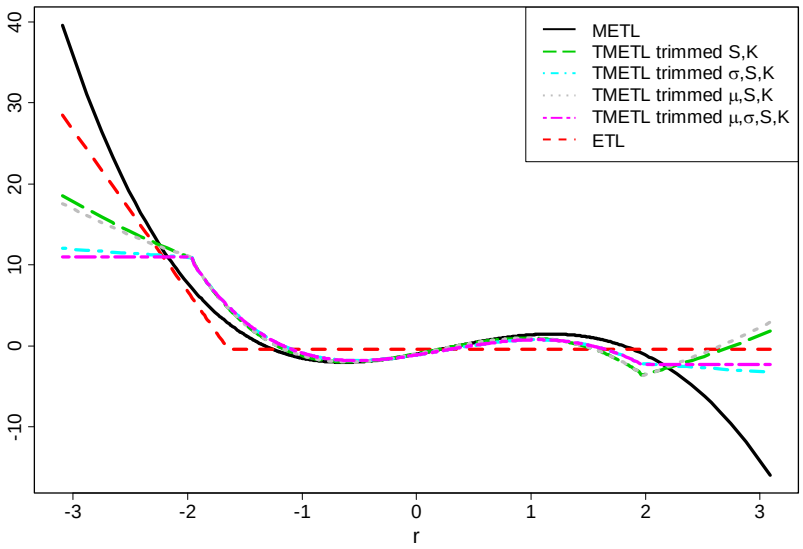

Figure 4.21 Influence functions for different combination of trimming methods: Standard Normal Distribution

The influence function for TMETL, when trimmed data is used in estimation of the mean, SD, skewness and kurtosis (Complete Trimmed METL, CTMETL) is constant beyond the trimming sites (as previously shown in Section 4.1), i.e., all sufficiently large outliers have the same influence. If we only trim skewness and kurtosis, the influence of large outliers on mean and SD will not be constant beyond where we trimmed. As the outlier goes further from the mean, it will always lead to larger risk estimate, no matter which tail this outlier is on. This is mainly the effect of untrimmed standard deviation. A very large loss would result in elevated risk, although the impact is not as large as the untrimmed version. A very large gain would also result in elevated risk if it is above where we trim on the right tail. However, if the gain is moderate, i.e. to the left of where we trim on the right tail, it results in decreased risk, just as METL does. This may have some merit, as we may reward moderate gains, but extreme large profit may be seen as a potential risk. Trimming on mean does not have a large impact on the estimate as trimming on the other moment estimates.

Our method of dealing with outliers via trimming is an “after the fact” method in the sense that method deals with one-dimensional outliers in the portfolio returns rather than directly in the multivariate asset returns. Boudt et. al. (2008) proposed a different way of limiting the influence of outliers by shrinking multivariate outliers in the multivariate returns. In their method, the mean vector and the covariance matrix are estimated using the subset of data that minimizes ellipsoid volume occupied by the data as in Rousseeuw (1985). Every return vector is subsequently ranked by its Mahalanobis distance $d_t^2 = \left(\mathbf{r}_t - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{r}_t - \boldsymbol{\mu}\right)$, which for a multivariate normal distribution with known mean vector and covariance, d_t^2 follows a chi-squared distribution with n degrees of freedom where n is the number of assets in the portfolio. The distances are often assumed to be approximately chi-squared with n degrees of freedom when the parameters are estimated and the returns are non-normal. Consequently, each return vector can be checked against larger of $\chi^2_{n,0.975}$ and the maximum distance at the cut-off point $d^2_{\lfloor(1-\alpha)T\rfloor}$. Those vectors with d_t that exceed such a threshold are flagged as outliers. They are subsequently scaled back to where the threshold is (see Boudt et. al. 2008 for details). The cleaning method is provided by Boudt and his colleagues in the *R* package *PerformanceAnalytics*.

Now we revisit the previous example of hedge fund A. The plot below shows the estimates of ETL by each method stated above.

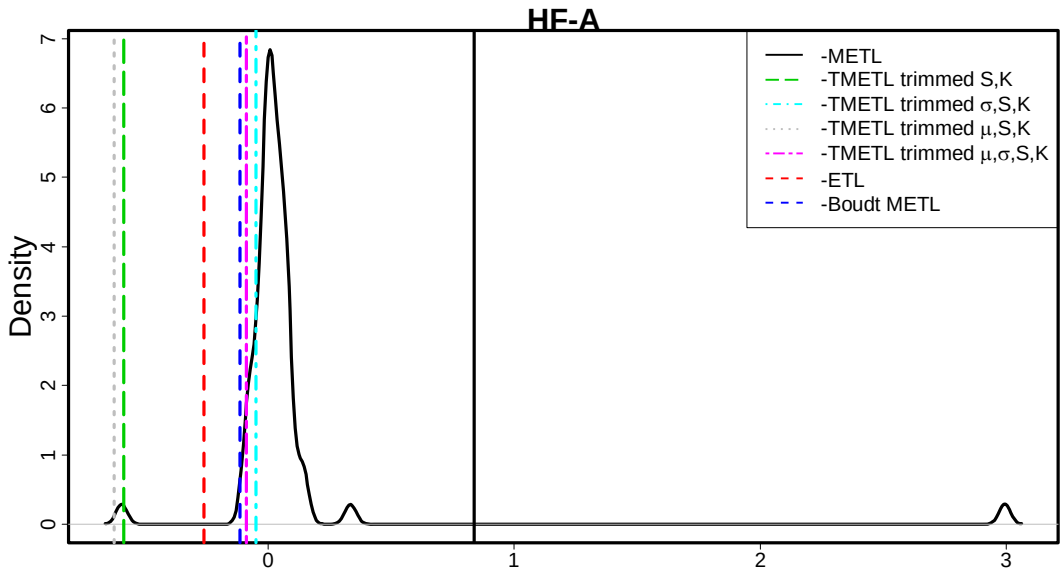

Figure 4.22 Negatives of ETL, METL and different combination of trimmed MEL for HF-A.

The METL for hedge fund A after applying the cleaning method used by Boudt (BMETL) is close to CTMETL method. Both are working on the cleaned/trimmed data for all estimates. The TMETL with only skewness and kurtosis trimmed gives the much larger estimate as shown previously. The real return for hedge fund A is -2.87 for the next period, far left out of the scope of the plot above. It seems TMETL with only skewness and kurtosis is the closest to the out-of-sample “truth” in this particular case.

The differences between trimmed ETL estimates and the Boudt cleaning method:

1. Trimming is a one dimension method after combining assets returns by the weight vector. Therefore it has to be done after weights are assigned. This has increased average optimization time for a 20-asset portfolio with 60 training periods from 3 minutes to 5 minutes on a 1.8 GHz CPU.

2. Cleaning is carried out directly on the multivariate asset returns and is done once before weights are selected. This preprocessing procedure only takes less than 1 minute.
3. Trimming removes some points from the portfolio returns. At least for the one-dimensional case, trimming having a smooth influence function of the same form as the Huber M-estimate. The Boudt et. al. (2008) cleaning method is a form of multivariate Winsorizing. It results in a discontinuous influence function. The latter is not so desirable.

Robust Portfolio Optimization Comparison

We revisit the portfolio optimization experiments for the following three robust methods: trimming with skewness and kurtosis (TMETL), complete trimming with mean, SD, skewness and kurtosis (CTMETL), and Boudt modified ETL based on their -cleaning method (BMETL). The names in the parentheses are their code names in the following tables and charts.

In Table 4.10 and 4.11, we display the summary statistics on respective out-of-sample performance measures for each portfolio optimization method across all 100 portfolios we formed. In Fig. 4.23 and Fig. 4.24, we display the kernel densities of these statistics. In Fig. 4.25 and 4.26, we display the box-plots of these statistics by comparing the difference on the same portfolio by different methods. For better presentation, some of the outliers are not shown (outside the plotting region).

Table 4.10 Summary Statistics of Performance Measures for Alternative Robust Methods

Annualized Mean Return				Diversification		
	BMETL	TMETL	CTMETL	BMETL	TMETL	CTMETL
Minimum	-0.050	-0.038	-0.066	0.614	0.647	0.567
25%	0.035	0.029	0.029	0.772	0.766	0.753
Median	0.077	0.070	0.071	0.807	0.81	0.789
Mean	0.081	0.074	0.078	0.794	0.796	0.776
75%	0.126	0.114	0.124	0.833	0.833	0.816
Maximum	0.485	0.269	0.319	0.866	0.871	0.857
Sharpe Ratio				Starr Ratio		
	BMETL	TMETL	CTMETL	BMETL	TMETL	CTMETL
Minimum	-0.572	-0.582	-0.617	-0.067	-0.061	-0.061
25%	0.062	0.006	0.013	0.008	0.001	0.001
Median	0.474	0.360	0.371	0.049	0.044	0.038
Mean	0.471	0.414	0.422	0.060	0.050	0.051
75%	0.806	0.706	0.724	0.100	0.086	0.093
Maximum	2.400	1.920	1.920	0.352	0.239	0.255

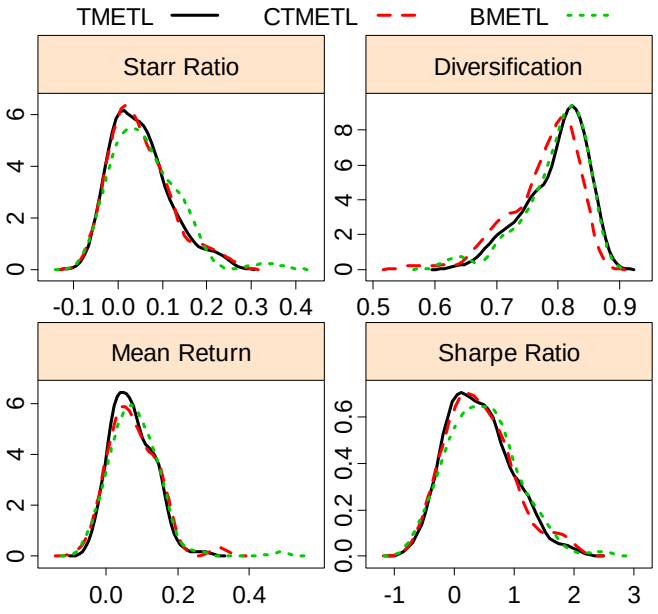

Figure 4.23 Kernel Densities of Performance Measures of Robust Methods

Table 4.11 Summary Statistics of Risk Measures for Alternative Robust Methods

Annualized Volatility				Semi. Std. Dev.		
	BMETL	TMETL	CTMETL	BMETL	TMETL	CTMETL
Minimum	0.071	0.069	0.073	0.075	0.074	0.080
25%	0.094	0.095	0.099	0.103	0.109	0.109
Median	0.110	0.112	0.114	0.126	0.128	0.133
Mean	0.121	0.116	0.127	0.130	0.131	0.144
75%	0.130	0.126	0.136	0.152	0.151	0.168
Maximum	0.857	0.330	0.542	0.269	0.252	0.344
Value at Risk				Expected Tail Loss		
	BMETL	TMETL	CTMETL	BMETL	TMETL	CTMETL
Minimum	0.013	0.016	0.016	0.038	0.035	0.035
25%	0.037	0.038	0.041	0.057	0.060	0.065
Median	0.046	0.049	0.050	0.074	0.078	0.084
Mean	0.048	0.049	0.051	0.077	0.079	0.087
75%	0.057	0.060	0.061	0.093	0.095	0.105
Maximum	0.093	0.102	0.102	0.159	0.167	0.200
Maximum Drawdown				Turnover		
	BMETL	TMETL	CTMETL	BMETL	TMETL	CTMETL
Minimum	0.082	0.076	0.092	0.088	0.083	0.091
25%	0.193	0.217	0.215	0.111	0.130	0.154
Median	0.260	0.281	0.295	0.123	0.153	0.181
Mean	0.264	0.276	0.293	0.130	0.157	0.184
75%	0.329	0.333	0.369	0.144	0.184	0.218
Maximum	0.528	0.488	0.530	0.218	0.261	0.328

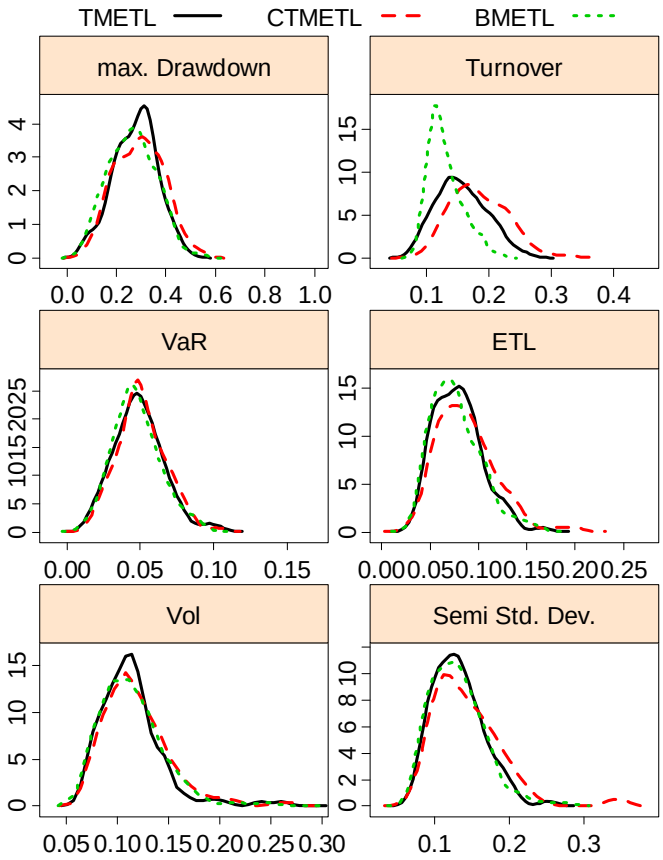

Figure 4.24 Kernel Densities of Risk Measures of Robust Methods

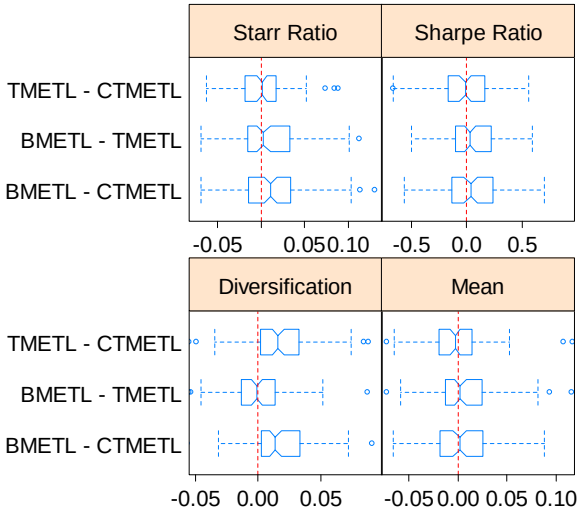

Figure 4.25 Notched Box-plots of Paired Difference for Performance Measures: Robust Methods

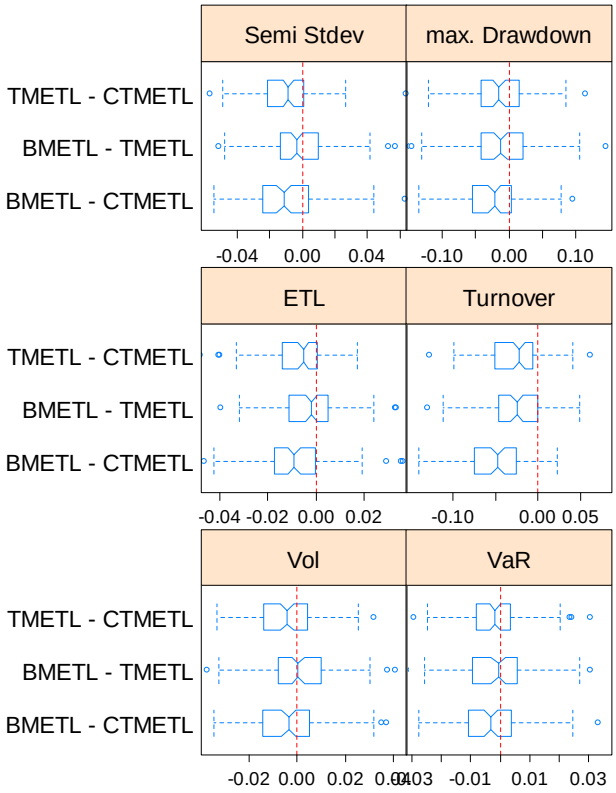

Figure 4.26 Notched Box-plots of Paired Difference for Risk Measures: Robust Methods

The three robust methods show very similar characteristics for all performance and risk measures. The BMETL method shows better Sharpe ratio and STARR ratio. Pair wise comparison indicates such difference is only marginal. CTMETL has the slightly worse diversification rate than the other two methods. Both TMETL and BMETL show smaller out-of-sample risks than CTMETL in each of the six measures on the pair wise comparison plot. BMETL shows slightly smaller out-of-sample ETL and turnover than the other two trimming methods. Overall BMETL shows the best control of risks. Between the two trimming measures, completely trimming (CMTEL) is worse than trimming only skewness and kurtosis in terms of risk control.

The following quantile cumulative return plot also shows that they all behave very similarly with CTMETL having a slight edge from about mid-2007 to late 2008. The TMETL method shows the worst performance for its 95% curve and best performance for its 5% curve during the recovery period after March 2009. This may be caused by the fact that it punishes more severely the worst and the best returns as we have seen from its influence function.

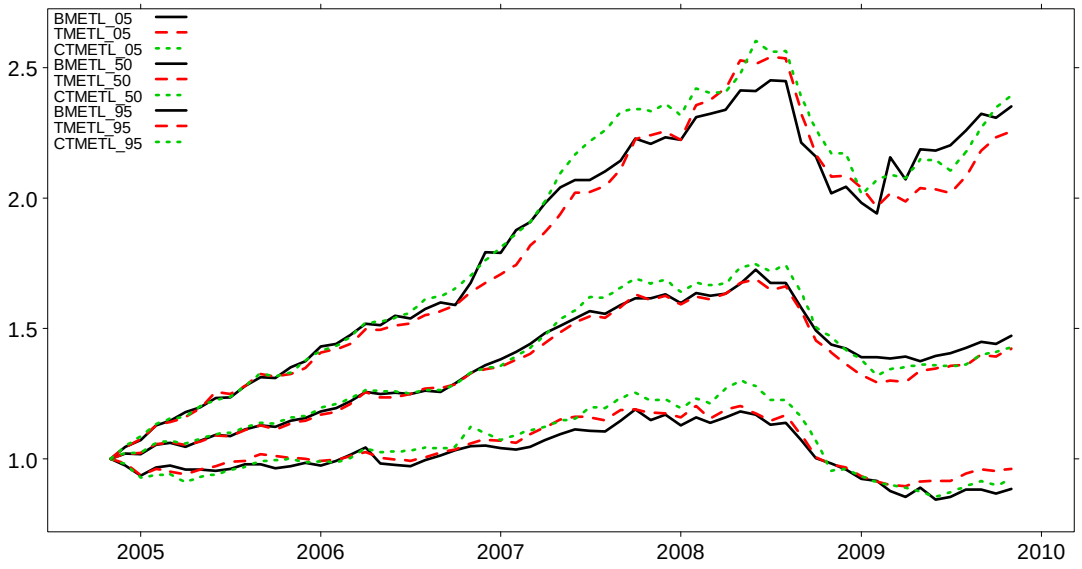

Figure 4.27 Quantile Cumulative Return Plots for the Robust Methods. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time.

Consistent with the results in Figure 4.28, the mean cumulative return of the CTMETL portfolios has a clear edge from mid-2007 to late 2008. TMETL has the worst performance during the down-turn from late 2008 through 2009. When the crash between the mid of 2008 and early 2009 are still in the training period, those with severe losses or early large recovery gains will not enter the TMETL portfolio early enough to enjoy the recovery, because both are punished as suggested by its influence function. Therefore the choice of which trimming method to use really relies on the market

environment. Their annualized relative difference is small (less than 0.7%) and highly subjective to magnitude of extreme returns in our data set.

Figure 4.28 Mean Cumulative Return Plots for the Robust Methods. Each curve represents the mean of the cumulative returns across all 100 portfolios at the different time.

STARR Ratio Optimization

In principle the portfolio manager would like to find a portfolio that optimizes the STARR ratio $\left(\mu - r_f\right)\big/METL$. However there is a basic problem that occurs even with ETL, namely ETL will be negative during extraordinary time periods when all returns are positive, and this can be the case even when a few of the smallest returns are negative but the average of the returns below VaR is positive. One can avoid this difficulty with ETL by using a centered version of ETL, i.e., a deviation measure version of ETL (see Rachev et. al., 2008). However, this solution is not guaranteed to work for METL because METL can be even larger than the mean, as we have witnessed in the first example in section. In that example given above, the resulting STARR ratio by METL is

negative, which coincidentally will not be chosen by the optimizer that seeks to maximize the STARR ratio. However, this does not mean the issue does not exist.

4.4.5 Conclusions

In this chapter we have shown METL is different from the conventional ETL. Through the study of its influence functions we find that METL is impacted by both the downside and upside of the distribution. Under some extreme circumstances when departure from normality is severe, e.g. one particular larger return, METL is dramatically underestimated and leads to heavy losses. The TMETL is much better is those cases.

In the portfolio optimization experiment we performed, we find METL and TMETL performs closely to MVaR and TMVaR, respectively. This is supported by the similarities in their influence functions. METL has a better control of risk but still not enough to prevent breakdowns. TMETL outperforms the METL in all risk measures. Upper bounded constraints can help METL without trimming.

Overall ETL still is the best performing method we find so far. The METL is a good complement, but suffers from impact of outliers. If we need to use METL, trimming method is very helpful. Among the different combination of trimming methods, we find trimming mean does not have large influence while trimming standard deviation does. Empirical experiments suggest trimming only skewness and kurtosis shows better control of risks than trimming all moments. In addition, the cleaning method developed by Boudt et. al. is also very good at defending against outliers.

Chapter 5
Portfolio Optimization with Skewed Fat-Tailed Distributions

In this chapter, we study portfolio optimization performance based on fitting parametric skewed fat-tailed distributions. For this purpose we will use the Cognity® system from *FinAnalytica*, which contains well-developed proprietary methods to fit stable distributions, skewed t-distributions and skewed-t copula. It is difficult and a huge task to implement accurate fat-tailed skewed distribution modeling including cross-sectional non linear dependence structure (Copula). Cognity® is the only commercial product to date that offers such capability. It has taken years to develop at *FinAnalytica.* We are very thankful to *FinAnalytica* to provide the Cognity® system for our research. Our study will use the same set of hedge fund returns and experimental conditions as in Chapters 3 and 4.

5.1 Parametric Model for Skewed Fat-Tailed Returns

Due to the asymmetry and fat tails in the return data, portfolios that use risk measures that rely on tail information can outperform those based on volatility by a large margin, as was shown in our previous experiments on the VaR and ETL families of risk measures in Chapters 3 and 4. Overall we found that the coherent risk measure ETL leads the group of tail risk measures with regard to optimized portfolio performance.

Both ETL based on the empirical distribution function of returns and modified ETL (METL) based on skewness and kurtosis corrections to ETL are non-parametric estimates of the unknown true ETL and METL. The question remains if we assume a fully parameterized distribution of the hedge fund returns, will that lead to a more accurate ETL, i.e., one that

more accurately reflects the skew and tail-fatness in the left tail of the portfolio returns distribution?

We provide an initial answer to this question by using parametric skewed and fat-tailed distribution fits to the returns and optimizing portfolios based on scenarios (Monte Carlo simulations) from the fitted distributions. Among the well-known models are skewed t-distributions, generalized hyperbolic distributions (Bibby and Sorensen, 2003), and stable distributions (Rachev and Mittnik, 2000, Rachev et. al., 2005).

There are a couple of choices for skewed t-distribution. The first is the skewed-t distribution defined by Fernandez and Steel in 1998. It uses one parameter to change the "tail fatness" in the regular student t-distribution as follows

$$f(x;n,\gamma)=\begin{cases} 2/\left(\gamma+1/\gamma\right)\cdot t\left(\gamma\cdot x;n\right) & \text{for } x<0 \\ 2/\left(\gamma+1/\gamma\right)\cdot t\left(1/\gamma\cdot x;n\right) & \text{for } x\geq 0 \end{cases} \tag{5.1}$$

where $t(\cdot;n)$ is the density of student-t distribution with n degrees of freedom. . It has the simple property that $f\left(\cdot;n,\gamma\right)$ and $f\left(\cdot;n,1/\gamma\right)$ have the same magnitude of skewness but opposite signs. The other alternatives include Jones and Faddy (2003) and Azzalini and Capitanio (2003). Aas and Haff in 2006 compared all these versions and claimed "all the three versions of the skew t-distribution presented in this section have two tails behaving as polynomials. This means that they should fit heavy-tailed data well, but they may not handle substantial skewness". They proposed the generalized hyperbolic (GH) skew student's t-distribution. It is a limiting case of the GH distribution. It has very complicated density form (interested readers can check Aas and Haff , 2006). However, it can also be represented by as a normal variance-mean mixture with the Generalized Inverse Gaussian (GIG) distribution, i.e.

$$X = \mu + \gamma W + Z\sqrt{W} \tag{5.2}$$

where μ is the location parameter, Z is a multivariate normal random vector, W is the GIG distribution and γ controls the asymmetry of the result. This is the skewed-t distribution used in *FinAnalytica* system.

The stable distribution does not have an explicit density form except for normal, Cauchy or a strictly positive stable random variable (see Rachev and Mittnik, 2000). Instead it is usually specified by its characteristic function

$$\varphi_X(t) = \begin{cases} \exp\left(-\sigma^\alpha |t|^\alpha \left[1 - i\beta \operatorname{sgn}(t) \tan\frac{\pi\alpha}{2}\right] + i\mu t\right), & \alpha \neq 1 \\ \exp\left(-\sigma|t|\left[1 - i\beta \operatorname{sgn}(t) \log|t|\right] + i\mu t\right), \alpha \neq 1 \end{cases} \tag{5.3}$$

where α is the tail fatness parameter, β is the skewness parameter and μ and σ are the location and scale parameters, respectively. The α parameter ranges from 0 to 2. When $\alpha = 2$, it becomes the normal distribution. The β parameter ranges from -1 to 1. When $\beta = 0$, the distribution is symmetric.

There are three methods for fitting stable distribution. The first is the quantile method. This is the fastest computation method. It is known to underestimate some of the parameters in some cases. The second is the Maximum Likelihood Estimation (MLE). This is the most accurate computation method, but also the most computationally intensive. The third is the regression method, a computation method that is midway between quantile and maximum likelihood estimation, both in terms of accuracy and speed.

With these characteristics, skewed t-distributions and stable distributions are versatile tools to model skewed and fat-tailed returns for a single asset. However one also needs a good model for the cross-section dependence between asset returns in a portfolio. The correlations between assets are usually nonlinear in the sense that they are more correlated when the market is falling than they are when market is rising. One solution that is becoming increasingly popular is the use of a copula. Copula is generally a multivariate distribution function defined on the cumulative probabilities on each dimension, i.e.

$$C\left(u_1, u_2, ..., u_n\right) \tag{5.4}$$

where $u_i = F_i\left(t\right) \in \left[0,1\right]$ for $i = 1..n$ .By Sklar's theorem (1996), for any multivariate cumulative distribution function $F\left(x_1, x_2, ..., x_n\right)$, there exists

$$C\left(F_1\left(x_1\right), F_2\left(x_1\right), ..., F_n\left(x_n\right)\right) = F\left(x_1, x_2, ..., x_n\right) \tag{5.5}$$

For example, if the multivariate cumulative distribution above is the multivariate normal distribution, we would have a Gaussian copula by (5.3), regardless of the marginal distributions for each dimension. However, usually Gaussian copula is a poor choice since it implies the extreme events are asymptotically independent (see Rachev, 2007). The Student t-copula, which is derived from multivariate Student's t-distribution is another choice which has the property that extreme events are asymptotically correlated. But it is symmetric on both ends. For financial returns, usually the assets tend to be more correlated in a downturn than they are in the upswing. The multivariate skewed t-distribution thus becomes a natural choice (see details in Rachev, 2007).

With both the marginal distribution and the dependence structure specified, we can glue all the univariate stable distribution together with the help of Copula structure. The Cognity® system developed from *FinAnalytica* uses a proprietary technique that fits the data to this model. After the parameters are all estimated, Monte Carlo examples can be generated to provide a much larger sample size (10,000 in our experiment) than the original historical data. The algorithm follows the following steps (See details in Embrechts, Lindskog and McNeil, 2001):

1. Generate random samples X following multivariate skewed t-distribution. This is to construct the copula as in (5.5)
2. Let $u_i = t_{\nu,\gamma}(x_i)$, where $t_{\nu,\gamma}(\cdot)$ is the skewed t-distribution for each dimension.
3. Map each dimension back to the marginal distribution fitted, i.e. $y_i = F_i^{-1}(u_i)$, where $F_i(\cdot)$ is the cumulative distribution function for the ith dimension, either stable or skewed-t distribution.

The generated Monte Carlo data then are used for ETL optimization using the linear programming paradigm specified in Chapter 4.

The whole process is parallel to the non-parametric historical ETL (HETL) optimization problem, except that we have incorporated a parametric model to generate a larger sample. Historical ETL has the advantage that it involves no distribution parameter estimation, and hence no estimation error due to this. It has the disadvantage that the historical distribution assigns zero probability to a return smaller (larger) than the smallest (largest) return. One hopes that by fitting a fat-tailed skewed distribution of the proper type the likely values below the smallest and above the largest return will be well represented by non-zero probabilities. However, a disadvantage is that there will be

estimation errors. Conditioned on the observed set of returns, this will represent pure bias in the limit as the simulation sample size goes to infinity. Unconditionally the estimation error will be in the form of variance plus bias. If the initial sample size is small, the tail fatness will be underestimated and results in a large bias than can't be overcome by however larger random samples generated in later Monte Carlo simulation. The only way to get estimates of variance and bias would be through bootstrap re-sampling of the entire process, which is not supported by Cognity® and not practical to implement.

5.2 Portfolio Optimization Results

The basic experiment settings will be the same as in Chapter 3 and 4. We optimize 100 portfolios, each with 20 randomly selected hedge funds, from 2005 to 2010. The portfolios are rebalanced monthly based on the preceding 60-month returns. All portfolios are optimized to be the global minimum risk portfolios with the long only and full investment constraints. The ETL portfolios are all set to 95% confidence level.

In this experiment we will focus on the following methods Minimum volatility (MinVol), HETL (HETL), Parametric ETL with Skewed t-distribution (SkewT), Parametric ETL with Stable distribution (Stable). The names in the parenthesis are the code names we use in the plots and tables.

In table 5.1 and 5.2, we display the summary statistics for respective out-of-sample performance and risk measures for each method across all 100 portfolios we formed. In Fig. 5.1 and 5.2, we display the kernel densities of these statistics. In Fig. 5.3 and 5.4, we display the box-plots of these statistics by comparing the difference on the same portfolio by different methods. For better presentation, some of the outliers are not shown (outside the plotting region).

Table 5.1 Summary Statistics of Performance Measures of Long Only Experiments

Annualized Mean Return					Diversification			
	MinVol	HETL	SkewT	Stable	MinVol	HETL	SkewT	Stable
Minimum	-0.137	-0.061	-0.064	-0.076	0.409	0.615	0.671	0.547
25%	-0.008	0.053	0.042	0.034	0.760	0.748	0.790	0.756
Median	0.033	0.107	0.082	0.068	0.814	0.782	0.816	0.794
Mean	0.031	0.108	0.094	0.083	0.789	0.772	0.808	0.781
75%	0.066	0.155	0.135	0.133	0.848	0.807	0.833	0.823
Maximum	0.225	0.415	0.408	0.311	0.893	0.856	0.876	0.859
Sharpe Ratio					Starr Ratio			
	MinVol	HETL	SkewT	Stable	MinVol	HETL	SkewT	Stable
Minimum	-1.220	-0.594	-0.708	-1.080	-0.142	-0.067	-0.077	-0.137
25%	-0.408	0.211	0.121	0.060	-0.041	0.026	0.014	0.007
Median	0.044	0.558	0.490	0.415	0.007	0.079	0.059	0.041
Mean	-0.005	0.647	0.500	0.451	0.008	0.084	0.066	0.061
75%	0.340	1.010	0.850	0.831	0.048	0.135	0.104	0.105
Maximum	1.150	2.420	1.880	1.800	0.284	0.337	0.307	0.257

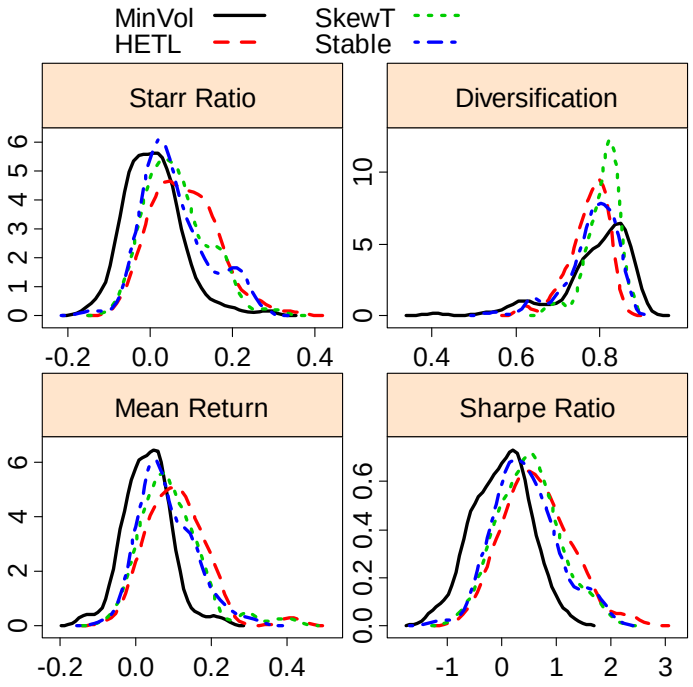

Figure 5.1 Kernel Densities of Performance Measures of Long Only Experiments

Table 5.2 Summary Statistics of Risk Measures of Long Only Experiments

Annualized Volatility					Semi. Std. Dev.			
	MinVol	HETL	SkewT	Stable	MinVol	HETL	SkewT	Stable
Minimum	0.057	0.066	0.070	0.065	0.064	0.072	0.080	0.069
25%	0.088	0.098	0.098	0.092	0.098	0.104	0.115	0.104
Median	0.104	0.113	0.113	0.112	0.113	0.126	0.134	0.127
Mean	0.118	0.133	0.133	0.123	0.118	0.127	0.135	0.130
75%	0.128	0.137	0.130	0.129	0.135	0.144	0.152	0.147
Maximum	0.359	0.723	0.632	0.422	0.206	0.228	0.314	0.328
Value at Risk					Expected Tail Loss			
	MinVol	HETL	SkewT	Stable	MinVol	HETL	SkewT	Stable
Minimum	0.020	0.011	0.024	0.015	0.038	0.035	0.037	0.030
25%	0.040	0.034	0.040	0.036	0.059	0.058	0.063	0.062
Median	0.049	0.046	0.052	0.047	0.069	0.073	0.077	0.076
Mean	0.050	0.045	0.052	0.048	0.073	0.076	0.080	0.078
75%	0.058	0.055	0.060	0.059	0.088	0.089	0.091	0.093
Maximum	0.098	0.096	0.099	0.093	0.139	0.147	0.186	0.185
Maximum Drawdown					Turnover			
	MinVol	HETL	SkewT	Stable	MinVol	HETL	SkewT	Stable
Minimum	0.098	0.075	0.105	0.069	0.051	0.064	0.092	0.098
25%	0.217	0.169	0.218	0.207	0.073	0.095	0.112	0.127
Median	0.271	0.247	0.289	0.281	0.080	0.105	0.121	0.138
Mean	0.280	0.240	0.281	0.276	0.079	0.108	0.122	0.137
75%	0.336	0.307	0.343	0.340	0.086	0.120	0.130	0.146
Maximum	0.542	0.435	0.570	0.487	0.102	0.155	0.171	0.182

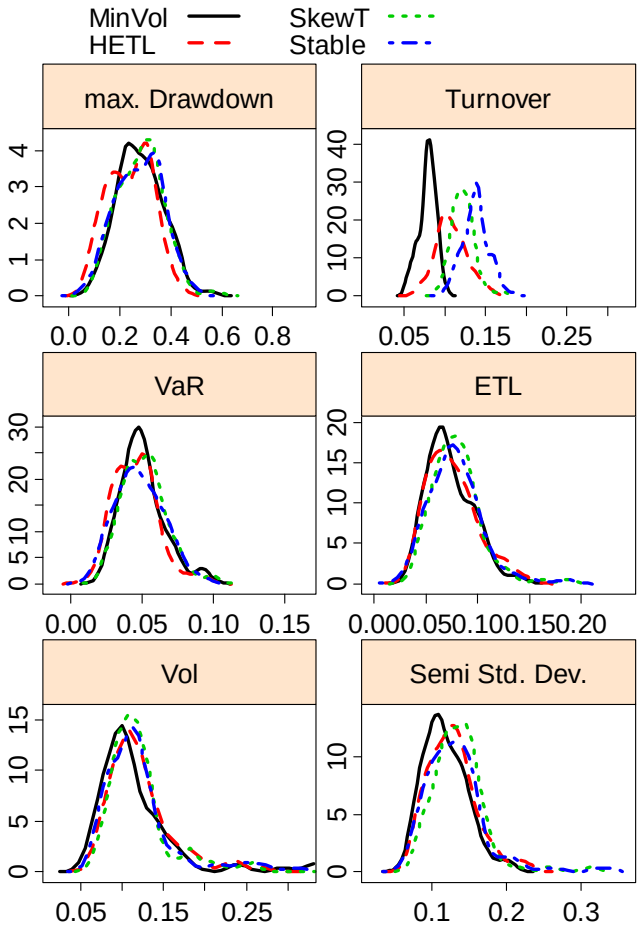

Figure 5.2 Kernel Densities of Risk Measures of Long Only Experiments

Combining results from these analyses, we observe the following facts:

1. Both HETL and parametric ETL (SkewT and Stable) outperform Minimum volatility by Starr Ratio, Sharpe Ratio and annualized mean return. HETL is the clear leader of the group. SkewT is slightly better than Stable.

2. HETL has the lowest risk profile among the three ETL methods. SkewT is slightly more risky than stable in most measures, except it has smaller turnover rate. MinVol has very low out-of-sample risks, comparable to HETL. It also has the lowest turnover rate.

The quantile cumulative return plot (Fig. 5.5) shows the cross sectional quantiles for cumulative returns across all 100 portfolios for each method. It shows that SkewT and Stable methods have larger volatility during the financial crisis for the best five portfolios. The downturn hit these two methods harder than HETL but it also recovers faster in late 2009. For the worst five portfolios, there is no such big difference between these two methods and HETL. The mean cumulative return plot (Fig. 5.6) also proves this point. We suspect the regime shifts in the market during 2008-2009 makes the parametric estimation more volatile. SkewT method has better overall mean return than Stable method in our experiments.

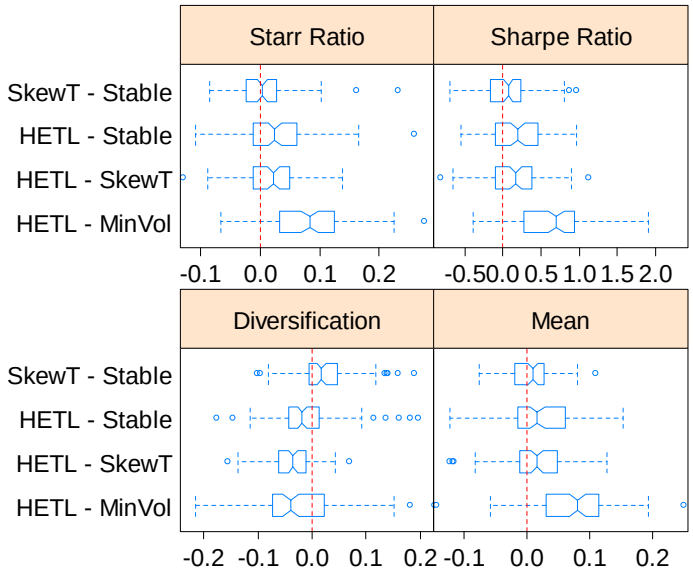

Figure 5.3 Notched Box-plots of Paired Difference for Performance Measures

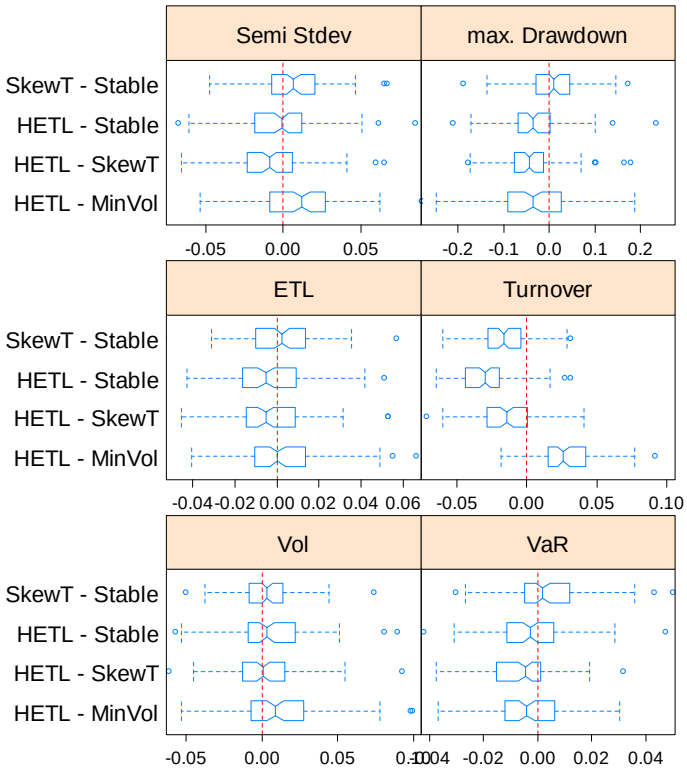

Figure 5.4 Notched Box-plots of Paired Difference for Risk Measures

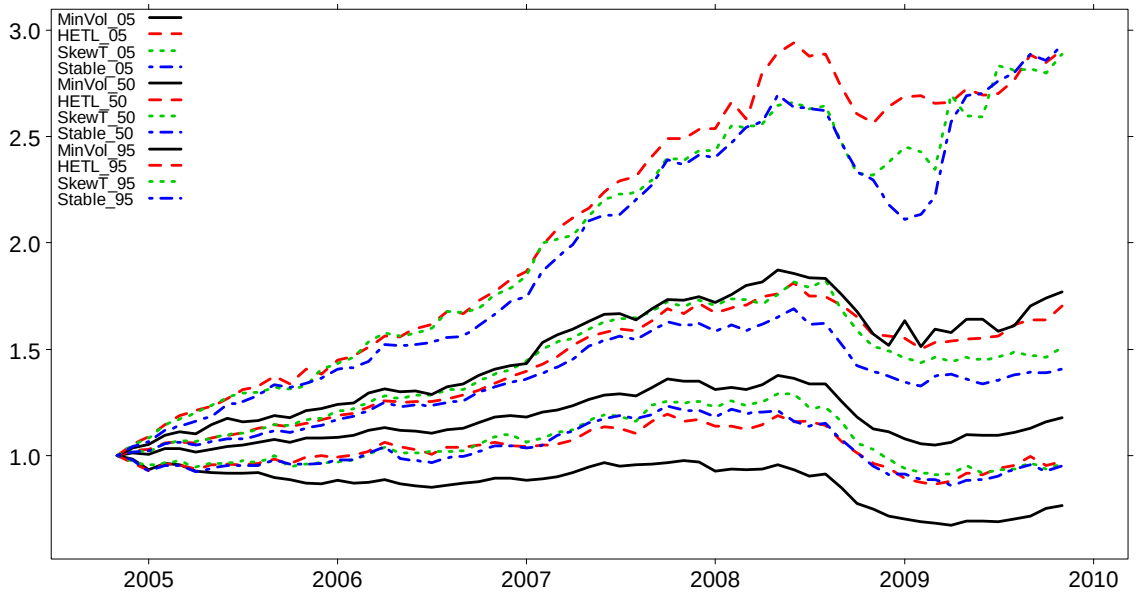

Figure 5.5 Quantile Cumulative Return Plot. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. ETL are all set at 95% confidence level.

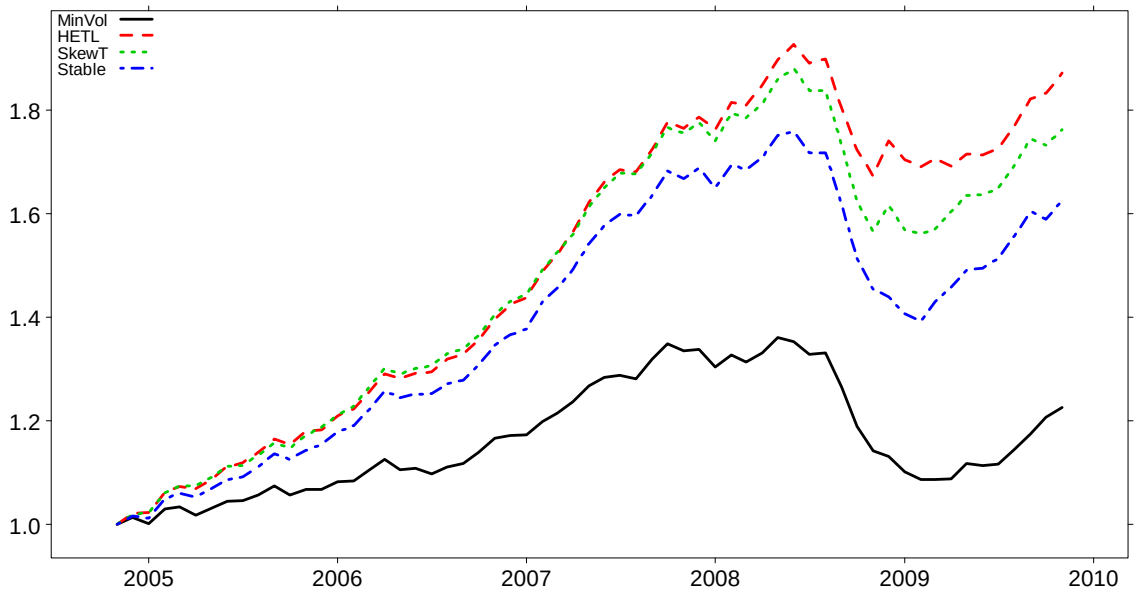

Figure 5.6 Mean Cumulative Return Plot. Each curve represents the mean of the cumulative returns across all 100 portfolios at the different time. ETL are all set at 95% confidence level.

5.3 Conclusions

In this chapter we have compared the parametric ETL risk measure with historical ETL in portfolio optimization. They generally perform the same, except that the parametric ETL methods use a much more complicated parametric model to fit the data and generate larger sample size in hope for more accuracy in the estimator. However, the large variance and bias in its estimation procedure could bring more volatile results.

Overall both parametric ETL (SkewT and Stable) and HETL outperform volatility as a risk measure in our experiments. Among the ETL measures, HETL method leads the whole group. It generates better performance and less risk. Parametric ETL methods have high volatility in the financial crisis period, especially for those well performing portfolios before 2008. The parametric ETL method that uses the skewed t-distribution outperforms the same method that uses the stable distribution. However, our analysis

shows this is at the expense of slightly more risks in the out-of-sample returns. The pair wise comparison between the same portfolios does not show significant difference between these two methods. We believe although the parametric methods have the advantage of looking beyond the range of historical returns, the limited original sample sizes makes it difficult to get an accurate estimation of the parametric model. The potential bias originated from the initial sample and estimation variability compromises the advantage of such methods. This may be the cause of a mediocre performance of parametric ETL method comparing to the simple historical ETL method.

Chapter 6
Utility Based Spectral Risk Measures

6.1 Introduction to Spectral Risk Measures

Acerbi (2002) introduced the notion of a spectral risk measure (SRM). It is a weighted average of the quantiles of the distribution function $F_W(w)$ of the end-of-period wealth random variable W. The weight function is designed so that more weight is given to larger losses that represent where the risk is. A spectral risk measure M_ϕ is formally defined by

$$M_\phi = -\int_0^1 F_W^{-1}(p)\phi(p)\,dp\,, \tag{6.1}$$

where the function $\phi(p)$ is called a *risk spectrum*. Acerbi showed that if $\phi(p)$ satisfies the conditions

$$\begin{aligned} &\phi(p) \geq 0 \\ &\phi'(p) \leq 0 \\ &\int_0^1 \phi(p)\,dp = 1\,. \end{aligned} \tag{6.2}$$

then M_ϕ is a coherent risk measure as defined by Artzner et. al. (1999). These conditions guarantee the risk spectrum is in effect a non-increasing probability density function. While these conditions form a low entrance bar for a SRM they also make good sense. First, investors want to avoid heavy losses and so it is desirable to put more weight on lower end of period wealth or lower gross return (equivalent to end-of-period wealth with \$1 invested at time zero), which is insured by the condition of a non-positive derivative of ϕ. Second, the weight function is normalized to unity so that translation in

wealth does not change the risk. Last, a SRM constructed following rules in (6.2) can also be viewed as an affine combination of ETL's at all quantile levels. Therefore it is easily a convex measure.

As simple as the SRM concept is, it is not clear how to choose a risk spectrum or the weight. The properties in (6.2) only give the basic properties of a weight function, but are not specific enough in reflecting an investor's risk preference. Utility functions on the other hand are the most commonly used tools to describe an investor's risk preference. Therefore having a method of translating a utility function representing a specific set of risk preferences to a SRM would be highly attractive. Acerbi (2004) proposed the exponential SRM based on the exponential utility function. Kevin Dowd (2008) proposed the power SRM based on power utility function. Although these spectral risk measures maintain a form similar to the utility function they are based on and inherit the respective parameters from that utility function, the translation method itself is quite ad hoc. There is currently no general recipe to directly translate an arbitrary utility function to an SRM. In this chapter, we will construct a method for such translation and discuss the properties of the resulting utility function based spectral risk measures.

We arrange the rest of this chapter in the following order. In 6.2, we consider a fairly obvious first method of mapping utility functions to spectral risk measures, provide a few examples, and show that the method is generally inadequate. In 6.3, we propose a method of mapping a utility function to a an SRM that works in a fairly general way. The utility function is combined with certain distribution functions for the translation process. We will start with the location-scale family for distribution functions, then generalize to more arbitrary forms and exhibit SRM's based on some common utility and distribution functions. In 6.4, we discuss the connection between utility based SRM and the second-

order stochastic dominance (SSD) and examine SRM under a pessimistic risk view. In 6.5, we provide a Bayesian view of utility based SRM and discuss use of an S-shaped utility. In 6.6, we derive the general form of influence function for our utility based SRM's.

6.2 Direct Utility Function Based SRM

Utility functions are based on wealth end-of-period wealth W or gross return R. Spectral risk measures are functionals defined on the space of quantile functions $F_W^{-1}(p)$ associated with the distribution function $F_W(w)$ of W and these functional are "indexed" by the probability based risk spectrum ("*p*-risk spectrum") $\phi(p)$. Our first task is to convert a SRM to a form that is defined on wealth distribution $F_W(w)$, and we do this by defining a "*w*-risk spectrum) as

$$\zeta(w) = \phi\left(F_W(w)\right). \tag{6.3}$$

Then we have

$$\begin{aligned} M_\phi &= -\int_{-\infty}^{\infty} w \cdot \phi\left(F_W(w)\right) dF_W(w) \\ &= -\int_{-\infty}^{\infty} w \cdot \zeta(w)\, dF_W(w) \\ &\triangleq M_\zeta \ . \end{aligned}$$

Assuming that W has a probability density function $f_W(w)$ we have

$$\zeta'(w) = \phi'\left(F_W(w)\right) f_W(w)$$

and

$$
\begin{aligned}
E\left(\zeta\left(w\right)\right) &= \int_{-\infty}^{\infty} \zeta\left(w\right) f_W\left(w\right) dw \\
&= \int_{-\infty}^{\infty} \phi\left(F_W\left(w\right)\right) f_W\left(w\right) dw \\
&= \int_0^1 \phi\left(p\right) dp \qquad .
\end{aligned}
$$

Thus it follows from (6.2) that M_ζ defines a coherent risk measure if

$$
\begin{aligned}
&\zeta\left(w\right) \geq 0 \\
&\zeta'\left(w\right) \leq 0 \\
&E\left(\zeta\left(w\right)\right) = 1 \, .
\end{aligned}
\tag{6.4}
$$

Combining (6.3) and (6.4), we made a simple transformation to change the domain of spectral risk measures from inverse probability distribution functions F_W^{-1} of wealth to probability distribution functions F_W of wealth, with the risk spectrum now depending on F_W (instead of p) To As an example of how this works consider the ETL (CVaR) risk measure defined by the *p*-risk spectrum.

$$
\phi\left(p\right) = \frac{1}{\alpha} 1\{0 \leq p \leq \alpha\}
$$

giving

$$
ETL_W\left(\alpha\right) = -\frac{1}{\alpha} \int_0^\alpha F_W^{-1}\left(p\right) dp \, .
$$

The corresponding *w*-risk spectrum is

$$
\zeta\left(w\right) = \frac{1}{\alpha} 1\{0 \leq F_W\left(w\right) \leq \alpha\}
$$

giving the alternative more familiar representation

$$
\begin{aligned}
ETL_W(\alpha) &= -\int_{-\infty}^{\infty} w \cdot \zeta(w)\, f_W(w) dw \\
&= -\frac{1}{\alpha}\int_{-\infty}^{\infty} w \cdot 1\{0 \le F_W(w) \le \alpha\}\, f_W(w) dw \\
&= -\frac{1}{\alpha}\int_{-\infty}^{F_W^{-1}(\alpha)} w \cdot f_W(w) dw \qquad .
\end{aligned}
$$

We note that the *w*-risk spectrum $\zeta(w)$ is dependent on the wealth distribution function while the *p*-risk spectrum $\phi(p)$ is purely based on cumulative probability p and is independent of the actual distribution of w. To maintain this distinction (6.3) indicates that $\zeta(w)$ cannot be an arbitrary function of w and instead must a function of F_W having the composition form (6.3). Can we remove the dependency on F_W in $\zeta(w)$? The short answer is no, as we shall see.

Having an SRM that is a function of wealth w, we are almost ready to use utility functions to define SRM's. Utility functions reflect an investor's satisfaction with various levels of wealth and the first derivative of a utility function (marginal utility) shows how fast the investor's satisfaction changes with changing wealth. As such marginal utility indicates how risky the current wealth level is. If the investor can become very unhappy with a small change in his or her wealth level, it reveals that the current level is very risky. On the other hand if a small change in wealth does not much change the investor's satisfaction then the position can be viewed as safe and comfortable, i.e., not very risky. With this consideration, focus on the derivative of utility functions $U(w)$ strictly non-satiated and risk-averse investors, i.e., utility functions that strictly increasing and concave. Assuming the first two derivatives exist, such utility functions satisfy

$$\begin{aligned} U'(w) &> 0 \\ U''(w) &< 0. \end{aligned} \tag{6.5}$$

If we define the w-risk spectrum as

$$\zeta(w) = U'(w) \tag{6.6}$$

then the first two conditions of (6.4) are satisfied:

$$\begin{aligned} \zeta(w) &> 0 \\ \zeta'(w) &< 0\ . \end{aligned} \tag{6.7}$$

Further if $E\left(U'(w)\right)$ exists then (6.7) implies that $E\left(U'(w)\right) > 0$ and setting

$$\zeta(w) = \frac{U'(w)}{E\left(U'(w)\right)} \tag{6.8}$$

results in satisfying the third condition of (6.5) as well:

$$E\left(\zeta(w)\right) = 1.$$

This results in the following SRM:

$$\begin{aligned} M_\zeta &= -\int_{-\infty}^{\infty} w\zeta(w)\, f_W(w)\, dx \\ &= -\int_{-\infty}^{\infty} x \frac{U'(w)}{E\left(U'(w)\right)}\, f_W(w)\, dx \\ &= \frac{E\left(-w \cdot U'(w)\right)}{E\left(U'(w)\right)} \quad . \end{aligned} \tag{6.9}$$

In summary, the standard properties of a utility function in (6.6) result in a w risk spectrum function given by (6.8) satisfies all conditions in (6.4). As a result, we created a mapping from utility functions to SRM measures (6.9).

We now consider trying to derive a SRM using (6.8) for the following two utility functions: a risk neutral utility function and an exponential utility function.

From Risk Neutral Utility to SRM

Under risk neutrality $U'(w) \equiv c$ is a constant, which results in

$$\zeta(w) = \frac{U'(w)}{E(U'(w))} \equiv 1$$

and so

$$M_\zeta = \frac{E(-w \cdot U'(w))}{E(U'(w))} = -E_{F_W}(w). \tag{6.10}$$

At first glance it may seem strange that the negative mean wealth becomes a risk measure. However, in a risk neutral world all investors choose a utility function with $U'(w) \equiv c$ for some c that could be different for different investors. $E_{F_W}(w)$ is nevertheless the same for all investors no matter what F_W is, for otherwise there would be an arbitrage opportunity. Hence the risk measure M_ζ should be the same for all investors in a risk neutral world (not the real world). This is hardly surprising since risk neutral investors do not differentiate different risks by definition.

Negative Exponential Utility

Now consider a negative exponential utility function defined on gross returns r :

$$U(r) = -e^{-\lambda r} \qquad \lambda > 0, \tag{6.11}$$

From (6.8), we have w-risk spectrum and corresponding SRM

$$\zeta(r) = \frac{U'(r)}{E\left(U'(r)\right)} = \frac{e^{-\lambda r}}{E\left(e^{-\lambda r}\right)} \tag{6.12a}$$

$$M_{\zeta} = -E\left(r \cdot \zeta(r)\right) = -\frac{E\left(re^{-\lambda r}\right)}{E\left(e^{-\lambda r}\right)}. \tag{6.12b}$$

Now suppose we have normal distribution for gross returns r

$$f_R(r) = \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left\{-\frac{(r-\mu_R)^2}{2\sigma_R^2}\right\},$$

where the combination of initial wealth and normal distribution mean and variance are such that the probability of a negative wealth is negligible (this is a common implicit assumption even in advanced textbooks, e.g., Pennacchi, 2008, pp. 48-9). Then

$$\begin{aligned} E\left(e^{-\lambda r}\right) &= \int_{-\infty}^{\infty} e^{-\lambda r} \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left\{-\frac{(r-\mu_R)^2}{2\sigma_R^2}\right\} dr \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left\{-\frac{\left(r-\mu_R+\lambda\sigma_R^2\right)^2}{2\sigma_R^2}\right\} \cdot \exp\left\{\frac{1}{2}\lambda^2\sigma_R^2 - \lambda\mu_R\right\} dr \\ &= \exp\left\{\frac{1}{2}\lambda^2\sigma_R^2 - \lambda\mu_R\right\} \end{aligned} \tag{6.13a}$$

and

$$E\left(r\cdot U'(r)\right)=E\left(re^{-\lambda r}\right)=\int_{-\infty}^{\infty}re^{-\lambda r}\frac{1}{\sqrt{2\pi}\sigma_R}\exp\left\{-\frac{\left(r-\mu_R\right)^2}{2\sigma_R^2}\right\}dr$$

$$=\int_{-\infty}^{\infty}r\frac{1}{\sqrt{2\pi}\sigma_R}\exp\left\{-\frac{\left(r-\mu_R+\lambda\sigma_R^2\right)^2}{2\sigma_R^2}\right\}\cdot\exp\left\{\frac{1}{2}\lambda^2\sigma_R^2-\lambda\mu_R\right\}dr \qquad (6.13b)$$

$$=\exp\left\{\frac{1}{2}\lambda^2\sigma_R^2-\lambda\mu_R\right\}\cdot\left(\mu_R-\lambda\sigma_R^2\right) \quad .$$

Consequently,

$$M_\zeta=-\frac{E\left(r\cdot U'(r)\right)}{E\left(U'(r)\right)}=-\left(\mu_R-\lambda\sigma_R^2\right) \qquad (6.13c)$$

It is easy to find that since σ_R^2 is not positive homogeneous the same is true of M_ζ and so is not a coherent risk measure. Thus our mapping (6.8) failed to generate a proper SRM. The reason is that although the resulting risk measure depends on F_r, it does not depend on r strictly through composition with F_r .

These two examples show the mapping from utility functions to risk spectrum is not always successful. On the opposite, given a risk measures and spectrum, can we find a utility function that gives rise to it?

V*aR*

We have

$$\zeta(w)=\delta_{F_W^{-1}(\alpha)}(w) \qquad (6.14)$$

where δ is the delta Dirac function (Fig. 6.1). It certainly is not a proper function and any sequence of approximating functions fail to everywhere deceasing. Hence VaR is not a

spectral risk measure and cannot be claimed to be coherent by virtue of being a spectral risk measure. There is no risk-averse utility function that would agree with (6.14) in the sense of (6.8), because it depends exogenously on the wealth distribution.

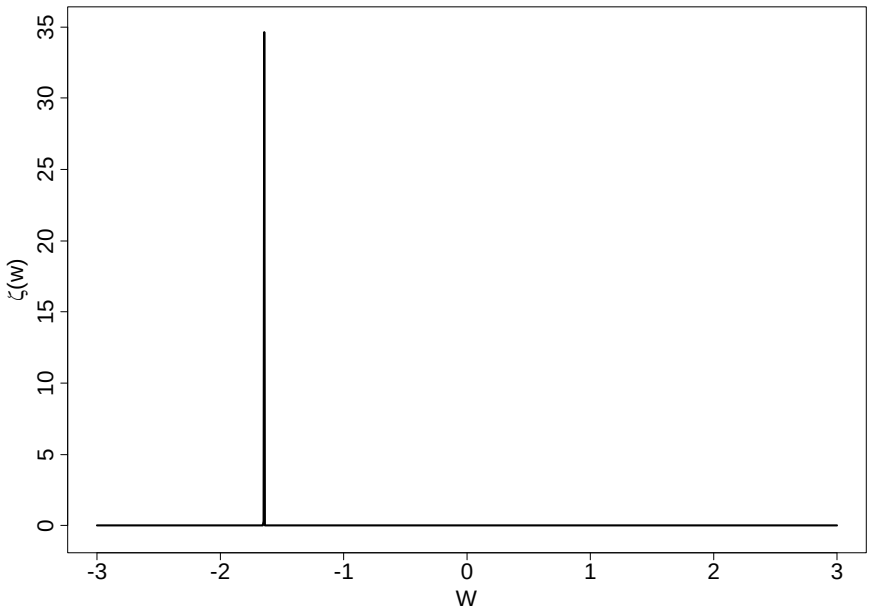

Figure 6.1 Approximation of $\zeta(w)$ (6.14) for $\mathrm{VaR}_{.95}$ under standard normal distribution

ETL

As in 6.1, we have

$$\zeta(w) = \frac{1}{\alpha} \cdot 1\left\{-\infty \le w \le F_W^{-1}(\alpha)\right\}, \tag{6.15}$$

which implies

$$\frac{U'(w)}{E(U'(w))} = \frac{1}{\alpha} \cdot 1\left\{-\infty \le w \le F_W^{-1}(\alpha)\right\}.$$

This is equivalent to

$$U(w) = \left(w - F_W^{-1}(\alpha)\right) \cdot 1\left\{-\infty \le w \le F_W^{-1}(\alpha)\right\}. \tag{6.16}$$

An investor who uses this utility function (Fig. 6.2) is not strictly non-satiated and risk averse. Instead he is non-satiated and risk neutral with respect to all wealth values below $F_W^{-1}(\alpha)$ and satiated and risk neutral with respect to all wealth values larger than $F_W^{-1}(\alpha)$. In a simple word, this investor is not strictly risk averse. However, weight function defined in (6.11) still satisfies all conditions in (6.4). ETL is a spectral risk measure.

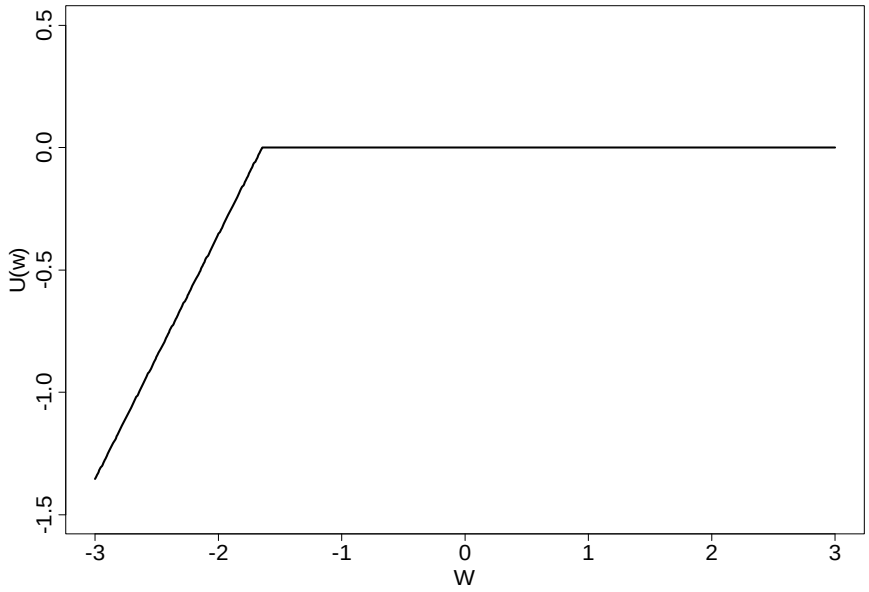

Figure 6.2 Utility Function for $\mathrm{ETL}_{0.95}$ (6.16) under standard normal distribution

6.3 Proper Utility Based SRM's

A possible fix to definition (6.8) is to define the w-risk spectrum as

$$\zeta(w) = \frac{U'\left(F_W(w)\right)}{E\left(U'\left(F_W(w)\right)\right)}, \tag{6.17}$$

so that $\zeta(w)$ has the required composition dependence on the wealth distribution F_W. However, this method has three problems.

P1) Weight function is distribution dependent and so we can't decide the weight function without knowing the distribution of F_W.

P2) Utility functions are generally defined on $[0,\infty)$ not on $[0,1]$. Hence the construction in (6.17) will only utilize a part of the utility function and which is quite unsatisfactory. (This is not addressed in Dowd, 2008).

P3) The derivative of utility function applied to F_W may not be integrable. For example, for the log utility function $U(w)=\log(w)$ we have

$$E\left[U'\left(F(w)\right)\right]=\int_0^1 \frac{1}{p}dp=\infty$$

and this prevents normalizing as in (6.8).

In spite of the above problems, it is highly appealing to use $U'(\bullet)$ as a link between the risk spectrum function and the utility function. In what follows we propose methods that retain a form closely related to (6.17) and are free of problems stated above. We achieve this by first considering some special distribution families and then generalize our results.

We proceed by replacing $F_W(w)$ in (6.17) with a more general $h(w)$ having the following properties:

1. $h(w)$ maps R^+ onto $(0,1)$
2. $E\left[U'\left(h(w)\right)\right] < \infty$
3. $h(w)$ is isotopic with respect to $F_W(w)$ in the following sense:

$$F_1(w) = F_2(w) \Leftrightarrow h_1(w) = h_2(w) \tag{6.18}$$

for any two cumulative distribution F_1 and F_2.

Conditions 1 and 2 are easy to satisfy by choosing for $h(w)$ a cumulative distribution function on the positive real line with a right tail that is not too fat. For condition 3, an interesting possibility is to choose

$$F_{W,h}(w) = F_0\left(h(w)\right) \tag{6.19}$$

where $F_0(w)$ is a suitable standard distribution with fixed distribution parameters. For two different choices of $h_i(w)$, $i = 1,2$, we write

$$F_i(w) = F_{W,i}(w) = F_0\left(h_i(w)\right),\ i = 1,2$$

The choice (6.19) also satisfies condition 3 because

$$\begin{aligned} &F_1(w) = F_2(w) \Rightarrow \\ &h_1(w) = F_0^{-1}\left(F_0\left(h_1(w)\right)\right) = F_0^{-1}\left(F_1(w)\right) \\ &= F_0^{-1}\left(F_2(w)\right) = h_2(w) \end{aligned}$$

and

$$\begin{aligned} &h_1(w) = h_2(w) \Rightarrow \\ &F_1(w) = F_0\left(h_1(w)\right) = F_0\left(h_2(w)\right) = F_2(w). \end{aligned}$$

Using (6.19) we have

$$U'\left(h\left(w\right)\right)=U'\left(F_0^{-1}\left(F_0\left(h\left(w\right)\right)\right)\right)=\left(U'\circ F_0^{-1}\right)\left(F_{W,h}\left(w\right)\right).$$

and so now define our w-risk spectrum as

$$\zeta\left(w\right)=\frac{U'\left(h\left(w\right)\right)}{E\left(U'\left(h\left(w\right)\right)\right)}=\frac{\left(U'\circ F_0^{-1}\right)\left(F_{W,h}\left(w\right)\right)}{E\left(\left(U'\circ F_0^{-1}\right)\left(F_{W,h}\left(w\right)\right)\right)}. \tag{6.20}$$

Now we only need to check that

$$\begin{aligned}\left(U'\circ F_0^{-1}\right)&\geq 0\\ \left(U'\circ F_0^{-1}\right)'&\leq 0.\end{aligned} \tag{6.20a}$$

The first condition is obvious since U' is non-negative and the second condition is satisfied because

$$\left(U'\circ F_0^{-1}\right)'=U''\frac{1}{f_0\left(F_0^{-1}\right)}\leq 0.$$

To summarize: If for all distribution functions $F_W\left(w\right)$ for which the risk measure is defined and finite we have a function $h\left(w\right)$ that satisfies (6.19), then the SRM

$$M_\zeta=-E\left(w\cdot\zeta\left(w\right)\right)=-\frac{E\left(w\cdot U'\left(h\left(w\right)\right)\right)}{E\left(U'\left(h\left(w\right)\right)\right)} \tag{6.21}$$

using w-risk spectrum (6.20) is a spectral risk measure.

Location-Scale Family

A location-scale family of wealth distributions has the form

$$F_W(w) = F_0\left(\frac{w-\mu}{\sigma}\right) \tag{6.22a}$$

where μ and σ are location and scale parameters and F_0 is fixed. If we set

$h(w) = \frac{w-\mu}{\sigma}$ we have

$$\zeta(w) = \frac{U'(h(w))}{E(U'(h(w)))} = \frac{U'\left(\frac{w-\mu}{\sigma}\right)}{E\left(U'\left(\frac{w-\mu}{\sigma}\right)\right)}. \tag{6.22b}$$

We note that

$$\begin{aligned} E\left(U'\left(\frac{w-\mu}{\sigma}\right)\right) &= \int_{-\infty}^{\infty} U'\left(\frac{w-\mu}{\sigma}\right)\frac{1}{\sigma} f_0\left(\frac{w-\mu}{\sigma}\right) dw \\ &= \int_{-\infty}^{\infty} U'(z)\, f_0(z)\, dz \end{aligned}$$

where $z = \frac{w-\mu}{\sigma}$. Thus the SRM is

$$\begin{aligned} M_\zeta &= -\frac{E\left(w \cdot U'\left(\frac{w-\mu}{\sigma}\right)\right)}{E\left(U'\left(\frac{w-\mu}{\sigma}\right)\right)} = -\frac{\int_{-\infty}^{\infty}(\mu+\sigma z)U'(z)\, f_0(z)\, dz}{\int_{-\infty}^{\infty} U'(z)\, f_0(z)\, dz} \\ &= -\left(\mu + \sigma\frac{E(zU'(z))}{E(U'(z))}\right). \end{aligned} \tag{6.23}$$

For the case of an exponential utility and normal distribution for w as in (6.14) we have

$$\frac{E\left(zU'(z)\right)}{E\left(U'(z)\right)} = \frac{\exp\left\{\frac{1}{2}\lambda^2\right\}\cdot(-\lambda)}{\exp\left\{\frac{1}{2}\lambda^2\right\}} = -\lambda.$$

This give as coherent risk measure the "mean-versus-risk" form

$$M_\zeta = -(\mu - \lambda\sigma). \tag{6.24}$$

with risk aversion parameter λ .

Since our method of constructing a spectral risk measure based on a utility function can be transformed into the standard form (6.1) with a spectral risk measure $\phi(p) = \zeta(w)$ as defined in (6.22), it follows that the risk measure (6.24) is coherent. We can formally state this result in the following theorem:

Theorem 6.1 For a random variable that follows a location-scale family distribution F_W as defined in (6.22a), its spectral risk measure M_ζ as in (6.23) that uses the w-risk spectrum as in (6.22b) is coherent.

Proof: (6.22b) can be easily rewritten as $\zeta(w) = \frac{U' \circ F_0^{-1}(p)}{E\left(U' \circ F_0^{-1}(p)\right)}$. Therefore (6.23) becomes

$$\begin{aligned} M_\zeta &= -E\left(w\cdot\zeta(w)\right) = -\int_0^1 F_W^{-1}(p)\cdot\frac{U' \circ F_0^{-1}(p)}{E\left(U' \circ F_0^{-1}(p)\right)}dp \\ &\triangleq -\int_0^1 F_W^{-1}(p)\cdot\phi(p)\,dp. \end{aligned} \tag{6.25}$$

As verified in (6.20a), we know $\phi(p)$ in (6.25) satisfies (6.2). Therefore M_ζ is coherent.

It is interesting that in the case of negative exponential utility and normal returns distributions, obtaining a coherent spectral risk measure derived from a utility function in the above reasonable way leads to mean-versus-volatility optimization to minimize risk, whereas the direct evaluation of expected utility in this case leads to mean-variance optimality.

Generalization to an Arbitrary Distribution

We now proceed to more generalized distributions. Suppose we define

$$\zeta(w) = \frac{U'\left(F_0^{-1}\left(F_W(w)\right)\right)}{E\left(U'\left(F_0^{-1}\left(F_W(w)\right)\right)\right)} \tag{6.26}$$

which is equivalent to

$$\phi(p) = \frac{U'\left(F_0^{-1}(p)\right)}{E\left(U'\left(F_0^{-1}(p)\right)\right)} \tag{6.27}$$

where F_0 is a cumulative distribution function we choose judiciously for the utility function we are working with. It does not need to have any association with the true probability distribution $F_W(w)$. Its only purpose is to map the cumulative probability p back to the space where the utility function is defined.

For example, we may choose $F_0 = \Phi$, the standard normal distribution if the utility function is defined on R and choose the standard lognormal distribution for F_0 if the utility function is defined on R^+ regardless of the actual underlying distribution $F_W(w)$.

In the following we will take $F_0(q)$ to be the distribution function of a random variable Q that takes on realized values q, and use this distribution function to map the original wealth variable w to q:

$$F_0(q) = F_W(w)$$

$$w = w(q) = F_W^{-1}\left(F_0(q)\right)$$

and

$$q = q(w) = F_0^{-1}\left(F_W(w)\right). \tag{6.28}$$

A similar expression holds with wealth replaced by returns with the usual assumption about negligible probability of negative returns.

Then we have,

$$U'\left(F_0^{-1}\left(F_W(w)\right)\right) = U'(q)$$

and

$$\begin{aligned} E\left(U'\left(F_0^{-1}\left(F_W(w)\right)\right)\right) &= \int_0^1 U'\left(F_0^{-1}(p)\right)dp \\ &= \int_{-\infty}^{\infty} U'(q)\, f_0(q)\, dq \\ &= E_Q\left(U'(q)\right) \end{aligned}$$

So we define the q-risk spectrum

$$\zeta(q) = \frac{U'(q)}{E_Q\left(U'(q)\right)}, \tag{6.29}$$

and can represent the SRM as

$$\begin{aligned} M_\zeta &= -\frac{E_Q\left(w(q)\cdot U'(q)\right)}{E_Q\left(U'(q)\right)} \\ &= -\frac{E_Q\left(F_W^{-1}\left(F_0(q)\right)\cdot U'(q)\right)}{E_Q\left(U'(q)\right)}. \end{aligned} \tag{6.30}$$

We emphasize that this new q-risk spectrum depends only on the fixed distribution F_0 and the utility function U. It does not depend on the distribution F_W or F_R of wealth or returns, and the only constraints on these functions is that F_0 has the same domain $U(\cdot)$ and $E_Q\left(U'(q)\right)$ is integrable. Thus we have avoided the three problems stated in 6.3.1 and have a class of utility function based spectral risk measures.

Now we give some examples of spectral risk measures in this general type.

Exponential Utility Function and Normal Distribution

With the negative exponential utility defined in (6.14) and F_0 set to the normal distribution, we have

$$M_\zeta = -\frac{E_\Phi\left(r_q \cdot e^{-\lambda q}\right)}{E_\Phi\left(e^{-\lambda q}\right)} = -e^{-\frac{1}{2}\lambda^2} \cdot E_\Phi\left(r_q \cdot e^{-\lambda q}\right), \tag{6.31}$$

where $\Phi(q) = F_R\left(r_q\right)$. This will further reduce to (6.24) if $F_R(r)$ is a normal distribution.

Power and Log Utility with Lognormal Distribution

Power utility function is defined as

$$U(w)=\frac{w^{\lambda}-1}{\lambda}, \tag{6.32}$$

for $w>0$, $\lambda<1$ and $\lambda\neq 0$. In the special case when $\lambda=0$, the power utility function reduces to the log utility function

$$U(w)=\log w. \tag{6.33}$$

In both cases, the first derivative has a uniform expression $U'(w)=w^{\lambda-1}$.

We choose F_0 to be the standard lognormal distribution where

$$\begin{aligned}\mu &= E\left(\log q\right)=0\\ \sigma^2 &= Var\left(\log q\right)=1.\end{aligned}$$

We also set $z=\log(q)$, hence z follows the standard normal distribution. We can derive SRM as follows

$$\begin{aligned}M_{\zeta} &= -\frac{E_Q\left(w\cdot q^{\lambda-1}\right)}{E_Q\left(q^{\lambda-1}\right)}=-\frac{E_{\Phi}\left(w_z\cdot e^{(\lambda-1)z}\right)}{E_{\Phi}\left(e^{(\lambda-1)z}\right)}\\ &= -e^{-\frac{1}{2}(\lambda-1)^2}\cdot E_{\Phi}\left(w_z\cdot e^{(\lambda-1)z}\right),\end{aligned} \tag{6.34}$$

where $\Phi(z)=F_0(q)=F_W(w_z)$. Generally this is does not have a closed form. Only when $E_{\Phi}\left(w_z\cdot e^{(\lambda-1)z}\right)$ can be simplified, we may be able to give a closed form of the risk. The normal distribution family given in the last section is a good example. If w also has

a lognormal distribution with parameters μ and σ , s.t. $\mu = E(\log w)$ and $\sigma^2 = Var(\log w)$. We have

$$w_z = F_W^{-1}(\Phi(z)) = e^{\mu+\sigma z}. \tag{6.35}$$

Now (6.34) further reduces to

$$\begin{aligned} M_\zeta &= -e^{-\frac{1}{2}(\lambda-1)^2} \cdot E_\Phi\left(e^{\mu+\sigma z} \cdot e^{(\lambda-1)z}\right) \\ &= -e^{\mu+\frac{1}{2}\sigma^2} \cdot e^{(\lambda-1)\sigma} \\ &= -\mu_W \cdot e^{(\lambda-1)\sigma} \end{aligned} \tag{6.36}$$

This shows larger the mean is, smaller the risk is. Also smaller the variance is, smaller is the risk. For the same variable, a more risk-averse investor who has a more negative λ would get higher risk (note $\mu_W > 0$ under lognormal).

Under log utility (6.36) reduces to

$$M_\zeta = -\mu_W \cdot e^{-\sigma} \tag{6.37}$$

At the first sight we may think (6.36) is not coherent. However, this is merely because the representation is under the original parameters. (6.36) is a direct result from (6.34) and hence is guaranteed to be coherent.

Exponential Utility with Empirical Distribution

What if we apply our utility based SRM to an arbitrary underlying distribution? Generally we won't have a closed form because $F_W^{-1}(F_0(q))$ is unknown. However, for well

behaved utility functions, we may get useful results by directly using the empirical distribution for our random variable w or r, when we calculate the risk.

For example, if we still use the exponential utility function and use the standard normal distribution as our exogenous interpreter for percentiles. We have the following results before we even consider the distribution for x

$$\begin{aligned} E_{\Phi}\left(U'(q)\right) &= \exp\left\{\frac{1}{2}\lambda^2\right\} \\ U'(q) &= e^{-\lambda q} \end{aligned} \quad . \tag{6.38}$$

Therefore if we have the empirical distribution

$$F_n\left(w_{(i)}\right) = \frac{i}{n} \tag{6.39}$$

for w, where $w_{(1)} \le w_{(2)} \le \ldots \le w_{(n)}$ is the order statistics. We can write

$$\begin{aligned} E\left(w \cdot U'(q)\right) &= \sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \exp\{-\lambda q\}\, d\Phi(q) \\ &= \sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \exp\{-\lambda q\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{q^2}{2}\right\} dq \\ &= \sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(q+\lambda)^2}{2}\right\} \exp\frac{\lambda^2}{2} dq \\ &= \sum_{i=1}^{n} \left[w_{(i)} e^{\lambda^2/2} \left\{ N_{-\lambda,1}\left(\Phi^{-1}(i/n)\right) - N_{-\lambda,1}\left(\Phi^{-1}\left((i-1)/n\right)\right)\right\}\right] \end{aligned} \tag{6.40}$$

where $N_{a,b}(\cdot)$ is the cumulative distribution function for normal distribution with $\mu = a, \sigma^2 = b$.

Consequently,

$$M_{\zeta} = -\frac{E_{\Phi}\left(w \cdot U'(q)\right)}{E_{\Phi}\left(U'(q)\right)} = -\frac{\sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \exp\{-\lambda q\} d\Phi(q)}{\exp\left\{\frac{1}{2}\lambda^2\right\}}$$

$$= -\sum_{i=1}^{n} w_{(i)} \int_{(i-1)/n}^{i/n} \exp\left\{-\lambda \Phi^{-1}(p) - \frac{1}{2}\lambda^2\right\} dp \qquad (6.41)$$

$$= \sum_{i=1}^{n} \left[\left(-w_{(i)}\right) \cdot s_i\right] \quad .$$

where s_i is the weight for i^{th} quantile.

$$s_i = \int_{(i-1)/n}^{i/n} \exp\left\{-\lambda \Phi^{-1}(p) - \frac{1}{2}\lambda^2\right\} dp$$

$$= \int_{(i-1)/n}^{i/n} \exp\left\{-\frac{1}{2}\left(\lambda + \Phi^{-1}(p)\right)^2 + \frac{1}{2}\Phi^{-1}(p)^2\right\} dp \qquad (6.42)$$

$$= \int_{(i-1)/n}^{i/n} \exp\left\{-\frac{1}{2}\left(\lambda + \Phi^{-1}(p)\right)^2 + \frac{1}{2}\Phi^{-1}(p)^2\right\} dp$$

We can view $\left(-w_{(i)}\right)$ as the losses generated by our distribution. Weights s_i can be calculated without any knowledge of the distribution of w and reflects the investor's risk preference by λ. We can make the following observations (Fig. 6.3):

1. The weight is decreasing from left tail to right tail.
2. When $\lambda = 0$, $s_i = \frac{1}{n}$. This again implies risk-neutral investors have the mean as their risk.
3. With larger λ, weights s_i decrease more rapidly.

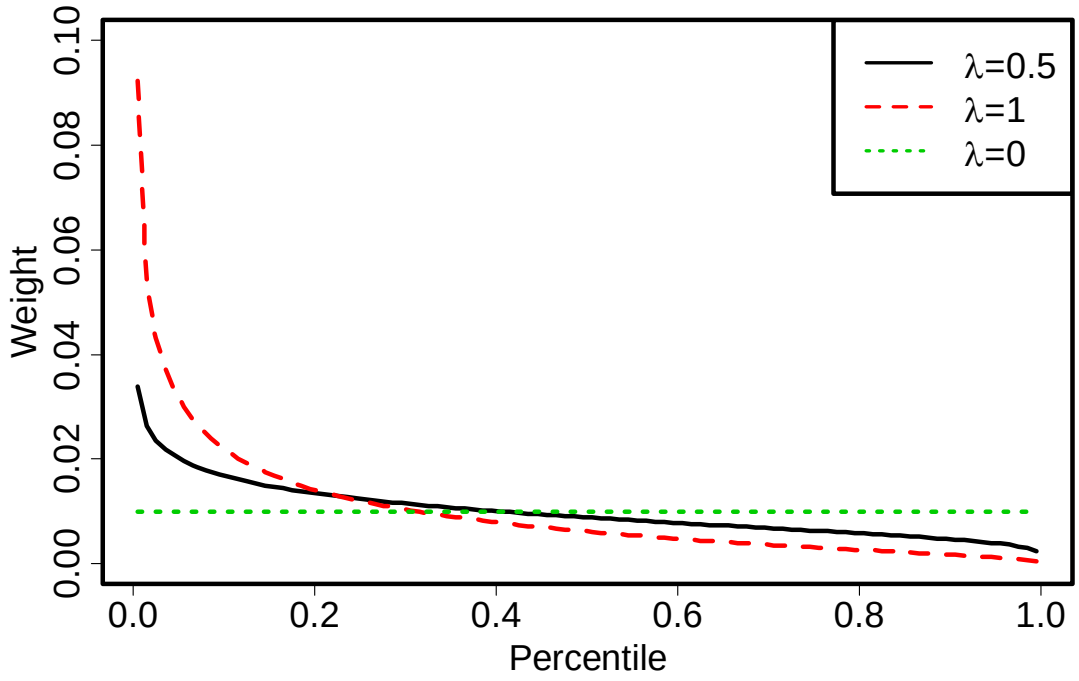

Figure 6.1 Weight vs. Percentile: for exponential utility ($\lambda = 0, 0.5, 1$) and empirical distribution (sample size=100).

Power and Log Utility with Empirical Distribution

If we have the power utility function and use the standard lognormal distribution as our exogenous interpreter for percentiles. We have the following results before we even consider the distribution for x

$$E_Q\left(U'(q)\right) = E_Q\left(q^{\lambda-1}\right) = E_z\left(e^{(\lambda-1)z}\right) = e^{\frac{1}{2}(\lambda-1)^2},$$

where $z = \log(q)$. Following the same routine,

$$
\begin{aligned}
E_Q\left(w\cdot U'(q)\right) &= \sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \exp\left\{(\lambda-1)z\right\} d\Phi(z) \\
&= \sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \exp\left\{(\lambda-1)z\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \sum_{i=1}^{n} w_{(i)} \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\left(z-(\lambda-1)\right)^2}{2}\right\} e^{\frac{1}{2}(\lambda-1)^2} dz \\
&= \sum_{i=1}^{n} \left[w_{(i)} e^{\frac{1}{2}(\lambda-1)^2} \left\{ N_{\lambda-1,1}\left(\Phi^{-1}(i/n)\right) - N_{\lambda-1,1}\left(\Phi^{-1}\left((i-1)/n\right)\right)\right\}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
M_\zeta &= -\frac{E_Q\left(w\cdot U'(q)\right)}{E_Q\left(U'(q)\right)} = \sum_{i=1}^{n}\left[\left(-w_{(i)}\right)\cdot s_i\right] \\
s_i &= N_{\lambda-1,1}\left(\Phi^{-1}(i/n)\right) - N_{\lambda-1,1}\left(\Phi^{-1}\left((i-1)/n\right)\right)
\end{aligned}
\tag{6.43}
$$

The weight s_i can be pre-calculated for samples with fixed sample size.

$$
\begin{aligned}
s_i &= \int_{\Phi^{-1}((i-1)/n)}^{\Phi^{-1}(i/n)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\left(z-(\lambda-1)\right)^2}{2}\right\} dz \\
&= \int_{(i-1)/n}^{i/n} \exp\left\{(\lambda-1)\Phi^{-1}(p) - \frac{1}{2}(\lambda-1)^2\right\} dp \\
&= \int_{(i-1)/n}^{i/n} \exp\left\{-\frac{1}{2}\left((\lambda-1)-\Phi^{-1}(p)\right)^2 + \frac{1}{2}\Phi^{-1}(p)^2\right\} dp
\end{aligned}
\tag{6.44}
$$

We can make the following observations (Fig. 6.4):

1. The weight is decreasing from left tail to right tail.
2. Since in power utility $\lambda \le 1$, the weight function above is of the same form of (6.42) in the exponential utility case.

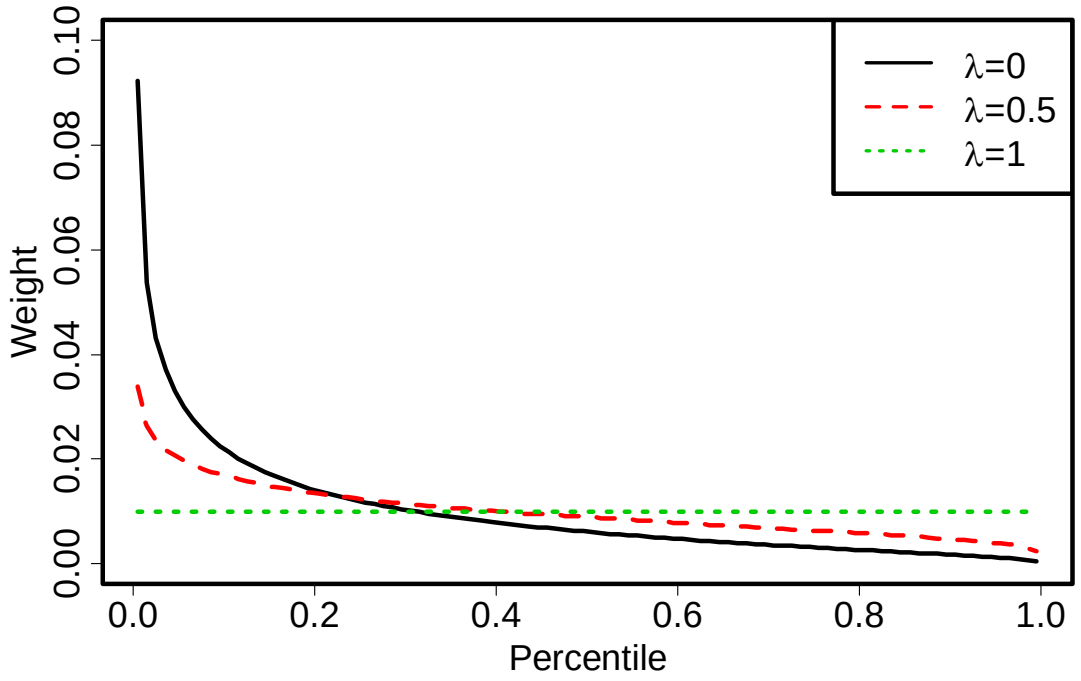

Figure 6.2 Weight vs. Percentile: for power utility ($\lambda = 0, 0.5, 1$) and empirical distribution (sample size=100).

Generalized Weight Function

Comparing the weight functions from the following two cases:

1. Negative Exponential Utility, Standard Normal Distribution (F_0) and Empirical Distribution (F_R).
2. Power/Logarithm Utility and Standard Lognormal Distribution (F_0) and Empirical Distribution (F_W).

we find they have a uniform form

$$s_i = N_{-\mu,1}\left(\Phi^{-1}\left(i/n\right)\right) - N_{-\mu,1}\left(\Phi^{-1}\left(\left(i-1\right)/n\right)\right) \qquad \mu \geq 0, i = 1 \ldots n \qquad (6.45)$$

where $N_{a,b}\left(\cdot\right)$ is the cumulative distribution function for normal distribution with $\mu = a, \sigma^2 = b$ and n is the sample size in the empirical distribution.

This weight function is free of any distribution parameters and solely depends on the risk-averse parameter v. If we have the non-shifted version of (6.45),

$$\Phi\left(\Phi^{-1}\left(i/n\right)\right)-\Phi\left(\Phi^{-1}\left(\left(i-1\right)/n\right)\right)=\frac{1}{n},$$

the result is for the risk neutral investors. (6.45) is a shifted version with the cumulative distribution function shifted to the left. In this way, the left tail gets more weight and the right tail gets less. We plot an illustration diagram in Fig. 6.5 to see this effect.

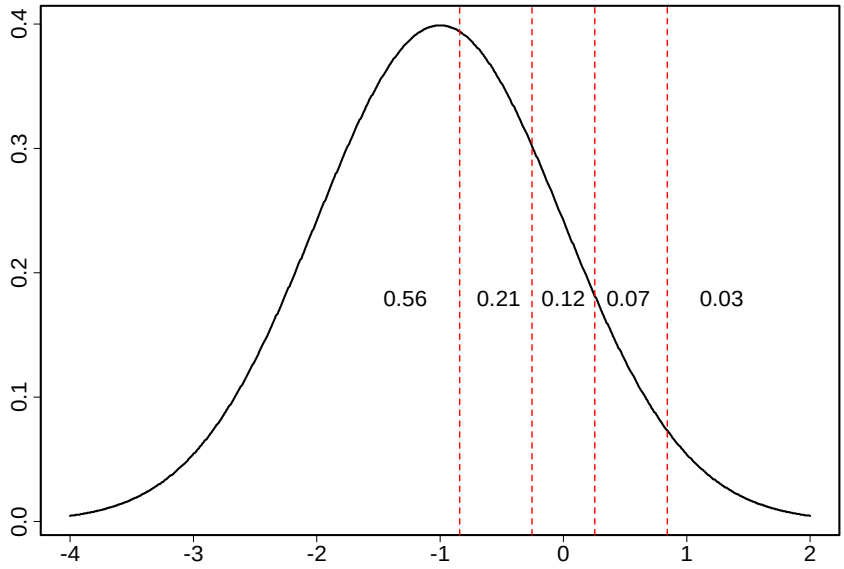

Figure 6.3 Illustration of (6.45): $\mu=1$ and $n=5$. The solid black curve is the shifted normal distribution and the red dashed line is the empirical quantiles mapped on the standard normal distribution. The marked numbers are the weights calculated using (6.45)

This effect implies we have essentially applied change-of-measure in our construction for SRM weights. We transformed the physical probability measure of the wealth or return to the risk-averse measure associated with the utility function. We will discuss the properties our utility based SRM with different perspectives in the following sections.

6.4 Second-Order Stochastic Dominance and Pessimistic Risk

Second-Order Stochastic Dominance

For random variables X and, X dominates Y by second-order stochastic dominance (SSD, written $X \succ_2 Y$) if

$$\int_{-\infty}^{x} F_X(t)\,dt \le \int_{-\infty}^{x} F_Y(t)\,dt, \qquad \forall x \in R \tag{6.46}$$

and the inequality is strict for at least one $x \in R$. It is proved in Levy (1998) that $X \succ_2 Y$ iff

$$E\left[U(X)\right] \ge E\left[U(Y)\right] \tag{6.47}$$

for all $U \in U_2$ and there exists at least one element in U_2 such the inequality is strict, where U_2 denotes the set of real-valued increasing, concave utility functions. Hence a risk-averse expected utility maximizing investor would always favor a risk measure that is isotonic with respect to second-order stochastic dominance, i.e.

$$X \succ_2 Y \Rightarrow Risk(X) \le Risk(Y) \tag{6.48}$$

De Giorgi(2004) discussed the standards that risk and reward measures should satisfy in order to be isotonic with respect to second order stochastic dominance. He proved that if g is non-decreasing and convex, then the risk measure R is isotonic with respect to the second-order stochastic dominance, where R is defined as

$$R = -\int x g'\left(\bar{F}_X(x)\right) dF_X(x) \tag{6.49}$$

and

$$\bar{F}_X(x) = 1 - F_X(x) \tag{6.50}$$

Comparing with our utility based SRM, we can have our SRM isotonic with respect to the second-order stochastic dominance if we define

$$g(p) = \frac{\int_0^p U'\left(F_0^{-1}(1-t)\right)dt}{\int_0^1 U'\left(F_0^{-1}(t)\right)dt}. \tag{6.51}$$

We can easily check that $g(0) = 0$ and $g(1) = 1$. Also we have

$$\begin{aligned} g'(p) &= \frac{U'\left(F_0^{-1}(1-p)\right)}{\int_0^1 U'\left(F_0^{-1}(t)\right)dt} \geq 0 \\ g''(p) &= \frac{U''\left(F_0^{-1}(1-p)\right)}{\int_0^1 U'\left(F_0^{-1}(t)\right)dt} \frac{-1}{f_0\left(F_0^{-1}(1-p)\right)} \geq 0. \end{aligned} \tag{6.52}$$

Hence g is non-decreasing and convex. We now can claim

$$\begin{aligned} R &= -\int x g'\left(\bar{F}_X(x)\right) dF_X(x) = -\int x \frac{U'\left(F_0^{-1}(t)\right)}{\int_0^1 U'\left(F_0^{-1}(t)\right)dt} dt \\ &= -\frac{\int_0^1 x U'\left(F_0^{-1}(t)\right)dt}{\int_0^1 U'\left(F_0^{-1}(t)\right)dt} \end{aligned} \tag{6.53}$$

is isotonic with respect to the second-order stochastic dominance. Therefore if an investment is preferred to another by all risk-averse expected utility maximizing investors, it is always less risky under our proposed SRM measure.

In De Giorgi(2004), it is also proved that such a measure like R is comonotoic[4] additive, which implies that it is not possible to diversify risk (under R) for comonotonic random variables (random variables that move together, not necessarily a linear relationship). No hedge is possible under this circumstance. Therefore our SRM also has this property.

Another key feature of our specific SRM is the utility associated distribution function F_0. We use it to map percentiles from real physical probability distributions to the utility function. It serves as a distortion function that translates the utility function based upon a standard F_0 for any arbitrary physical probability distribution. This feature is in some sense closer to the cumulative prospect theory developed by Kahneman and Tversky (1992), where the physical probability is distorted by a weight function and the reference point is different for different investors. The key difference is that we still maintain the utility theory's traditional risk-averse feature.

Pessimistic Risk

A different perspective from above is as the following, if we define

$$\upsilon(p) = \frac{\int_0^p U'\left(F_0^{-1}(t)\right)dt}{\int_0^1 U'\left(F_0^{-1}(t)\right)dt}, \tag{6.54}$$

then we can rewrite our risk measure as

[4] Two random variables X and Y are said to be comonotonic, if $(X,Y) =_d \left(F_X(U), F_Y(U)\right)$, where $=_d$ stands for equality in distribution and U is a uniform random variable . See Jan Dhaene, Steven Vanduffel and Marc Goovaerts (2007) for details

$$
\begin{aligned}
R &= -\int x \frac{U'\left(F_0^{-1}(p)\right)}{\int_0^1 U'\left(F_0^{-1}(p)\right)dp} dp = -\int x\upsilon'(p)\,dp \\
&= -\int x d\upsilon(p) \\
&= -\int F_X^{-1}(p)\, d\upsilon(p) \\
&= -E_\upsilon(x)
\end{aligned}
\tag{6.55}
$$

This is a type of Choquet expected utility. In Choquet expected utility theory, investors are allowed to distort their preferences from the original probability distribution of events. A simple expected utility

$$E_X\left[U(x)\right] = \int_{-\infty}^{\infty} U(x)\, dF(x) = \int_0^1 U\left(F^{-1}(p)\right)dp$$

can be rewritten as

$$E_{X,\upsilon}\left[U(x)\right] = \int_0^1 U\left(F^{-1}(p)\right)d\upsilon(p)$$

where we replace the Lebesgue measure dp by the distortion probability measure $d\upsilon$. In particular, a general class of Choquet risk measure is defined as

$$\rho(X) = -\int_0^1 F_X^{-1}(p)\, d\upsilon(p)$$

Therefore (6.55) also is a Choquet risk measure. Bassett and etl (2004) defined pessimistic measures as the special case of the Choquet risk measures where υ is concave and decreasing on $[0,1]$. The measure υ in (6.55) is non-decreasing and concave ($\upsilon'(p) \geq 0, \upsilon''(p) \leq 0$). Therefore it is also a pessimistic measure.

6.5 S-Shaped Utility and Bayesian View

In section 6.3, we showed that the utility based SRM under some circumstances is calculated under a shifted normal distribution, hence a change-of-measure. In section 6.4, we showed it can also be viewed as a pessimistic measure. The measure changing procedure essentially shifts more mass on the original distribution to the left tail. The driving force of this procedure is the risk-averse utility function. We now take a different perspective to examine another interesting class of utility functions: the S-shaped utility functions. It is not risk-averse. Therefore the risk measure discussed below is not a coherent or spectral risk measure. Nevertheless, our procedure of constructing utility based SRM still can be applied.

Usually the S-shaped utility function is asymmetric. Just for illustration purpose and simplicity, we would limit to a simpler version of S-shaped Utility. Suppose

$$\begin{aligned} U &= \Phi \\ U' &= \varphi, \end{aligned} \tag{6.56}$$

the standard normal distribution and density functions. Our risk profile looks like the following (Fig. 6.6)

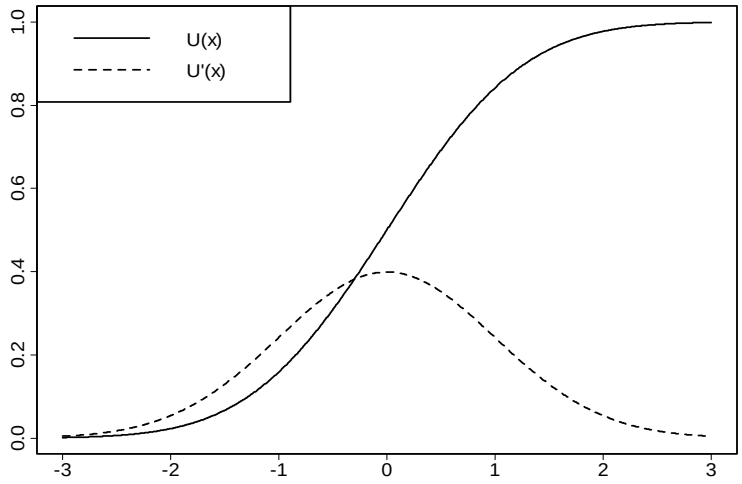

Figure 6.4 S-Shaped Utility Function

The investors always prefer more to less, but are risk loving when having losses and are risk averse when having profits. They are risk neutral (indifferent) when they are even (or at a chosen target). Such a utility function will encourage risk-seeking behavior (concentrated investments) most when we are close on to our target. The further away from that level, we will be more conservative (diversified investments).

This view is best captured under our utility based SRM. Suppose we select the S-shaped utility function and standard normal distribution to interpret physical returns

$$U(r) = N_{\mu,\sigma^2}(r)$$
$$F_0(q) = \Phi(q) = F_R(r).$$

Following (6.55),

$$
\begin{aligned}
R &= -E_{\upsilon}(x) \\
&= -\int x \frac{N'_{\mu,\sigma^2}\left(F_0^{-1}(p)\right)}{\int_0^1 N'_{\mu,\sigma^2}\left(F_0^{-1}(p)\right)dp} dp \\
&= -\int F_R^{-1}\left(\Phi(q)\right) \frac{N'_{\mu,\sigma^2}(q)\varphi(q)}{\int N'_{\mu,\sigma^2}(q)\varphi(q)\,dq} dq.
\end{aligned}
\tag{6.57}
$$

Note $N'_{\mu,\sigma^2}(q) = N'_{q,\sigma^2}(\mu)$, hence

$$
\begin{aligned}
&\frac{N'_{\mu,\sigma^2}(q)\varphi(q)}{\int N'_{\mu,\sigma^2}(q)\varphi(q)\,dq} \\
&= \frac{N'_{\mu,\sigma^2}(\mu)\varphi(q)}{\int N'_{\mu,\sigma^2}(\mu)\varphi(q)\,dq} \\
&= \Pr(q \mid \mu)
\end{aligned}
\tag{6.58}
$$

is the posterior probability of q given μ and assuming σ^2 is known. Consequently,

$$
\begin{aligned}
R &= -\int F_R^{-1}\left(\Phi(q)\right) \cdot \Pr(q \mid \mu) \cdot dq \\
&= -E\left[F_R^{-1}\left(\Phi(q)\right) \mid \mu\right] \\
&= -E\left[r \mid \mu\right]
\end{aligned}
\tag{6.59}
$$

which is the negative posterior expected return given our S-shaped utility function! When we try to minimize R, we maximize the posterior expected value. In this case, we are trying to shrink towards the mean μ, since we have assumed the normal prior distribution. We view μ as our priori view of target returns and assume the standard normal distribution for q. We may optimize our portfolio such that the return r to be as close to μ as we can. This is in agreement with the Bayesian analysis for portfolio optimization. It does not encourage investing in extreme good or bad portfolios. Instead it is most interested in reaching the predefined target.

6.6 *Influence Function of SRM*

In this section, we study the influence function of utility based SRM. It is a direct approach to understand how this risk measure responds to the changes in the return distribution. We will first go over the concepts of influence functions and work through the general form of influence functions for utility based SRM. We will then give two concrete cases under ETL and exponential utility.

Define the mixture distribution

$$F_\gamma = (1-\gamma)F + \gamma \cdot \delta_c, \quad 0 \le \gamma < 1 \tag{6.60}$$

where δ_c is a point mass probability distribution located at the contamination point c.

Referring to the functional $\theta(F)$ loosely by $\hat{\theta}$, the influence function (IF) of $\hat{\theta}$ at F is defined by

$$\mathrm{IF}\left(c;\hat{\theta},F\right) = \lim_{\gamma \downarrow 0} \frac{\theta\left(F_\gamma\right) - \theta(F)}{\gamma} = \frac{d}{d\gamma}\theta\left(F_\gamma\right)\Big|_{\gamma=0}. \tag{6.61}$$

For the utility based SRM, we have

$$\begin{aligned} \hat{\theta} = M_\zeta &= -\frac{E_Q\left(r \cdot U'(q_r)\right)}{E_Q\left(U'(q_r)\right)} \\ &= -\frac{E_R\left(r \cdot U'\left(F_0^{-1}\left(F_R(r)\right)\right)\right)}{E_R\left(U'\left(F_0^{-1}\left(F_R(r)\right)\right)\right)}, \end{aligned} \tag{6.62}$$

where $q_r = F_0^{-1}\left(F_R(r)\right)$. Therefore

$$\mathrm{IF}(c; M_\zeta, F_R) = -\frac{d}{d\gamma}\frac{E_Q\left(r \cdot U'\left(q_{\gamma,r}\right)\right)}{E_Q\left(U'\left(q_{\gamma,r}\right)\right)}\bigg|_{\gamma=0} \quad , \tag{6.62}$$

where $q_{\gamma,r} = F_0^{-1}\left(F_{R,\gamma}\left(r\right)\right) = F_0^{-1}\left(\left(1-\gamma\right)F_R + \gamma \cdot \delta_c\right)$.

First we note that

$$\frac{d}{d\gamma}E_Q\left(U'\left(q_{\gamma,r}\right)\right)\bigg|_{\gamma=0} = 0, \tag{6.63}$$

since the expected value of U' is constant despite of changing $F_{R,\gamma}$ (see examples in section 6.3). Equation (6.62) becomes

$$\mathrm{IF}(c; M_\zeta, F_R) = -\frac{\frac{d}{d\gamma}E_Q\left(r \cdot U'\left(q_{\gamma,r}\right)\right)\Big|_{\gamma=0}}{E_Q\left(U'\left(q_r\right)\right)}. \tag{6.64}$$

We then derive the IF for $U'\left(q_r\right)$ as follows

$$\begin{aligned} IF\left(c; U'\left(q_r\right), F_R\right) &= \frac{d}{d\gamma}U'\left(q_{\gamma,r}\right)\bigg|_{\gamma=0} \\ &= U''\left(q_r\right) \cdot IF\left(c; q_r, F_R\right) \quad , \end{aligned} \tag{6.65}$$

In order to get $IF\left(c; q_r, F\right)$, we construct the following identity

$$C_r = \int_{-\infty}^{r} dF_{R,\gamma}. \tag{6.66}$$

(6.66) is a constant denoting the tail probability of the mixture distribution up to r. Recall the calculus of taking derivatives under the integral

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f\left(x,t\right)dt = \int_{a(x)}^{b(x)} f_x\left(x,t\right) \cdot dt + b'\left(x\right) f\left(x, b\left(x\right)\right) - a'\left(x\right) f\left(x, a\left(x\right)\right). \tag{6.67}$$

Taking derivatives with respect to γ on both sides of (6.66) and evaluate at $\gamma = 0$,

$$\begin{aligned} C_r &= \int_{-\infty}^{r} dF_{R,\gamma} = \int_{-\infty}^{q_{\gamma,r}=F_0^{-1}\left(F_{R,\gamma}(r)\right)} dF_0 \\ 0 &= IF\left(c;q_r,F_R\right)\cdot f_0\left(q_r\right) + \int_{-\infty}^{r} d\left[\delta_c - F_R\right] \\ 0 &= IF\left(c;q_r,F_R\right)\cdot f_0\left(q_r\right) - F_R\left(r\right) + I_{(-\infty,r]}\left(c\right) \\ 0 &= IF\left(c;q_r,F_R\right)\cdot f_0\left(q_r\right) - F_0\left(q_r\right) + I_{(-\infty,r]}\left(c\right) \\ IF\left(c;q_r,F_R\right) &= -\frac{I_{(-\infty,r]}\left(c\right) - F_0\left(q_r\right)}{f_0\left(q_r\right)}. \end{aligned} \tag{6.68}$$

In (6.68), we performed the first part of (6.67) by using the identity $C_r = \int_{-\infty}^{q_{\gamma,r}=F_0^{-1}\left(F_{R,\gamma}(r)\right)} dF_0$ and calculated the rest of (6.67) by using $C_r = \int_{-\infty}^{r} dF_{R,\gamma}$.

Now we can proceed with

$$\begin{aligned} &\frac{d}{d\gamma} E_Q\left(r\cdot U'\left(q_{\gamma,r}\right)\right)\Big|_{\gamma=0} \\ &= \frac{d}{d\gamma} E_R\left(r\cdot U'\left(F_0^{-1}\left(F_{R,\gamma}\left(r\right)\right)\right)\right)\Big|_{\gamma=0} \\ &= \int r\cdot IF\left(c;U'\left(q_r\right),F_R\right) dF_R \\ &\qquad + \int r\cdot U'\left(F_0^{-1}\left(F_R\left(r\right)\right)\right)\cdot d\left[\delta_c\left(r\right) - F_R\left(r\right)\right] \\ &= \int r\cdot IF\left(c;U'\left(q_r\right),F_R\right) dF_R \\ &\qquad + c\cdot U'\left(F_0^{-1}\left(F_R\left(c\right)\right)\right) - E_R\left(r\cdot U'\left(q_r\right)\right). \end{aligned} \tag{6.69}$$

The first term can be expanded with (6.68) to

$$
\begin{aligned}
&\int r \cdot IF\left(c; U'(q_r), F_R\right) dF_R(r) \\
&= -\int r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)} dF_R(r) \\
&= -\int r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)} dF_0(q_r) \\
&= -E_Q\left(r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)} \right)
\end{aligned}
$$

Finally, plug in the influence function (6.64)

$$
\begin{aligned}
\mathrm{IF}(c; M_\zeta, F_R) &= -\frac{\frac{d}{d\gamma} E_Q\left(r \cdot U'\left(q_{\gamma,r}\right)\right)\Big|_{\gamma=0}}{E_Q\left(U'(q_r)\right)} \\
&= \frac{1}{E_Q\left(U'(q_r)\right)} E_Q\left(r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)} \right) \\
&\quad - c \cdot U'\left(F_0^{-1}\left(F_R(c)\right)\right) \Big/ E_Q\left(U'(q_r)\right) + E_R\left(r \cdot U'(q_r)\right) \Big/ E_Q\left(U'(q_r)\right) \\
&= \frac{1}{E_Q\left(U'(q_r)\right)} \cdot \left\{ E_Q\left(r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)} \right) - c \cdot U'(q_c) \right\} - M_\zeta,
\end{aligned}
\tag{6.70}
$$

where q_c is shorthand for $F_0^{-1}\left(F_R(c)\right)$.

6.6.1 Influence Function for ETL

In the case of ETL, the utility based SRM takes the following assumptions

$$
\begin{aligned}
U(t) &= t \cdot I_{(0,\alpha]}(t) - \alpha \\
U'(t) &= I_{(0,\alpha]}(t) \\
U''(t) &= \delta_\alpha \\
F_0(q) &= q \\
f_0(q) &= 1
\end{aligned}
\qquad , \tag{6.71}
$$

where the utility function is two-piece linear function and the associated F_0 is uniform.

Following (6.70), we get

$$
\begin{aligned}
&\mathrm{IF}(c; M_\zeta, F) \\
&= \frac{1}{E_Q\left(U'(q_r)\right)} \cdot \left\{ E_Q\left(r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)} \right) - c \cdot U'(q_c) \right\} - M_\zeta, \\
&= \frac{1}{\alpha} \cdot \left\{ -VaR_{1-\alpha} \cdot \left(I_{(-\infty,-VaR_{1-\alpha}]}(c) - \alpha \right) - c \cdot I_{(-\infty,-VaR_{1-\alpha}]}(c) \right\} - ETL_{1-\alpha} \\
&= VaR_{1-\alpha} - ETL_{1-\alpha} - \frac{I_{(-\infty,-VaR_{1-\alpha}]}(c)}{\alpha} \cdot \left(VaR_{1-\alpha} + c \right)
\end{aligned}
\tag{6.72}
$$

where VaR and ETL are defined on the original space or r or w. Note when $F_0(q_r) = q_r = \alpha = F_R(r)$, we have $r = -VaR_{1-\alpha}$. We also used the identity $I_{(-\infty,-VaR_{1-\alpha}]}(c) = I_{(0,\alpha]}(q_c)$.

We plot (6.72) using $\alpha = 0.95$ and assume standard normal distribution for F_R in Fig. 6.7.

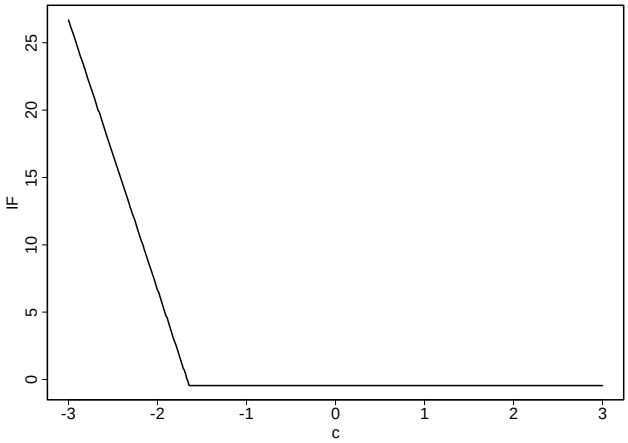

Figure 6.5 Influence Function of ETL: $\alpha = 0.95, F_R = \Phi$

We see that the influence vanishes above the 5% quantile on F_R. The influence below this point is linearly increasing as the loss becomes larger. This indicates heavier losses would result in a much larger ETL risk.

6.6.2 Exponential Utility Derived SRM

In the case of exponential utility, we have the following assumptions

$$\begin{aligned} U(t) &= -e^{-\lambda t} \\ U'(t) &= \lambda e^{-\lambda t} \\ U''(t) &= -\lambda^2 e^{-\lambda t} \\ F_0(q) &= \Phi(q) \\ F_0(r) &= N_{\mu,\sigma^2}(r), \end{aligned} \tag{6.73}$$

where we assumed normal distribution for both F_0 and F_R.

We first derive the parts that will be used in (6.70)

$$E_Q\left(U'(q_r)\right) = E_Q\left(\lambda e^{-\lambda q}\right) = \lambda \exp\left\{\frac{1}{2}\lambda^2\right\}$$

$$c \cdot U'(q_c) = c \cdot \lambda \exp\left\{-\lambda \frac{c-\mu}{\sigma}\right\}$$

and

$$E_Q\left(r \cdot U''(q_r) \cdot \frac{I_{(-\infty,r]}(c) - F_0(q_r)}{f_0(q_r)}\right)$$

$$= \int (\mu + q_r\sigma) \cdot U''q_r \cdot \left[I_{(-\infty,r]}(c) - \Phi(q_r)\right] dq_r$$

$$= \int (\mu + q_r\sigma) \cdot U''(q_r) \cdot \left[I_{[q_c,\infty)}(q_r) - \Phi(q_r)\right] dt$$

$$= \int_{q_c}^{\infty} (\mu + q_r\sigma) \cdot U''(q_r)\, dq_r - \int (\mu + q_r\sigma) \cdot U''(q_r) \cdot \Phi(q_r)\, dq_r$$

$$= r \cdot U'(q_r)\Big|_{q_c}^{\infty} - \sigma \cdot \int_{q_c}^{\infty} U'(q_r) dq_r$$

$$- r \cdot U'(q_r) \cdot \Phi(q_r)\Big|_{-\infty}^{\infty} + \int r \cdot U'(q_r)\, d\Phi(q_r)$$

$$+ \sigma \cdot \left(\Phi(q_r) \cdot U(q_r)\Big|_{-\infty}^{\infty} - \int U(q_r)\, d\Phi(q_r)\right)$$

$$= -cU'(q_c) + \sigma \cdot U(q_c) + \left(\mu + \sigma \cdot E_\Phi\left(q_r \cdot U'(q_r)\right)\right) - \sigma \cdot E_\Phi\left(U(q_r)\right).$$

Finally we have

$$
\begin{aligned}
&\mathrm{IF}(c; M_\zeta, F) \\
&= \frac{1}{E_Q\left(U'(q_r)\right)} \cdot \left\{ E_Q\left(r \cdot U''(q_r) \cdot \frac{I_{(-\infty, r]}(c) - F_0(q_r)}{f_0(q_r)} \right) - c \cdot U'(q_c) \right\} - M_\zeta, \\
&= \frac{1}{\lambda \exp\left\{\frac{1}{2}\lambda^2\right\}} \cdot \left\{ \begin{array}{l} \sigma \cdot U(q_c) + \left(\mu + \sigma \cdot E_\Phi\left(q_r \cdot U'(q_r)\right)\right) \\ -\sigma \cdot E_\Phi\left(U(q_r)\right) - 2c \cdot \lambda \exp\left\{-\lambda \frac{c-\mu}{\sigma}\right\} \end{array} \right\} + (\mu - \lambda\sigma) \\
&= \frac{1}{\lambda \exp\left\{\frac{1}{2}\lambda^2\right\}} \cdot \left\{ -\sigma \cdot \exp\left(-\lambda \frac{c-\mu}{\sigma}\right) + \left(\mu - \sigma \cdot \lambda^2 \exp\left\{\frac{1}{2}\lambda^2\right\}\right) \right. \\
&\qquad \left. + \sigma \cdot \exp\left\{\frac{1}{2}\lambda^2\right\} - 2c \cdot \lambda \exp\left\{-\lambda \frac{c-\mu}{\sigma}\right\} \right\} + (\mu - \lambda\sigma) \\
&= \mu\left(1 + \frac{1}{\lambda} e^{-0.5\lambda^2}\right) - \sigma\left(2\lambda - \frac{1}{\lambda}\right) - \left(2c + \frac{\sigma}{\lambda}\right) e^{-\lambda \cdot (c-\mu)/\sigma - 0.5\lambda^2}
\end{aligned}
\tag{6.74}
$$

If we take the exponential SRM above with $\lambda = 1$ and pick standard normal for F_R, we get

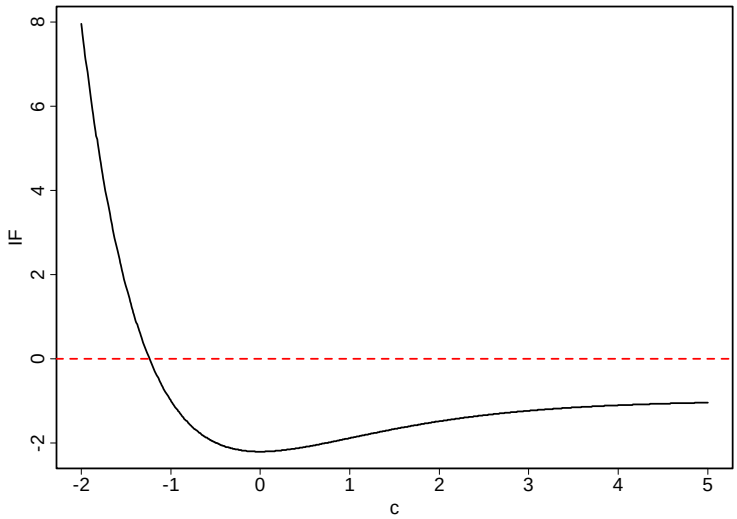

Figure 6.6 Influence Function of Exponential Utility Based SRM

Based on Fig. 6.8, we find the downside is much riskier than the upside under this risk measure. The influence on the upside tends to level off above certain threshold. This shows that the investor is indifferent in terms of risk after seeing returns above that level.

This agrees with the negative exponential utility function. On the downside, the utility function changes fast even with little change in returns. This is also reflected in this influence function. The investor feels much safer even if the loss is only slightly smaller.

We also find the influence is minimized at 0. This means that the risk is reduced when more positive returns are added, but the largest reduction effect is at 0. More positive returns will still reduce risk but less effective. This is an example of SRM under Bayesian view. We set $F_0 = \Phi$ in our assumptions. This implies we view the best target return is 0. That's where the strongest mitigation effect happens.

Chapter 7
Portfolio Optimization with SRM, SMCR and SSD

In this chapter, we discuss the portfolio optimization results with the following methods: Spectral Risk Measure (SRM), Second Moment Coherent Measure (SMCR) and Semi-Standard Deviation (SSD). In 7.1, we will discuss applying the general convex programming method, especially the Second-Order Cone Programming (SOCP), to some portfolio optimization problems. In 7.2, we will discuss the optimization results.

7.1 Optimization Methods

The general optimization problem usually can be framed as

$$\begin{aligned} &\text{minimize} && f_0(x) \\ &\text{subject to} && f_i(x) \le 0, \quad i = 1,\ldots,m \\ &&& h_i(x) = 0, \quad i = 1,\ldots,p. \end{aligned} \tag{7.1}$$

If all f_i are convex and all h_i are affine, then the problem (7.1) is a convex programming problem. It will have only one minimum. Therefore any local minimum is a global minimum. A lot of techniques have been developed to solve this type of problems. (7.1) is also sub-typed into different problems, such as linear programming (LP), quadratic programming (QP), second-order cone programming (SOCP) and etc.

7.1.1 Second-Order Cone Programming (SOCP)

A set $S \subset R^n$ is a convex cone if

$$x, y \in S, \quad \lambda, \mu \ge 0 \Rightarrow \lambda x + \mu y \in S \tag{7.2}$$

It contains all rays passing through its points which emanate from the origin. A second-order cone is the Euclidean norm cone, i.e.

$$S = \left\{ (\mathbf{x}, t) \,|\, \|\mathbf{x}\| \le t \right\} \tag{7.3}$$

where $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$.

A SOCP problem can be formulated as

$$\begin{array}{lll} \text{minimize} & f^T x & \\ \text{s.t.} & \|A_i x + b_i\| \le c_i^T x + d_i & i = 1 \ldots L \end{array} \tag{7.4}$$

where $\|\cdot\|$ is the Euclidian norm.

7.1.2 SMCR and SSD

In Krokhmal (2007), the global minimum SMCR portfolio can be obtained by solving the following SOCP problem,

$$\begin{array}{lll} \text{minimize} & \eta + \dfrac{1}{1-\alpha}\dfrac{1}{\sqrt{T}} t & \\ \text{s.t.} & \sum_{i=1}^{n} w_i = 1 & \\ & w_i \ge 0 & i = 1 \ldots N \\ & s_j \ge 0 & j = 1 \ldots T \\ & s_j \ge -\eta - \sum_{i=1}^{N} w_i r_{ij} & j = 1 \ldots T \\ & t \ge \left(s_1^2 + \ldots + s_T^2 \right)^{1/2} & \end{array} \tag{7.5}$$

where α represents the confidence level and η is the cutoff point, which changes with different weights $\mathbf{w}$. N and T are the number of assets and training periods, respectively.

Krokhmal (2007) explained that the SMCR measure can be solved via the Second-Order Cone Programming (SOCP). We formally present the problem by matching (7.5) to (7.4) as follows,

$$x = \begin{bmatrix} \eta & t & w_1 \dots w_N & s_1 \dots s_T \end{bmatrix}$$
$$f = \begin{bmatrix} 1 & \dfrac{1}{\alpha\sqrt{T}} & 0 \dots 0 & 0 \dots 0 \end{bmatrix} \tag{7.6}$$

and

$$A = \begin{bmatrix} 0_{(2+N+2T)\times(2+N)} & 0_{(2+N+2T)\times T} \\ 0_{T\times(2+N)} & I_{T\times T} \end{bmatrix} \qquad b = 0_{(2+N+3T)\times 1}$$
$$c = \begin{bmatrix} 0 & 0 & 1_{1\times N} & 0_{1\times T} \\ 0 & 0 & -1_{1\times N} & 0_{1\times T} \\ 0_{N\times 1} & 0_{N\times 1} & I_{N\times N} & 0_{N\times T} \\ 0_{T\times 1} & 0_{T\times 1} & 0_{T\times N} & I_{T\times T} \\ 1_{T\times 1} & 0_{T\times 1} & R_{T\times N} & I_{T\times T} \\ 0 & 1 & 0_{1\times N} & 0_{1\times T} \end{bmatrix} \qquad d = \begin{bmatrix} -1 \\ 1 \\ 0_{N\times 1} \\ 0_{T\times 1} \\ 0_{T\times 1} \\ 0 \end{bmatrix} \tag{7.7}$$

The last T rows of A and b form the norm inequality constraint and correspond to the last row of C and d. The rest of constraints are 1-to-1 by row.

The *Rdonlp2* package in R uses *donlp2* algorithm (non-linear programming method developed by P. Spellucci (1997)) to solve non-linear programming problems. We applied *Rdonlp2* to the problem (7.6) and (7.7) and successfully solved (7.5).

Semi-Standard Deviation can be defined as

$$SSD = \sqrt{\int_{-\infty}^{\mu} (\mu - r)^2 \, dF(r)} \tag{7.8}$$

where μ is the mean of return r. The global minimum SSD portfolio can be obtained by solving the following SOCP problem

$$\begin{aligned}
&\text{minimize} && t && \\
&\text{s.t.} && \sum_{i=1}^{n} w_i = 1 && \\
& && w_i \geq 0 && i = 1 \ldots N \\
& && s_j \geq 0 && j = 1 \ldots T \\
& && s_j \geq \sum_{i=1}^{N} w_i \left(\bar{r}_{i.} - r_{ij} \right) && j = 1 \ldots T \\
& && t \geq \left(s_1^2 + \ldots + s_T^2 \right)^{1/2} &&
\end{aligned} \tag{7.9}$$

where s_i's are auxiliary variables. t is the semi-standard deviation once minimized and rest variables are defined the same way as previously. Formally, we match (7.9) with (7.4) as follows

$$\begin{aligned}
x &= \begin{bmatrix} t & w_1 \ldots w_N & s_1 \ldots s_T \end{bmatrix} \\
f &= \begin{bmatrix} 1 & 0 \ldots 0 & 0 \ldots 0 \end{bmatrix}
\end{aligned} \tag{7.10}$$

and

$$\begin{aligned}
A &= \begin{bmatrix} 0_{(2+N+2T)\times(1+N)} & 0_{(2+N+2T)\times T} \\ 0_{T\times(1+N)} & I_{T\times T} \end{bmatrix} \qquad b = 0_{(2+N+3T)\times 1} \\
c &= \begin{bmatrix} 0 & 1_{1\times N} & 0_{1\times T} \\ 0 & -1_{1\times N} & 0_{1\times T} \\ 0_{N\times 1} & I_{N\times N} & 0_{N\times T} \\ 0_{T\times 1} & 0_{T\times N} & I_{T\times T} \\ 0_{T\times 1} & R_{T\times N} - 1_{T\times 1} \times \bar{R}_{1\times N} & I_{T\times T} \\ 1 & 0_{1\times N} & 0_{1\times T} \end{bmatrix} \qquad d = \begin{bmatrix} -1 \\ 1 \\ 0_{N\times 1} \\ 0_{T\times 1} \\ 0_{T\times 1} \\ 0 \end{bmatrix}
\end{aligned} \tag{7.11}$$

It can also be solved by *Rdonlp2*.

7.1.3 SRM

The utility based spectral risk measures that we derived in Chapter 6 are convex measures because it's created following the rules in Acerbi (2002). In fact it can always be expressed as the linear combination of ETL as follows

$$\begin{aligned} SRM &= \sum_{i=1}^{N} w_i r_{(i)} = \sum_{i=1}^{N} \left[\left(w_i - w_{i+1} \right) \sum_{j=1}^{i} r_{(j)} \right] \\ &= \sum_{i=1}^{N} \left[\left(w_i - w_{i+1} \right) \cdot i \cdot ETL_{1-i/N} \right] \\ &\triangleq \sum_{i=1}^{N} \left[\pi_i \cdot ETL_{1-i/N} \right] \end{aligned} \tag{7.12}$$

where we define $w_{N+1} = 0$. Note $\pi_i \triangleq w_i - w_{i+1} \geq 0$ by Acerbi's rules. Therefore SRM is a convex combination of convex measures, hence convex. It can be solved by *Rdonlp2*, too.

7.2 Portfolio Optimization Results

7.2.1 Confidence Level

We targeted for 95% confidence level for all tail risk measures tested in Chapter 3 and 4. In this experiment, we will target both 95% and 75% confidence levels.

In order to match that risk confidence level for different risks, we try to match SMCR and SRM to the level of ETL. For SMCR, based on Krokhmal's (2007) research,

$$SMCR_{\alpha} \geq ETL_{\alpha^*} \tag{7.13}$$

when $\alpha^* = 2\alpha - \alpha^2$. If we want to match SMCR to ETL at 95% confidence level, we can set $\alpha^* = 0.95$, and solve $\alpha^* = 2\alpha - \alpha^2$ to get $\alpha = 0.776$. If we set $\alpha^* = 0.75$, we find the corresponding level for SMCR is $\alpha = 0.5$.

For SRM, we don't have an available method to pick the corresponding confidence level. However, under exponential utility and normality assumptions,

$$SRM = -\left(\mu - \lambda\sigma\right) \tag{7.14}$$

Also under standard normal distribution, SRM, ETL and VaR can all be expressed in the form of (7.14), except for different λ

$$\lambda = \begin{cases} \text{exponential utility function parameter} & \text{SRM} \\ \varphi\left(\Phi^{-1}\left(1-\alpha\right)\right)\Big/\left(1-\alpha\right) & \text{ETL}. \\ \Phi^{-1}\left(1-\alpha\right) & \text{VaR} \end{cases} \tag{7.15}$$

Based on this view, we can match SRM to ETL at the 95% confidence level, by setting

$$\begin{aligned} \lambda &= \phi\left(\Phi^{-1}\left(1-\alpha\right)\right)\Big/\left(1-\alpha\right) \\ &= \phi\left(\Phi^{-1}\left(1-0.95\right)\right)\Big/\left(1-0.95\right) = 2.06 \end{aligned} \tag{7.16}$$

Therefore we choose $\lambda = 2.06$ to match the 95% confidence level ETL. The SRM weights are calculated using the formula we derived in Chapter 6 and assuming $n = 60$ for 5 years of monthly returns.

$$s_i = N_{-\lambda,1}\left(\Phi^{-1}\left(i/n\right)\right) - N_{-\lambda,1}\left(\Phi^{-1}\left(\left(i-1\right)/n\right)\right), \qquad i = 1\ldots 60$$

We show the SRM weights obtained and compare with the ETL weighting in the plot below. The SRM weights are higher than ETL's weight for the heavy losses and goes down sharply. The same procedure gives $\lambda = 1.27$ to match 75% confidence level ETL.

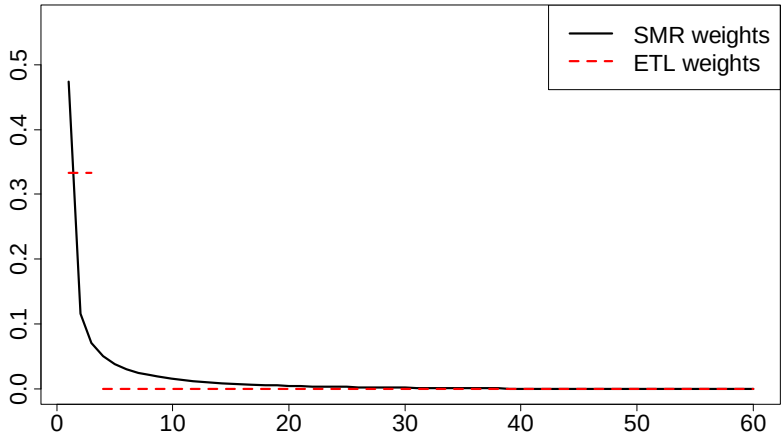

Figure 7.1 Weights of SMR and ETL at 95% confidence level

7.2.2 Experiments with 95% Confidence Level

In the following experiment, we will focus on the following 5 methods Minimum volatility (MinVol), Historical ETL (HETL), SRM (SRM), SSD (SSD), SMCR (SMCR). The names in the parenthesis are the code names we use in the plots and tables.

In table 7.1 and 7.2, we display the summary statistics on respective out-of-sample performance measures for each portfolio optimization method across all 100 portfolios we formed. In Fig. 7.2 and 7.3, we display the kernel densities of these statistics. In Fig. 7.4 and 7.5, we display the box-plots of these statistics by comparing the difference on the same portfolio by different methods. For better presentation, some of the outliers are not shown (outside the plotting region).

Table 7.1 Summary Statistics of Performance Measures of Experiments with 95% Confidence Level.

Annualized Mean Return						Diversification				
	MinVol	SSD	HETL	SMCR	SRM	MinVol	SSD	HETL	SMCR	SRM
Minimum	-0.137	-0.144	-0.061	-0.024	-0.016	0.409	0.397	0.615	0.607	0.597
25%	-0.008	-0.002	0.053	0.051	0.057	0.760	0.750	0.748	0.751	0.745
Median	0.033	0.036	0.107	0.104	0.105	0.814	0.806	0.782	0.788	0.781
Mean	0.031	0.042	0.108	0.106	0.110	0.789	0.779	0.772	0.779	0.774
75%	0.066	0.085	0.155	0.150	0.149	0.848	0.848	0.807	0.813	0.810
Maximum	0.225	0.251	0.415	0.461	0.511	0.893	0.889	0.856	0.860	0.855
Sharpe Ratio						Starr Ratio				
	MinVol	SSD	HETL	SMCR	SRM	MinVol	SSD	HETL	SMCR	SRM
Minimum	-1.220	-1.270	-0.594	-0.467	-0.522	-0.142	-0.162	-0.067	-0.054	-0.064
25%	-0.408	-0.321	0.211	0.198	0.261	-0.041	-0.037	0.026	0.025	0.031
Median	0.044	0.067	0.558	0.619	0.611	0.007	0.008	0.079	0.077	0.080
Mean	-0.005	0.073	0.647	0.627	0.662	0.008	0.019	0.084	0.082	0.085
75%	0.340	0.445	1.010	0.994	1.000	0.048	0.067	0.135	0.131	0.134
Maximum	1.150	1.540	2.420	2.690	2.590	0.284	0.277	0.337	0.333	0.376

Table 7.2 Summary Statistics of Risk Measures of Experiments with 95% Confidence Level.

Annualized Volatility						Semi. Std. Dev.				
	MinVol	SSD	HETL	SMCR	SRM	MinVol	SSD	HETL	SMCR	SRM
Minimum	0.057	0.056	0.066	0.065	0.067	0.064	0.063	0.072	0.074	0.074
25%	0.088	0.087	0.098	0.098	0.097	0.098	0.094	0.104	0.103	0.106
Median	0.104	0.104	0.113	0.113	0.114	0.113	0.111	0.126	0.124	0.126
Mean	0.118	0.124	0.133	0.133	0.133	0.118	0.115	0.127	0.125	0.127
75%	0.128	0.128	0.137	0.134	0.133	0.135	0.131	0.144	0.142	0.144
Maximum	0.359	0.458	0.723	0.845	0.901	0.206	0.212	0.228	0.206	0.241
Value at Risk						Expected Tail Loss				
	MinVol	SSD	HETL	SMCR	SRM	MinVol	SSD	HETL	SMCR	SRM
Minimum	0.020	0.015	0.011	0.010	0.010	0.038	0.036	0.035	0.034	0.036
25%	0.040	0.036	0.034	0.036	0.035	0.059	0.054	0.058	0.060	0.061
Median	0.049	0.045	0.046	0.046	0.043	0.069	0.067	0.073	0.072	0.075
Mean	0.050	0.048	0.045	0.046	0.045	0.073	0.071	0.076	0.074	0.075
75%	0.058	0.058	0.055	0.055	0.054	0.088	0.084	0.089	0.082	0.088
Maximum	0.098	0.099	0.096	0.102	0.097	0.139	0.146	0.147	0.128	0.141
Maximum Drawdown						Turnover				
	MinVol	SSD	HETL	SMCR	SRM	MinVol	SSD	HETL	SMCR	SRM
Minimum	0.098	0.081	0.075	0.066	0.067	0.051	0.054	0.064	0.081	0.069
25%	0.217	0.204	0.169	0.176	0.174	0.073	0.080	0.095	0.106	0.109
Median	0.271	0.264	0.247	0.249	0.255	0.080	0.088	0.105	0.119	0.123
Mean	0.280	0.269	0.240	0.239	0.241	0.079	0.088	0.108	0.123	0.124
75%	0.336	0.328	0.307	0.289	0.303	0.086	0.095	0.120	0.136	0.134
Maximum	0.542	0.555	0.435	0.423	0.430	0.102	0.119	0.155	0.196	0.191

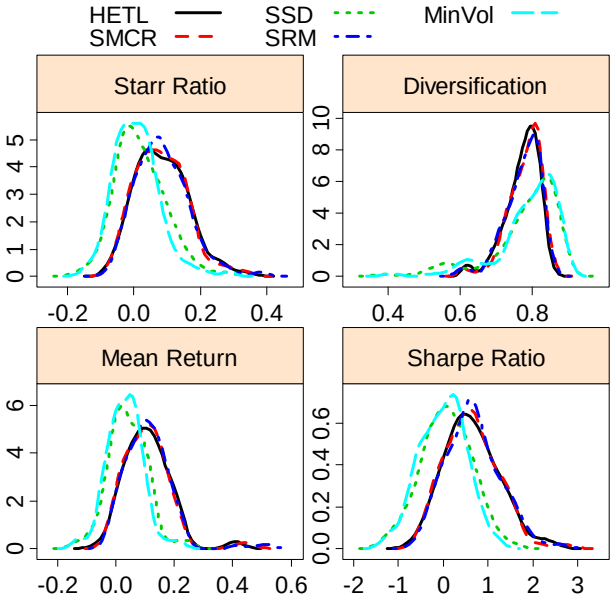

Figure 7.2 Kernel Densities of Performance Measures of Long Only Experiments

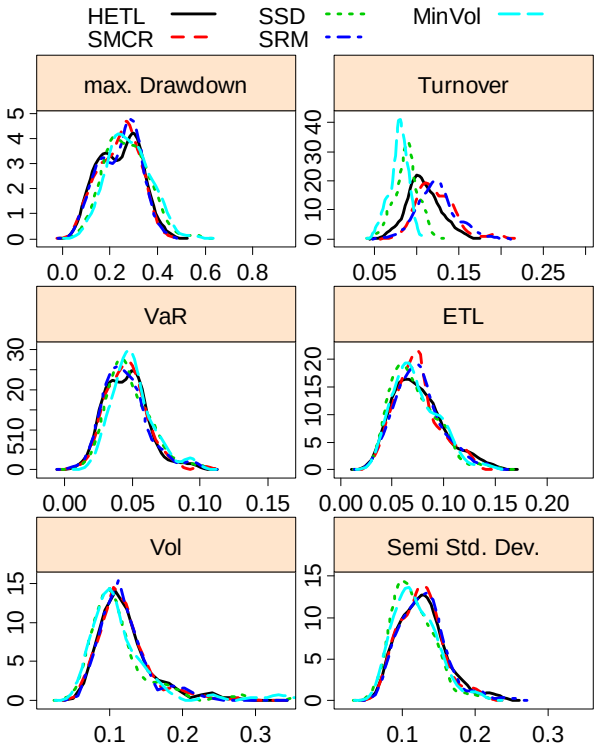

Figure 7.3 Kernel Densities of Risk Measures of Long Only Experiments

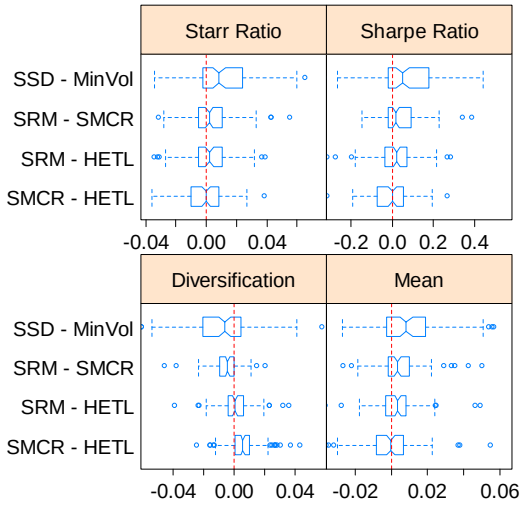

Figure 7.4 Notched Box-plots of Paired Difference for Performance Measures

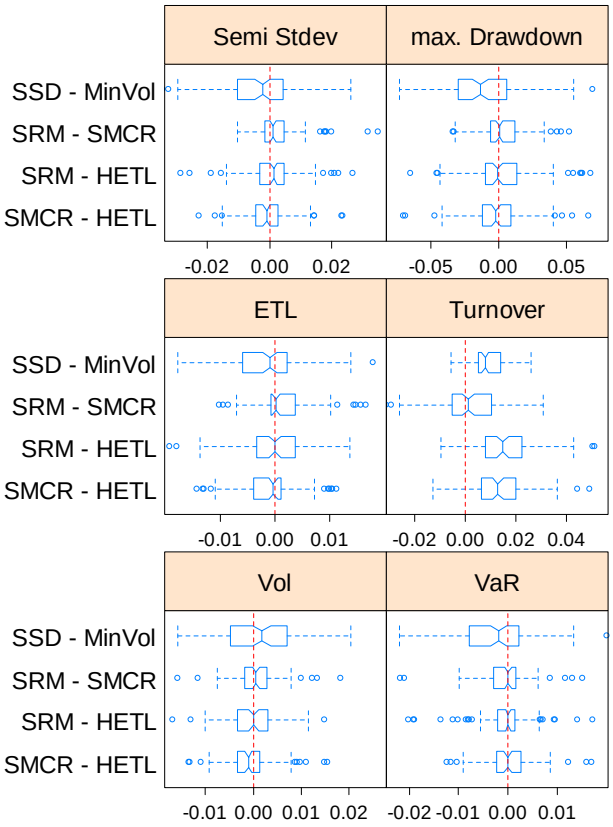

Figure 7.5 Notched Box-plots of Paired Difference for Risk Measures

Combining results from these analyses, we observe the following facts:

1. SSD method clearly outperforms Minimum volatility method in all performance measures except diversification. It also has smaller out-of-sample risks except volatility and turnover rate.

2. HETL, SMCR and SRM are measuring the 95% level tail risk. They perform much better than SSD, although there are little difference among these three methods.

3. Among SRM, SMCR and HETL, SRM performs slightly better. SRM's worst portfolio averages -1.6% per annum, whereas SMCR is -2.4% and HETL is -6.1%. SRM also has the best average annualized return. This is probably because SRM uses the whole distribution. Therefore it catches early upswing signals. SMCR and HETL only use the left tail of the distribution.

4. MinVol and SSD have smaller turnover rates than the HETL, SMCR and SRM group. Among the latter group, HETL has the smallest turnover rate.

In summary, the tail risk group (HETL, SMCR, SRM, SSD) shows better control of risk on most aspects. It is less volatile on both upside and downside. SRM is leading the group slightly in terms of the annualized returns. The risk measures are almost the same for HETL, SMCR and SRM.

The quantile cumulative return plot (Fig. 7.6) and average cumulative return plot (Fig. 7.7) show that all these measures are not affected by the erratic returns that plagued the modified VaR/ETL families in chapter 3 and 4. SRM is slightly in the lead in the recovery period. SMCR and ETL do not have much difference.

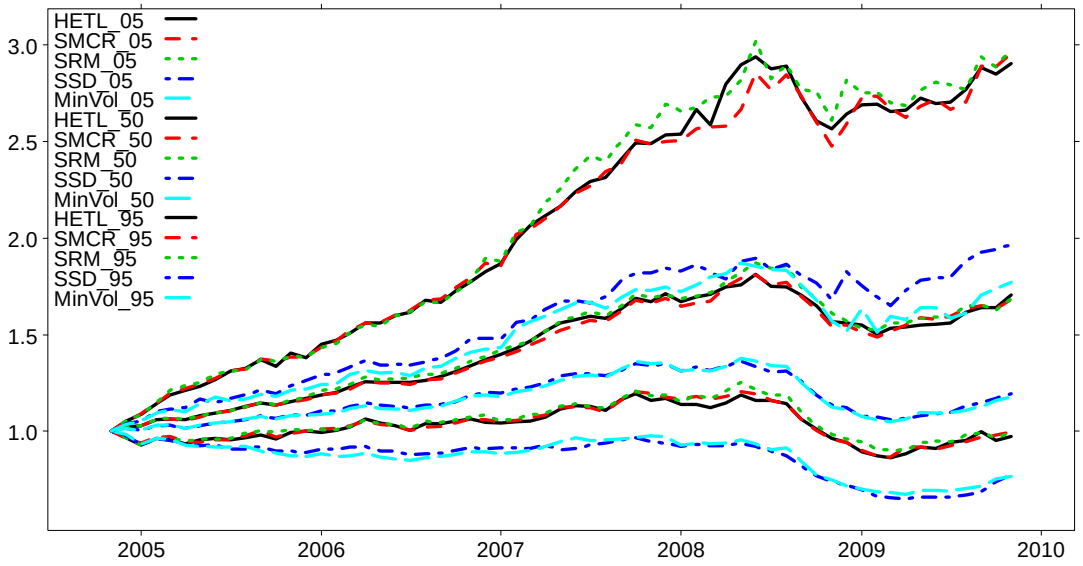

Figure 7.6 Quantile Cumulative Return Plots. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. Confidence levels are set to 95% for HETL, SRCM and SRM.

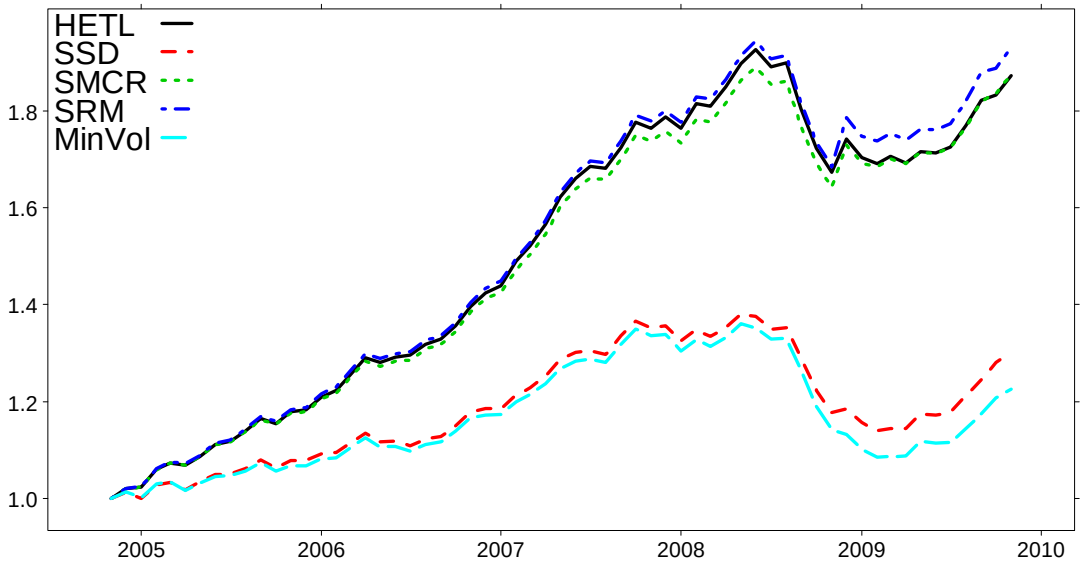

Figure 7.7 Average Cumulative Return Plots. Each curve represents the average of the cumulative returns across all 100 portfolios at the different time. Confidence levels are set to 95% for HETL, SRCM and SRM.

7.2.3 Experiments with 75% Confidence Level

In this section we reduce our tail risk measures confidence level to 75%. In the last experiment, the 95% confidence level is equivalent to considering only 3 points in the tail for the 60-period training data. By reducing the confidence level, we can compare every method with more data under consideration. SSD and MinVol are unaffected by confidence level change, hence not included in this section.

We summarize our results in Fig. 7.8 to Fig. 7.10 using the same pattern as in the last section.

It is clear that SRM is in the lead in terms of average returns. It has smaller risks (ETL, VaR, volatility, max. drawdown) and better performance (returns, Sharpe ratio) than HETL method. It even has smaller turnover rate. It outperforms SMCR in all these measures too, except volatility. At 95% level, essentially all three methods are dealing with the last three points on the left tail. Although SRM considers the whole distribution, the weights attributed to the upside in minimal. No matter how the weights are arranged to those three points, the outcomes of all three methods are not that different. At 75% level, there are 15 points on the left tail under consideration. The different risk measures now show more divergence in their portfolio optimization results. In particular, the SRM method that considers the whole distribution now has a much more flat weight spectrum than it had at the 95% confidence level. This may give a better balance between controlling risk and pursuing better return.

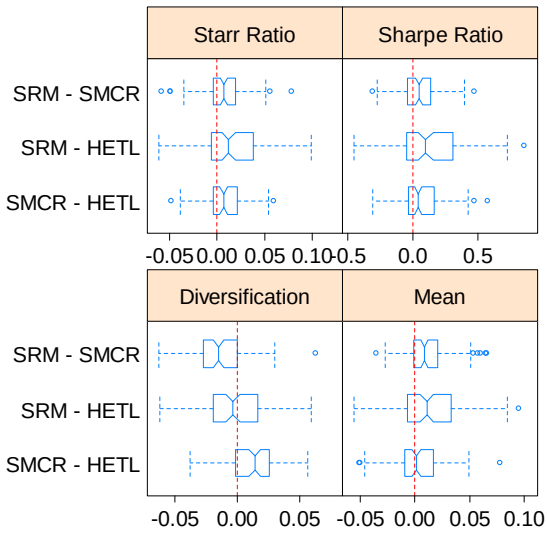

Figure 7.8 Notched Box-plots of Paired Difference for Performance Measures

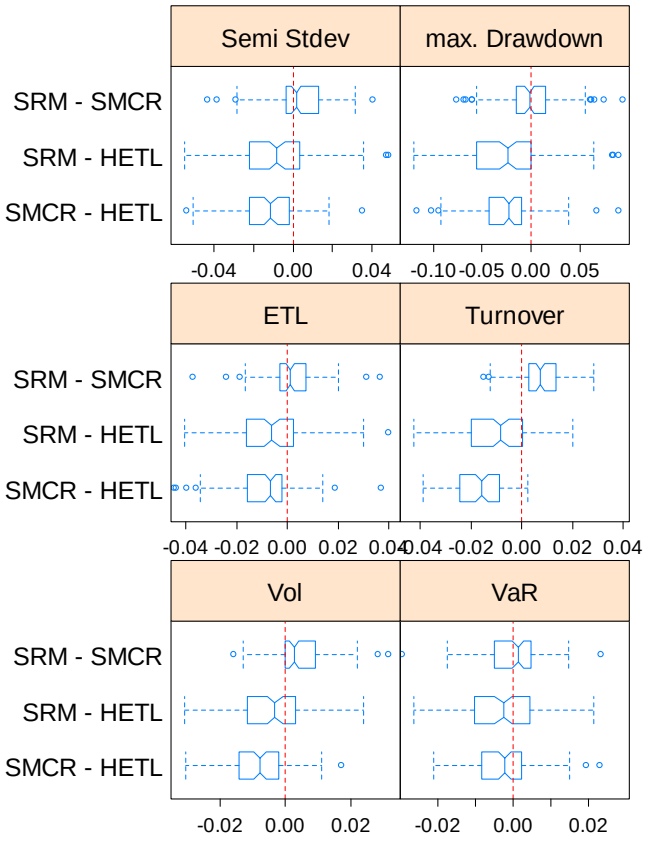

Figure 7.9 Notched Box-plots of Paired Difference for Risk Measures

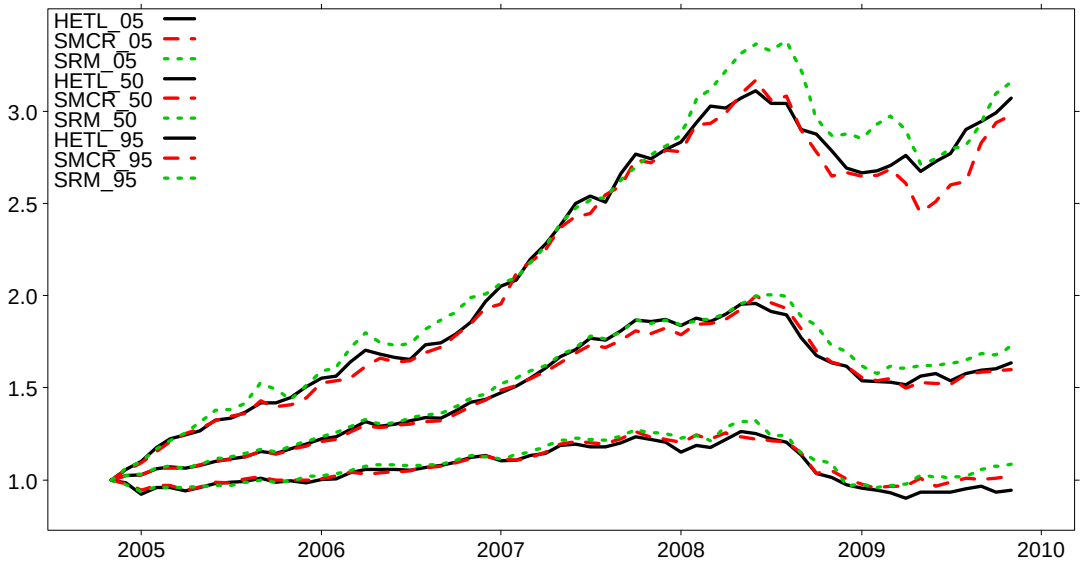

Figure 7.10 Quantile Cumulative Return Plots. Each curve represents the specified quantile of the cumulative returns across all 100 portfolios at the different time. Confidence levels are set to 75% for HETL, SRCM and SRM.

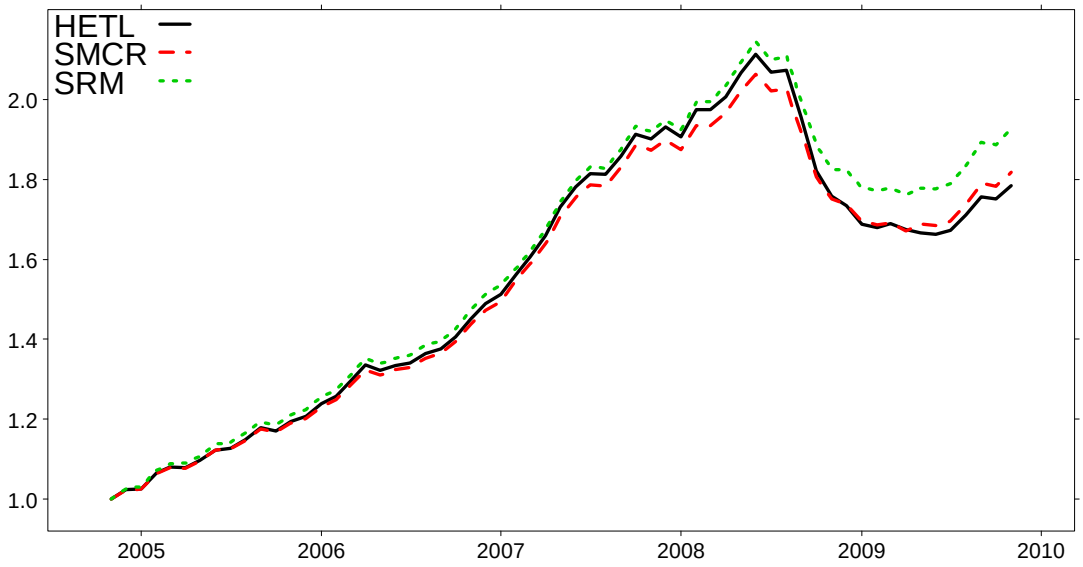

Figure 7.11 Average Cumulative Return Plots. Each curve represents the average of the cumulative returns across all 100 portfolios at the different time. Confidence levels are set to 75% for HETL, SRCM and SRM.

7.3 Conclusions

In this chapter we have compared several tail risk measures in portfolio optimization. They generally assign heavier weights to more severe losses. When the confidence level is set to be very high, i.e. very few data points are used in estimating the tail risks, HETL, SMCR and SRM perform very similarly. When the confidence is lower, SRM is in the lead. It shows better control of tail risks and better returns.

SSD and MinVol is another group of comparison. SSD is clearly better than minimum variance method. This happens when the upside variance is large. SSD does not account that as risk and gives better performance. However, both methods are not as good as the other tail risk measures discussed above. They lag behind in terms of performance by a big margin.

Bibliography

Aas, Kjersti and Hobæk Haff, Ingrid (2006), “The Generalised Hyperbolic Skew Student’s t-distribution”, *Journal of Financial Econometrics*, 4(2).

Acerbi, C. (2002), “Spectral Measures of Risk: a Coherent Representation of Subjective Risk Aversion”, *J. Bank. Finance*, 26(7), 1487–1503.

Acerbi, C. and Tasche, D. (2002), “On the Coherence of Expected Shortfall”, *J. Bank. Finance*, 26(7), 1487–1503.

Arditti, F. (1971), “Another Look at Mutual Fund Performance”, *Journal of Financial and Quantitative Analysis*, 6, 909-12.

Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1999), “Coherent Measures of Risk”, *Math. Finance*, 9(3), 203–228.

Athayde De M and Flôres G. (1997), “Introducing Higher Moments in the CAPM: Some Basic Ideas”, Working Paper.

Athayde G. and Flôres R. (2004), “Finding A Maximum Skewness Portfolio – A General Solution to Three Moments Portfolio Choice”, *Journal of Economic Dynamics and Control*, 7, 1335-1352

Azzalini, A. T., Dal_Cappello and S. Kotz (2003), “Log-skew-normal and Log-skew-t Distributions as Models for Family Income Data”, *J. Income Distribution*, 11, no. 3-4, 12-20.

Bali, T.G and Gokcan, S. (2004), “Alternative Approaches to Estimating VaR for Hedge Fund Portfolios”, Intelligent Hedge Fund Investing: An Introduction, Risk Books.

Barone-Adesi, G. (1985), “Arbitrage Equilibrium with Skewed Asset Returns”, *Journal of Financial and Quantitative Analysis*, 20, 299-313

Barone-Adesi, G., Gagliardini, P. and Urga, G. (2003), “Testing Asset Pricing Models with Coskewness”, *Journal of Business & Economic Statistics*

Berenyi, Z. (2002), “Measuring Hedge Fund Risk with Multi-Moment Risk Measures”, Working Paper.

Bassett, Gilbert W., Koenker, Roger W. and Kordas, G. (2004), “Pessimistic Portfolio Allocation andChoquet Expected Utility”, *Journal of Financial Econometrics*, 2(4), 477-492

Bibby, B.M., and M. Sørensen (2003), “Hyperbolic Processes in Finance”, in S. Rachev (ed.) Handbook of Heavy Tailed Distributions in Finance, Elsevier Science, Amsterdam. 211–248.

Boudt, K., Peterson, B. and Croux, C. (2008), “Estimation and Decomposition of Downside Risk for Portfolios with Non-normal Returns”, *Journal of Risk*, 11(2), 79-103

Bowley, A.L. (1920), “Elements of Statistics”, New York: Charles Scribner's Sons.

Boyer, B., Mitton, T. and Vorkink, K. (2009), “Expected Idiosyncratic Skewness”, *The Review of Financial Studies.*

Brys, G., Hubert, M., and Struyf, A. (2003a), “A Comparison of Some New Measures of Skewness, Developments in Robust Statistics”, ICORS 2001, eds. R. Dutter, P. Filzmoser, U. Gather and P.J. Rousseeuw, Springer-Verlag: Heidelberg, 98–113.

Brys, G., Hubert, M. and Struyf, A. (2004), “A Robust Measure of Skewness”, *J. Comput. Graphical Statist.,* 13 (4), 1-22.

Brys, G., Hubert, M., and Struyf, A. (2006), “Robust Measures of Tail Weight”, *Computational Statistics & Data Analysis,* 50(3), 733-759

Christie-David and R., Chaudry, M. (2001), “Coskewness and Cokurtosis in Futures Markets”, *Journal of Empirical Finance*, 8, 55-81.

Chung, P., Johnson, H. and Schill M. (2006), “Asset Pricing When Returns Are Non Normal: Fama-French Factors Vs Higher Order Systematic Co-Moments”, *Journal of Business*.

Conrad, J., Dittmar, R. and Ghysels, E. (2008), “Ex Ante Skewness and Expected Stock Returns”, Working Paper.

Copeland, T.E., Weston, J.F. (1988), “Financial Theory and Corporate Policy”, Addison-Wesley

Cornish E., Fisher R. (1938), “Moments and Cumulants in the Specification of Distributions”, Revue De L'Institut International De Statistique.

Cornish E., Fisher R. (1960), “the Percentile Points of Distributions Having Known Cumulants”, *Technometrics.*

Cotter A., and Dowd K. (2006), “Extreme spectral risk measures: An application to futures clearinghouse margin requirements”, *Journal of Banking & Finance,* 30, 3469–3485

Crow and Siddiqui (1967), “Robust Estimation of location”, *J. Amer. Statist. Assoc.*, 62, 353-389.

Dhaene J., Vanduffel S., and Goovaerts M. (2007), “Comonotonicity”

Dittmar, R. (2002), “Nonlinear Pricing Kernels, Kurtosis Preference and Evidence from the Cross-Section of Equity Returns”, *Journal of Finance*.

Dolmas, Jim (2005). “Trimmed Mean PCE Inflation”, *Research Department Working Paper 0506.* Federal Reserve Bank of Dallas.

Fama, E. and French, E. (1995), “Size and Book to Market Factors in Earnings and Returns”, *Journal of Finance*, 50, 131–155.

Fama, E. and French, E. (1996), “Multifactor Explanations of Asset Pricing Anomalies”, *Journal of Finance,* 51, 55–84.

Fang, H. and Lai, T. (1997), “Co-Kurtosis and Capital Asset Pricing”, *Financial Review*, 32, 293–307.

Favre, L. and Galeano, J.A. (2002), “Mean-Modified Value-At-Risk Optimization with Hedge Funds”, *Journal of Alternative Investments*, 21-25.

Fernandez, C. and Steel, M. F. J. (1998), “On Bayesian Modeling of Fat Tails and Skewness”, *J. Am. Statist. Assoc.* 93, 359–371.

Fishburn, P.C. (1977), “Mean-risk Analysis with Risk Associated with Below Target Returns”, *The American Economic Review,* 67, 116–126

Francis, J. (1975), “Skewness and Investors’ Decisions”, *Journal of Financial and Quantitative Analysis*, 10, 163-72.

Friend, I. and Westerfield, R. (1980), “Co-Skewness and Capital Asset Pricing”, *Journal of Finance*, 30, 897–913.

Füss, R., Kaiser, D.G., Adams, Z. (2007), “Value at Risk, Garch Modeling and the Forecasting of Hedge Fund Return Volatility”, *Journal of Derivatives and Hedge Funds*, 13, 2-25.

Galagedera, D., Henry, D. and Silvapulle, P. (2004), “Empirical Evidence on the Conditional Relation between Higher-order Systematic Co-Moments and Security Returns”, *Quarterly Journal of Business and Economics*, 42, 121–137.

Gamba, A. and Rossi, F. (1998), “Mean-Variance-Skewness Analysis in Portfolio Choice and Capital Markets”, *Ricerca Operativa*, 28, 5-46.

Giorgi E.D., (2005), “Reward–risk Portfolio Selection and Stochastic Dominance”, *Journal of Banking & Finance*, 29(4), 895-926

Groeneveld, R.A. and Meeden, G. (1984), "Measuring Skewness and Kurtosis," *The Statistician*, 33, 391-399.

Groeneveld, R.A. (1991), “An Influence Function Approach to Describing the Skewness of a Distribution,” *The American Statistician*, 45, 97–102.

Hampel, F.R. (1968), “Contribution to the Theory of Robust Estimation”, Ph.D. thesis, University of California, Berkeley.

Hampel, F.R. (1974), “The Influence Curve and Its Role in Robust Estimation”, *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), “Robust Statistics: the Approach Based on Influence Functions”, Wiley, New York.

Hansen, L. (1982), “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica*, 50, 1029–1054.

Hanoch, G and H. Levy (1969), "The Efficiency Analysis of Choices Involving Risk", *The Review of Economic Studies*, 36(3), 335-346

Harvey, C. (2000), “The Drivers of Expected Returns in International Markets”, *Emerging Markets Quarterly*, 17.

Harvey, C., Liechty, J., Liechty, M. and Mueller, P. (2004), “Portfolio Selection with Higher Moments”, *Social Science Research Network Working Paper Series.*

Harvey, C., Siddique, A. (2000), “Conditional Skewness in Asset Pricing Tests”, *Journal of Finance*, 3, 1263-1295.

Hinkley, D.V. (1975), “On Power Transformations to Symmetry”, *Biometrika*, 62, 101–111.

Hogg, R.V. (1974), "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory", *Journal of the American Statistical Association*, 69, 909-923.

Hyndman and Fan (1996), "Sample Quantiles in Statistical Packages", *The American Statistician*, 50(4):361-365

Huber, P. (1981), Robust Statistics, Wiley, New York.

Hwang, S. and Satchell, S. (1999), “Modeling Emerging Market Risk Premia Using Higher Moments”, *International Journal of Finance & Economics*, 4(4), 271 - 296.

Ingersoll, J. (1975), “Multidimensional Security Pricing”, *Journal of Financial and Quantitative Analysis*, 10, 785-798.

Jaschke, S.R. (2001), “The Cornish-Fisher-Expansion in the Context of Delta-Gamma-Normal Approximations”, *Journal of Risk*, 4(4), 33-52

Jones, M. C. and Faddy, M. J. (2003), “A Skew Extension of the t-distribution with Applications”, *J. Roy. Statist. Soc B*, 65, 159-174.

Jorion, P. (1997), “Value at Risk: the New Benchmark For Controlling Market Risk”, Mcgraw-Hill.

Jurcenzko, E. and Maillet, B. (2002), “The Four-Moment Capital Asset Pricing Model: Some Basic Results”, Edhec Publications.

Kim T., White H. (2003), “On More Robust Estimation of Skewness and Kurtosis: Simulation and Application to the S&P500 Index”, *Finance Research Letters*, 1, 56-73.

Kendall, M.G. and Stuart, A. (1977), “The Advanced Theory of Statistics”, Vol. 1, London: Charles Griffin and Co.

Kooli, M., Serge, P.A. and Gueyie, J.P. (2005), “Hedge Funds in a Portfolio Context: A Mean–Modified Value at Risk Framework, Derivatives Use”, *Trading and Regulation,* 10, 373-383.

Kraus, A. and Litzenberger, R. (1976), “Skewness Preference and the Valuation of Risky Assets”, *Journal of Finance*, 31, 1085-1100.

Krokhmal, P.A. (2007), “Higher Moment Coherent Risk Measures”, *Quantitative Finance*, 7(4)

Lamm, Jr., and Mcfall, R. (2003), “Asymmetric Returns and Optimal Hedge Fund Portfolios”, *Journal of Alternative Investments*, 9-21.

Levy, H. (1998), “Stochastic Dominance: Investment Decision Making under Uncertainty”, Kluwer Academic Publishers

Lim, K. (1989), “A New Test of the Three-Moment Capital Asset Pricing Model”, *Journal of Financial and Quantitative Analysis*, 24, 205–216.

Lin, B. and Wang, J. (2003), “Systematic Skewness in Asset Pricing: An Empirical Examination of the Taiwan Stock Market”, *Applied Economics*, 35, 1877 – 1887.

Lintner, J. (1965), “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets”, *Review of Economics and Statistics*, 47, 13-37.

Malevergne, Y. and D. Sornette (2005a), “High-Order Moments and Cumulants of Multivariate Weibull Asset Returns Distributions: Analytical Theory and Empirical Tests: II”, *Finance Letters*, 3(1), 54-63

Malevergne,Y. and Sornette, D. (2005b), “Higher-Moment Portfolio Theory: Capitalizing on Behavioral Anomalies of Stock Markets”, *The Journal of Portfolio Management*, 31, 49-55.

Markowitz, H. (1952), “Portfolio Selection”, *Journal of Finance.*

Maronna, R. A., Martin, R. D. and Yohai, V. J. 2006. Robust Statistics: Theory and Methods. Wiley, London.

Martellini, L., Vaissié M. and Ziemann V. (2005), “Investing in Hedge Funds: Adding Value Through Active Style Allocation Decisions”, *Edhec Risk and Asset Management Research Centre Publication*

Martellini, L. and Ziemann, V. (2008), “Improved Forecasts of Higher-Order Co-Moments and Implications For Portfolio Selection”, *Edhec Publications.*

Mina, J. and Ulmer, A. (1999), “Delta-Gamma Four Ways”, Riskmetrics Group.

Moors, J. J. A. (1988), “A Quantile Alternative for Kurtosis”, t*he Statistician*, 37, 25-32.

Newey, W.K. and West, K.D. (1987), “A Simple Positive-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimator”, *Econometrica*, 55, 703-708.

Ogryczak, W. and Ruszczynski, A. (1999), “From Stochastic Dominance to Mean-Risk Models: Semideviations as Risk Measures”, *European Journal of Operational Research*, Elsevier, 116(1), 33-50

Peiro, A. (2003), “Skewness in Financial Returns”, *Journal of Banking & Finance*, 23, 847-862

Pflug, G. (2000), “Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk”, *Probabilistic Constrained Optimization: Methodology and Applications*, edited by S. Uryasev

Piessens R., De Doncker-Kapenga E. and Überhuber, (1983) “C.. QUADPACK: a Subroutine Package for Automatic Integration”, Springer.

Post, T., Vliet, P. and Levy, H. (2008), “Risk Aversion and Skewness Preference”, *Journal of Banking & Finance*, 32, 1178–1187

Ranald, A. and Favre. L. (2003), “How to Price Hedge Funds from Two- to Four-Moment CAPM”, *Edhec Publications.*

Rachev, S.T. and S. Mittnik (2000), “Stable Paretian Models in Finance”, *Series in Financial Economics and Quantitative Analysis*, John Wiley.

Rachev, S.T. (2003 Editor), “Handbook of Heavy Tailed Distributions in Finance”, North Holland Handbooks of Finance, Elsevier.

Rachev, S.T., C. Menn, and F. Fabozzi (2005), “Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management”, *Portfolio selection, and Option Pricing*, John Wiley.

Rachev, S., D. Martin, B. Racheva and S. Stoyanov (2007), “Stable ETL Optimal Portfolios and Extreme Risk Management”, *Risk Assessment: Decisions in Banking and Finance*, Springer.

Rachev S. T, Racheva B., Stoyanov S. and Fabozzi F. J. (2007) ‘Risk Management and Portfolio Optimization for Volatile Markets”, *Technical Report*

Rachev S. T, Stoyanov S. and Fabozzi F. J. (2008) “Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization”, John Wiley.

Rockafellar, R.T. and Uryasev, S. (2000), “Optimization of Conditional Value-at-Risk”. *J. Risk*, 2, 21–41.

Rockafellar, R.T. and Uryasev, S. (2002), “Conditional Value-at-Risk for General Loss Distributions”, *J. Bank. Finance*, 26(7) 1443–1471.

Ronchetti, E. and Trojani, F. (2000), “Robust Inference with GMM Estimators”, *Journal of Econometrics*, 101, 37-69.

Rothschild, M and J. Stiglitz (1970), "Increasing risk. I. A Definition", *Journal of Economic Theory*, 2 ,225-243

Samuelson, P. A. (1970), “A Fundamental Approximation Theory of Portfolio Analysis in Terms of Means, Variance, and Higher Moments”, *Review of Economic Studies*, 37, 537-542.

Sears, S. and Wei, J. (1985), “Asset Pricing, Higher Moments and the Market Risk Premium: A Note”, *Journal of Finance*, 40, 1251-1253.

Sears, S., Wei, J. (1988), “The Structure of Skewness Preferences in Asset Pricing Models with Higher Moments: An Empirical Test”, *Financial Review*, 23, 25-38.

Seier, E., and D. Bonett (2003), “Two Families of Kurtosis Measures”, Metrika, 58(1), 59–70.

Sharpe, W. (1964), “Capital Asset Prices: A theory of Market Equilibrium Under. Conditions of Risk”, Journal of Finance, 19, 425-442.

Sklar, A. (1996), “Random Variables, Distribution Functions, and Copulas - a Personal Look Backward and Forward published in Distributions with Fixed Marginals and Related Topics”, edited by L. Rüschendorf, B. Schweizer, and M.D. Taylor, Institute of Mathematical Statistics, Hayward, CA, 1-14.

Smith, D. (2007), “Conditional Coskewness and Asset Pricing”, Journal of Empirical Finance, 14, 91-119.

Spellucci, P. (1998), “An SQP Method for General Nonlinear Programs Using only Equality Constrained Subproblems”, Mathematical Programming, (82)3, 4

Appendix A

Sequential Quadratic Programming

The basic Sequential Quadratic Programming (SQP) uses the following algorithm. Suppose a nonlinear programming problem is as follows

$$\begin{aligned} \min. \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0 \end{aligned}$$

The Lagrangian for the problem is

$$L(x,\lambda) = f(x) - \lambda g(x)$$

SQR uses the quadratic model to approximate this objective function. Given a current estimate x_k and a search direction δ, the original problem becomes

$$\begin{aligned} \min. \quad & L(x_k,\lambda_k) + \nabla L(x_k,\lambda_k)^T \delta + \frac{1}{2}\delta^T \nabla_{xx} L(x_k,\lambda_k)\delta \\ \text{s.t.} \quad & \nabla g(x_k)^T \delta + g(x_k) = 0 \end{aligned}$$

By solving this quadratic programming problem, we will get the improved x_k, λ_k and use them for the next iteration.

Vita

Minfeng Zhu was born in Changzhou, China. He currently lives in Seattle, WA. He earned his Bachelor's degree in Biochemistry from Nanjing University and a Master of Science in Statistics from the University of Washington. In 2010, he graduated with a Doctor of Philosophy in Statistics also from the University of Washington.