

Expected Shortfall Semi-Scale T-Distribution M-Estimator

R. Douglas Martin¹, Shengyu Zhang²

¹Department of Applied Mathematics and Statistics, University of Washington, Seattle, USA

²BECU, Tukwila, Washington, USA

Email: doug@amath.washington.edu, syzhang0401@gmail.com

How to cite this paper: Martin, R.D. and Zhang, S.Y. (2023) Expected Shortfall Semi-Scale T-Distribution M-Estimator. *Journal of Mathematical Finance*, 13, 483-500. <https://doi.org/10.4236/jmf.2023.134029>

Received: September 13, 2023

Accepted: November 27, 2023

Published: November 30, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The influence function of parametric t-distribution expected shortfall (ES) estimators has an approximately symmetric shape, for which large positive returns indicate large losses. We avoid this risk estimator's unacceptable feature by introducing an ES semi-scale M-estimator for t-distributions, for which the usual t-distribution scale parameter is replaced by a semi-scale parameter. We derive the influence function of the ES semi-scale M-estimator, and show that its influence function has large values only for large negative returns as one expects, and only very small typically negative values for positive returns. The computation of an ES semi-scale M-estimator is shown to be a simple modification of a parametric t-distribution ES maximum-likelihood estimator (MLE), in which the scale MLE is replaced by a semi-scale estimator. We also derive the asymptotic variance expression for the ES semi-scale M-estimator, and show that its standard error is not very much larger than that of the t-distribution ES maximum-likelihood estimator.

Keywords

Risk, Expected Shortfall, Semi-Scale, MLE, M-Estimator, Influence Function

1. Introduction

The early development of portfolio risk management focused on the Value-at-Risk (VaR) measure of risk, whose popularity owes a lot to JP Morgan Risk Metrics document (1996) [1]. Subsequently, a set of risk measure coherence axioms was introduced by Artzner *et al.* (1999) [2], who showed that VaR did not satisfy these axioms, and that the expected shortfall (ES) measure of the average losses beyond the VaR is a coherent risk measure. See also, Sections 2.3 and 8.1 of McNeil *et al.* (2015) [3]. Over the years, ES has become an increasingly popular

risk measure of choice, in part due to its properties for portfolio optimization established by Rockafellar and Uryasev (2000) [4], who used the term CVaR in place of ES, and due to its support by Basel regulatory guidelines for use by banks.

There exist two basic variants of ES estimators, nonparametric ES estimators which do not depend on returns distribution assumptions, and parametric ES estimators which depend on an assumed returns distribution. Nonparametric ES estimators are the average of a small fraction of the smallest ordered portfolio returns, and the ES risk measure that gives rise to such an estimator is known to satisfy the risk coherence axioms. The parametric normal distribution ES risk measure is a linear combination of the distribution mean and the distribution standard deviation, and this linear combination fails to be a coherent risk measure because it fails to satisfy the coherence monotonicity axiom, which states that if random return R_1 is greater or equal to random return R_2 , then the risk for R_1 is less or equal to the risk for R_2 . In terms of the ES risk estimator obtained by replacing the distribution mean and standard deviation with sample estimates, this failure manifests itself in the fact that one large return can give rise to an increase in ES risk, which is quite non-intuitive. For a reasonable risk measure, increasing returns should intuitively result in decreasing risk. Fischer (2003) [5] showed that a simple fix to the lack of coherence of parametric normal distribution ES is to replace the standard deviation by the lower standard semi-deviation, “semi-deviation” for short.

The fact that returns are often non-normal led to a literature focus on parametric ES for fat-tailed non-normal distributions, the simplest of which are t-distributions. Furthermore, parametric ES estimators based on maximum-likelihood parameter estimators (MLE’s) have the attractive feature that they achieve the minimum large-sample variance when returns in fact have the assumed non-normal distribution. With this in mind, Martin and Zhang (2019) [6], henceforth MZ2019 [6], studied the comparative behavior of both parametric normal and parametric t-distribution ES risk measures and estimators, in terms of the behavior of their *influence functions* which are extensively treated in the robust statistics literature. It turns out that the influence functions of both parametric normal and parametric t-distribution based on parameter MLE’s are approximately symmetric functions of a return, which reflects the failure of these risk measures to satisfy the coherence monotonicity axiom.

In MZ2019 [6] it is shown that the monotonicity of a risk measure influence function is a sufficient condition to ensure that the risk measure satisfies the coherence monotonicity axiom. Thus, our goal was to modify the parametric t-distribution ES in such a way that its influence function is monotonic. Motivated by the Fischer (2003) [5] result that replacement of the standard deviation by semi-deviation results in a coherent mean semi-deviation risk measure, we will show that replacement of the t-distribution MLE scale estimator by a semi-scale estimator results in ES influence functions that are quite close to being monotonic,

the more so for t-distribution tail probabilities that are larger than commonly recommended values.

The remainder of the paper is organized as follows. Section 2 briefly reviews nonparametric and normal distribution expected shortfall, and their influence functions. Section 3 introduces general parametric risk measures and influence functions, and discusses the t-distribution special case. Section 4 introduces a new t-distribution ES Semi-Scale M-estimator, derives its influence function formula, and displays its typical influence function shape. Section 5 describes a practical implementation of the ES Semi-Scale M-estimator. Section 6 derives the asymptotic variance formula for the ES Semi-Scale M-estimator and shows that, as a function of the ES tail probability, its asymptotic standard errors are not very much greater than those of a t-distribution ES maximum likelihood estimator. Section 7 contains concluding comments. The detailed derivation of some mathematical results may be found in the Appendix of an early draft version of this paper, which is available at SSRN (<https://ssrn.com/abstract=4605604>). An early version of MZ2019 [6], which includes its Appendices, may be found at SSRN (<https://ssrn.com/abstract=2747179>).

2. Nonparametric and Normal Distribution Parametric Expected Shortfall and Influence Functions

This section reviews some basic material about nonparametric and normal distribution parametric shortfall, and their influence functions, which is discussed in more detail in MZ2019 [6].

2.1. Nonparametric and Normal Distribution Parametric Expected Shortfall Formulas

The integral form of expected shortfall with risk as a positive quantity is

$$ES_{\gamma}(F) = -\frac{1}{\gamma} \int_{r \leq q_{\gamma}(F)} r \cdot dF(r) \quad (2.1)$$

where F is an unrestricted returns distribution function and $q_{\gamma}(F)$ is the tail probability γ quantile functional. A nonparametric estimator of ES is obtained by replacing the unknown distribution function F in the $ES_{\gamma}(F)$ formula by the empirical distribution function $F_n(r)$ which has a jump of size n^{-1} at each of the return values r_1, r_2, \dots, r_n . For purposes of providing a formula for the ES estimator, we let $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(n)}$ be the ordered values of the observed returns, and let $\lceil x \rceil$ be the smallest integer greater or equal to x . Then the nonparametric estimator \widehat{ES}_{γ} formula is:

$$\widehat{ES}_{\gamma} = -\frac{1}{\lceil n\gamma \rceil} \sum_{i=1}^{\lceil n\gamma \rceil} r_{(i)} \quad (2.2)$$

In typical risk management applications, the choice of γ will be 0.01, 0.025 or 0.05, *i.e.*, 1%, 2.5% or 5%. For such applications, the ordered returns in the above summation are that small fraction of all returns which have the most negative single period returns, thereby resulting in a positive ES estimator. The

greater such losses are, the greater will be the associated positive ES estimator.

A parametric ES risk measure is obtained by replacing F with a parametric distribution function F_{θ} where θ is a vector of the distribution parameters. In the case of a normal distribution, the parameters are μ and σ , and straightforward evaluation of the integral (2.1) results in the *normal distribution parametric* ES formula

$$ES_{\gamma}(\mu, \sigma) = -\mu + \frac{\phi(z_{\gamma})}{\gamma} \cdot \sigma \quad (2.3)$$

where ϕ is the standard normal density function, and z_{γ} is the standard normal distribution γ quantile¹. In this case the parametric ES estimator is:

$$ES_{\gamma}(\hat{\mu}, \hat{\sigma}) = -\hat{\mu} + \frac{\phi(z_{\gamma})}{\gamma} \cdot \hat{\sigma} \quad (2.4)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the normal distribution sample mean and sample standard deviation maximum likelihood estimators.

Likewise, for a t-distribution with parameters $\theta = (\mu, s, \nu)$, evaluation of the integral (2.1) results in the parametric t-distribution ES formula:

$$ES_{\gamma}(\mu, s, \nu) = -\mu + \frac{g_{\gamma, \nu}}{\gamma} \cdot s \quad (2.5)$$

where

$$g_{\gamma, \nu} \triangleq \sqrt{\frac{\nu}{\nu-2}} \cdot f_{\nu-2} \left(\sqrt{\frac{\nu-2}{\nu}} \cdot q_{\gamma, \nu} \right) \quad (2.6)$$

with $f_{\nu-2}$ the standard t-density with $\nu-2$ degrees of freedom, and $q_{\gamma, \nu}$ the tail probability γ quantile for the standard t-distribution with ν degrees of freedom². The parametric t-distribution ES estimator is obtained by replacing the parameters μ, s, ν with their t-distribution maximum-likelihood estimates.

The above nonparametric ES estimator is fundamentally different in character than the above parametric ES estimators, in that the former depends only on a small fraction of the most negative ordered returns, whereas the latter depends on all the returns for their parameter estimates. Furthermore, nonparametric ES is known to be a coherent risk measure, whereas parametric ES is not a coherent risk measure. In particular, parametric ES fails to satisfy the coherence monotonicity axiom. In data-oriented estimator terms, monotonicity means that if any return data value is decreased the risk should always increase or remain constant, and conversely. For the nonparametric risk estimator (2.2), this condition clearly holds since the decrease of any return among the lowest γ percent of the returns results in an increase in risk, and conversely, the decrease or increase of any return in the largest $1-\gamma$ percent of the returns results in no change in the risk. On the other hand, monotonicity does not hold for the parametric normal distribution ES because while the decrease of a return value below the mean will

¹See for example MZ2019 [6] or Jorion (2007) [7].

²Zhang (2016) [9] shows that the expression (2.5) is equivalent to the one in McNeil *et al.* (2015) [3].

decrease both terms in Equation (2.3), a decrease of a large return above the mean can decrease the second term while increasing the first term, in such a way that the risk decreases.

2.2. Nonparametric and Parametric Normal Distribution ES Influence Functions

The influence function was introduced by Hampel (1974) [8] as a basic tool for the study of robust parameter estimators³. As we shall see, the shape of a risk measure influence function provides an immediate intuitive understanding of the impact that gains and losses, positive and negative returns respectively, have on a parametric risk estimator. The influence function is defined using a mixture distribution of the form

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

where $\delta_r(x)$ is a point mass probability distribution located at r . The influence function of nonparametric $ES_\gamma(F)$ is defined as the following directional (Gateaux) derivative of the ES functional at $F(x)$ in the direction of the point mass distribution $\delta_r(x)$:

$$IF_{ES_\gamma}(r; F) = \lim_{p \rightarrow 0} \frac{ES_\gamma(F_p) - ES_\gamma(F)}{p} = \left. \frac{d}{dp} ES_\gamma(F_p) \right|_{p=0}.$$

As such, it has the intuitive interpretation as the asymptotic form of the difference quotient for an ES estimator evaluated at a fixed set of n returns r_1, r_2, \dots, r_n and at an augmented set of $n + 1$ returns r_1, r_2, \dots, r_n, r with r variable, divided by n^{-1} .

Using the integral representation of $ES_\gamma(F)$ in equation (2.1) it is shown in MZ2019 [6] that the nonparametric influence function of ES is

$$IF_{ES_\gamma}(r; F) = -q_\gamma(F) - ES_\gamma(F) - \frac{I_{(-\infty, q_\gamma(F)]}(r)}{\gamma} (r - q_\gamma(F)) \tag{2.7}$$

where $I_{(-\infty, q_\gamma(F)]}(r)$ is the indicator function of the interval $(-\infty, q_\gamma(F)]$. The above influence function expression can be evaluated for any distribution F , and the left-hand plot in **Figure 1** displays the nonparametric ES influence function for the case where $\gamma = 5\%$, with $F(x)$ a normal distribution with mean 0.12 and standard deviation 0.24. The resulting influence function has a striking piece-wise linear monotonic decreasing shape.

As for the influence function of the parametric normal distribution ES given by Equation (2.3), we first note that this ES functional representation $ES_\gamma(\mu(F), \sigma(F))$ is the following linear combination of the mean functional $\mu(F)$ and the standard deviation functional $\sigma(F)$:

$$ES_\gamma(\mu(F), \sigma(F)) = -\mu(F) + \frac{\phi(z_\gamma)}{\gamma} \cdot \sigma(F).$$

³See Hampel *et al.* (1986) [10] for detailed discussions of influence functions, and their applications.

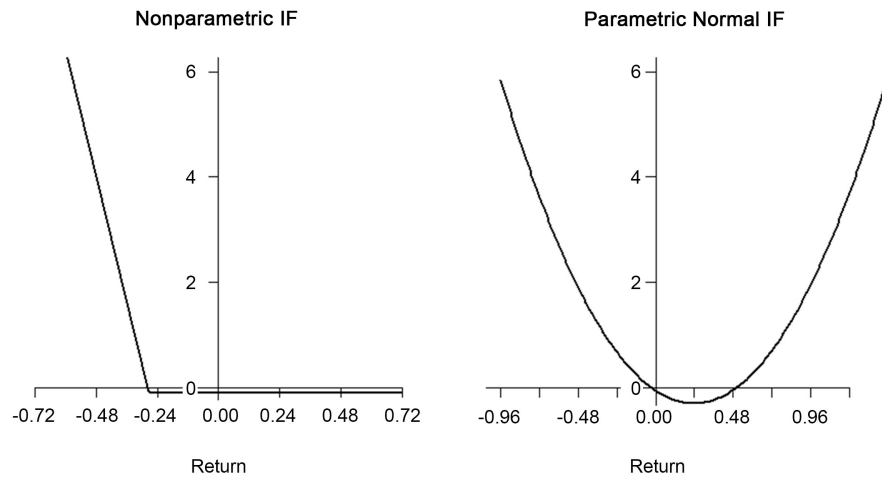


Figure 1. Nonparametric and parametric tail probability $\gamma = 5\%$ ES influence function for a normal distribution with mean 0.12 and standard deviation 0.24.

It follows that the influence function of parametric normal distribution ES is the corresponding linear combination of the influence functions of $\mu(F)$ and $\sigma(F)$. It is well-known in the robust statistics literature that influence function of the mean is

$$IF_{\mu(F)}(r) = r - \mu$$

and the influence function of the standard deviation is⁴:

$$IF_{\sigma(F)}(r) = \frac{(r - \mu)^2 - \sigma^2}{2\sigma}.$$

Thus, the influence function of parametric normal distribution ES is

$$IF_{ES,\gamma}(r; \mu, \sigma) = -(r - \mu) + \frac{\phi(z_\gamma)}{\gamma} \cdot \frac{(r - \mu)^2 - \sigma^2}{2\sigma} \tag{2.8}$$

and it is displayed in the right-hand plot of **Figure 1**, with the same mean and standard deviation values as for the nonparametric ES influence function in the left-hand plot.

The shape of the parametric ES influence function is symmetric about the value $\mu + \gamma\sigma/\phi(z_\gamma) = 0.24$, and in this regard is strikingly different from the desirable monotonic character of the nonparametric ES influence function, which is either constant, or strictly increasing, for decreasing return r , beyond the tail probability $\gamma = 0.05$ quantile. On the other hand, the parametric normal distribution influence function is quite non-monotonic, with increasing return values much larger than 0.24 indicating increased risk, which is quite unnatural.

3. Parametric Risk Measures and Influence Functions

This section first discusses general parametric risk measures and their general influence function formula, as well as a specific influence function formula for the

⁴See for example Section 3 of Zhang, Martin and Christidis (2021) [11].

case where maximum-likelihood parameter estimators are used. Then it presents the MZ2019 [6] influence function formula for parametric t-distribution expected shortfall.

3.1. General Parametric Risk Measures and Influence Functions

Let $\rho(\boldsymbol{\theta}) = \rho(F_{\boldsymbol{\theta}})$ be a parametric risk measure defined by a fixed parametric family of univariate distribution functions $F_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. It will suffice for our applications in this paper to assume that $F_{\boldsymbol{\theta}}(r)$ is continuous and strictly increasing with quantile function $q_{\gamma}(F_{\boldsymbol{\theta}}) = F_{\boldsymbol{\theta}}^{-1}(\gamma)$. One obtains a parametric risk measure estimator by first choosing a parameter estimator $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ and then using it to obtain a risk estimator $\hat{\rho}_n = \rho(\hat{\boldsymbol{\theta}})$, where we have suppressed the subscript n on $\hat{\boldsymbol{\theta}}_n$ for notational convenience. We assume that $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$ is based on independent and identically (i.i.d.) returns r_1, r_2, \dots, r_n , with a general distribution function F that will ultimately be equal to a parametric family $F_{\boldsymbol{\theta}}$ of t-distributions for our parametric ES maximum-likelihood estimator.

An estimator $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$ is assumed to be obtained from a functional $\boldsymbol{\theta}(F)$ by the plug-in-rule

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}(F_n) \quad (3.1)$$

where F_n is the empirical distribution function of a set of returns r_1, r_2, \dots, r_n . Correspondingly, we represent the risk measure functional as

$$\rho(\boldsymbol{\theta}) = \rho(\boldsymbol{\theta}(F)) \quad (3.2)$$

and the risk measure estimator is

$$\hat{\rho}_n = \rho(\boldsymbol{\theta}(F_n)). \quad (3.3)$$

We tacitly assume that the returns are independent and identically distributed (i.i.d.) with a parametric distribution $F_{\boldsymbol{\theta}}$. Then when the usual Fisher consistency condition $\boldsymbol{\theta}(F_{\boldsymbol{\theta}}) = \boldsymbol{\theta}$ holds, the estimator $\hat{\boldsymbol{\theta}}_n$ will under reasonable conditions converge in probability to the true parametric distribution parameter $\boldsymbol{\theta}$. Correspondingly, the risk measure estimator $\hat{\rho}_n = \rho(\boldsymbol{\theta}(F_n))$ will converge in probability to $\rho(\boldsymbol{\theta})$. Note however that, when the returns come from $F \neq F_{\boldsymbol{\theta}}$ then $\boldsymbol{\theta}(F_n)$ will converge in probability to $\boldsymbol{\theta}(F)$, but typically $\boldsymbol{\theta}(F) \neq \boldsymbol{\theta}$.

For deriving parametric risk measure influence functions, the distribution function F is replaced by the parametric distribution function $F_{\boldsymbol{\theta}}$, which results in the parametric mixture distribution

$$F_{\boldsymbol{\theta},p} = (1-p)F_{\boldsymbol{\theta}} + p \cdot \delta_r, \quad 0 \leq p < 1. \quad (3.4)$$

Then, referring to the risk measure loosely by its estimator $\hat{\rho}$, the influence function of $\hat{\rho}$ is defined as

$$IF_{\hat{\rho}}(r; F_{\boldsymbol{\theta}}) = \lim_{p \downarrow 0} \frac{\rho(\boldsymbol{\theta}(F_{\boldsymbol{\theta},p})) - \rho(\boldsymbol{\theta}(F_{\boldsymbol{\theta}}))}{p} = \left. \frac{d\rho(\boldsymbol{\theta}(F_{\boldsymbol{\theta},p}))}{dp} \right|_{p=0}. \quad (3.5)$$

The parameter estimator vector influence function $IF_{\hat{\theta}}(r) = IF_{\hat{\theta}}(r; F)$ has the components

$$IF_{\hat{\theta}_k}(r; F_{\theta}) = \left[\frac{\partial \theta_k(F_{\theta, p})}{\partial p} \right]_{p=0} \tag{3.6}$$

With $\nabla \rho(\theta) = (\rho_1(\theta), \dots, \rho_K(\theta))'$ the gradient of $\rho(\theta)$, where $\rho_k(\theta) = \partial \rho(\theta) / \partial \theta_k$, the chain rule gives:

$$\begin{aligned} IF_{\hat{\rho}}(r; F) &= \left[\nabla \rho(\theta(F_p))' \cdot \frac{\partial \theta(F_p)}{\partial p} \right]_{p=0} \\ &= \nabla \rho(\theta(F_p))' \cdot IF_{\hat{\theta}}(r) \\ &= \sum_{k=1}^K \rho_k(\theta) \cdot IF_{\hat{\theta}_k}(r) \end{aligned} \tag{3.7}$$

The above is a general expression valid for any consistent parameter estimator $\hat{\theta}$ having an influence function $IF_{\hat{\theta}}(r)$. For a given parametric risk measure the above expression can be used for a variety of choices of parameter estimators, and for a given parameter estimator the expression can be used for a variety of parametric risk measures.

In the frequently occurring case where the parameter estimator is an MLE, it was shown in Hampel *et al.* (1986) [10] that $IF_{\hat{\theta}}(r)$ has the special form

$$IF_{\hat{\theta}}(r; \theta) = I_{\theta}^{-1} \cdot \psi(r; \theta) \tag{3.8}$$

where I_{θ} is the information matrix, and $\psi(r; \theta)$ is the score function vector⁵. In this case the IF formula (3.7) has the general form:

$$IF_{\hat{\rho}}(r; F) = \nabla \rho(\theta(F_p))' I_{\theta}^{-1} \psi(r; \theta). \tag{3.9}$$

3.2. Parametric t-Distribution Expected Shortfall Influence Function

For a parametric t-distribution ES based on maximum-likelihood parameter estimates, the formula (3.9) becomes

$$IF_{ES}(r; \mu, s, \nu) = \nabla \rho_{ES}(\mu, s, \nu) \cdot I_{\mu, s, \nu}^{-1} \cdot \psi_{ES}(r; \mu, s, \nu) \tag{3.10}$$

where $\nabla \rho_{ES}(\mu, s, \nu)$ is the gradient vector of $\rho_{ES}(\mu, s, \nu) = ES_{\gamma}(\mu, s, \nu)$ given by Equations (2.5) and (2.6), $I_{\mu, s, \nu}$ is the Fisher information matrix, and $\psi_{ES}(r; \mu, s, \nu)$ is the t-distribution score function. The risk measure gradient vector is given by

$$\nabla \rho_{ES}(\mu, s, \nu) = \left(-1, \frac{g_{\gamma, \nu}}{\gamma}, \frac{s}{\gamma} \cdot \frac{\partial g_{\gamma, \nu}}{\partial \nu} \right)'$$

with the partial derivative approximated by a finite difference quotient. The $\psi_{ES}(r; \theta)$ score vector is

⁵With $l(\theta) = \ln f(r; \theta)$ the log-likelihood for a single observation, the vector score function $\psi(r; \theta)$ has components $\psi_{\theta_k}(r) = \partial l(\theta) / \partial \theta_k, k = 1, 2, \dots, K$, and the $K \times K$ information matrix $I_{\theta} = E_r[\psi(r; \theta)\psi'(r; \theta)]$ has elements $I_{k,j} = E_r[\psi_{\theta_k}(r) \cdot \psi_{\theta_j}(r)]$.

$$\begin{bmatrix} \psi_\mu(r) \\ \psi_s(r) \\ \psi_\nu(r) \end{bmatrix} = \begin{bmatrix} \frac{(v+1)(r-\mu)}{vs^2 + (r-\mu)^2} \\ \psi_s(r) \\ \frac{1}{2} \left(\Omega\left(\frac{v+1}{2}\right) - \Omega\left(\frac{v}{2}\right) \right) - \frac{1}{2} \log \left(1 + \frac{(r-\mu)^2}{vs^2} \right) + \frac{s}{2v} \psi_s(r) \end{bmatrix} \quad (3.11)$$

where the formula of the *scale score function* $\psi_s(r)$ is

$$\psi_s(r) = \frac{v}{s} \cdot \frac{(r-\mu)^2 - s^2}{vs^2 + (r-\mu)^2}$$

And $\Omega(v) = \frac{d \log \Gamma(v)}{dv}$ is the digamma function. The formula for the information matrix $I_{\mu,s,\nu}$ was derived by Lucas (1997) [12] and may be found in MZ2019 [6].

Note that $IF_{ES}(r; \mu, s, \nu)$ depends on r solely through its appearance in the three components $\psi_\mu(r)$, $\psi_s(r)$, $\psi_\nu(r)$ of the score function vector, and for large r this dependence is symmetric about μ . In this regard, the behavior of $IF_{ES}(r; \mu, s, \nu)$ has a quite undesirable non-monotonic approximately symmetric behavior similar to that of the parametric normal distribution influence function in Equation (2.8) and the right-hand plot of **Figure 1**. This behavior is displayed in the left-hand plot of **Figure 2**, which is discussed further in the next section in comparison with an alternative parametric t-distribution ES M-estimator with semi-scale, and its influence function.

4. Parametric T-Distribution Semi-Scale Expected Shortfall M-Estimator and Influence Function

In this Section we derive a semi-scale modified version of the parametric t-distribution expected shortfall estimator, whose influence function does not suffer from the approximately symmetric behavior of the parametric t-distribution expected shortfall exhibited in the previous Section. Since the semi-scale parameter estimator is not an MLE, the MLE results of the previous Section no longer apply. However, the following closely related *M-estimator*, as discussed in Hampel *et al.* (1986) [10] can be used in place of an MLE.

A general M-estimator functional $\theta(F)$ is defined as the solution for θ of the vector equation

$$E_F[\boldsymbol{\psi}(r; \boldsymbol{\theta})] = \int \boldsymbol{\psi}(r; \boldsymbol{\theta}) dF(r) = 0 \quad (4.1)$$

where $\boldsymbol{\psi}(r; \boldsymbol{\theta})$ is a vector-valued general “score function”, which depends on a vector parameter $\boldsymbol{\theta}$ and a scalar return variable r^6 . An M-estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_n$ is obtained from a set of returns r_1, r_2, \dots, r_n by replacing F in (2.11) with the empirical distribution F_n :

⁶The above equation becomes an MLE estimating equation for the special choice $\boldsymbol{\psi}_{MLE}(r; \boldsymbol{\theta}) = -d \ln f(r; \boldsymbol{\theta}) / d\boldsymbol{\theta}$, but here we move beyond the MLE.

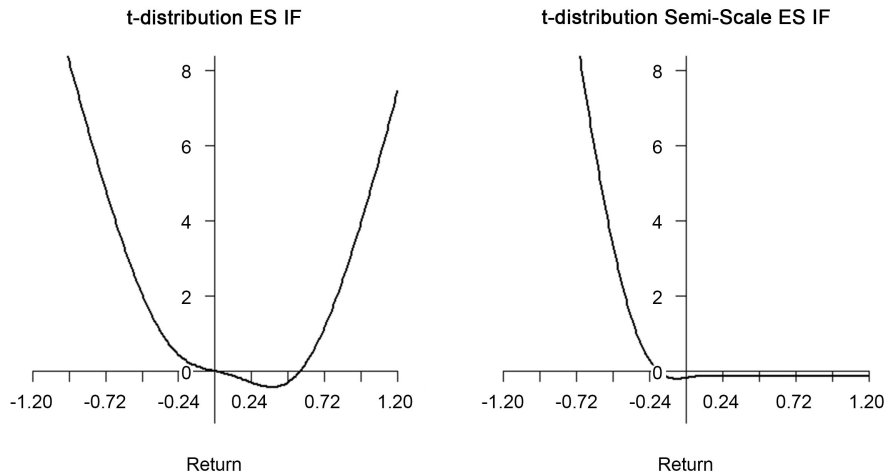


Figure 2. t-Distribution parametric symmetric and semi-scale ES influence function with tail probability $\gamma = 5\%$ for t-distribution with $\gamma = 10$ degrees of freedom, mean 0.12 and standard deviation 0.24.

$$\sum_{i=1}^n \psi(r_i; \hat{\theta}) = 0. \tag{4.2}$$

For our application where the distribution of r is in a parametric family $F_{\theta}, \theta \in \Theta$, we assume the Fisher consistency condition

$$E_{F_{\theta}} [\psi(r; \theta)] = \int \psi(r; \theta) dF_{\theta}(r) = 0, \theta \in \Theta \tag{4.3}$$

In other words, the expected value of the score function at the true parameter value is zero. Correspondingly, an M-estimator defined by (4.2) converges in probability to a solution of the asymptotic estimating Equation (4.3).

For many estimation problems $\psi(r; \theta)$ may be defined as the derivative $\psi(r; \theta) = d\rho(r; \theta)/d\theta$. Of a loss function $\rho(r; \theta)$. In particular the log-likelihood function is a choice.

It is shown in Hampel *et al.* (1986) [10] that the influence function of an M-estimator is given by

$$IF(r; \theta(F)) = M^{-1}(\psi, F) \psi(r; \theta(F)) \tag{4.4}$$

where the M-matrix $M(\psi, F)$ is the $p \times p$ matrix

$$M(\psi, F) = - \int \left[\frac{\partial}{\partial \theta} \psi(r; \theta) \right]_{\theta = \theta(F)} dF(r). \tag{4.5}$$

Combining (3.7) and (4.4) we have the following general expression for the influence function of a parametric risk measure based on a general M-estimator of the unknown parameters:

$$\begin{aligned} IF_{\rho}(r; F) &= \nabla \rho(\theta(F))' \cdot IF_{\theta}(r) \\ &= \nabla \rho(\theta(F))' \cdot M^{-1}(\psi, F) \psi(r; \theta(F)) \end{aligned} \tag{4.6}$$

Note that the gradient vector in the above expression depends only on the risk measure chosen and the M-estimator functional that represents the asymptotic value of the M-estimator. For parametric distributions and consistent estimators,

this gradient only depends on the parametric risk measure and the distribution parameters. However, for any given distribution the influence function part of the above expression depends only on the choice of M-score function ψ .

We seek to modify the t-distribution MLE score function of the previous Section so that the scale score function is a semi-scale score function, and the expected value of the resulting M-score function is still zero. One way to do this is as follows. Define the semi-scale score function by

$$\begin{aligned} \psi(r) &= \begin{bmatrix} \tilde{\psi}_\mu(r) \\ \tilde{\psi}_s(r) \\ \tilde{\psi}_v(r) \end{bmatrix} \\ &= \begin{bmatrix} \psi_\mu(r) \\ \tilde{\psi}_s(r) \\ \frac{1}{2} \left(\Omega\left(\frac{v+1}{2}\right) - \Omega\left(\frac{v}{2}\right) \right) - \frac{1}{2} \log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{vs^2} \right) - c + \frac{s}{2v} \tilde{\psi}_s(r) \end{bmatrix} \end{aligned} \tag{4.7}$$

where the location score function $\psi_\mu(r)$ is the same as for a t-distribution MLE as given by (3.11), and $\tilde{\psi}_s(r) = \tilde{\psi}_s(r; \mu, s, v)$ is the semi-scale score function

$$\tilde{\psi}_s(r) = \frac{(v+1)(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{vs^3 + s \cdot (r-\mu)^2} - \frac{1}{2s} \tag{4.8}$$

where the constant c remains to be determined.

In order to ensure consistency of the M-estimator defined by the above M-score function, the latter must satisfy the zero-expectation condition (4.2) under a t-distribution for returns. This is already the case for $\psi_\mu(r)$, and we now show this is also the case for $\tilde{\psi}_s(r)$.

In order to see that the expected value of $\tilde{\psi}_s(r)$ is zero, note that from the t-distribution MLE score function (3.11) we have

$$\begin{aligned} s \cdot \psi_s(r) &= \frac{(v+1)(r-\mu)^2}{vs^2 + (r-\mu)^2} - 1 \\ &= \frac{(v+1)(r-\mu)^2}{vs^2 + (r-\mu)^2} \cdot I_{(-\infty, \mu]}(r) + \frac{(v+1)(r-\mu)^2}{vs^2 + (r-\mu)^2} \cdot I_{(\mu, \infty)}(r) - 1. \end{aligned} \tag{4.9}$$

Since the expected value of $\psi_s(r)$ is zero the sum of the first two terms on the right is one, and from the symmetry of the t-distribution about μ the expected value of each of these first two terms must be equal, and hence equal to one-half. Thus

$$\begin{aligned} E[\tilde{\psi}_s(r)] &= \frac{1}{s} E \left[\frac{(v+1)(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{vs^2 + (r-\mu)^2} - \frac{1}{2} \right] \\ &= \frac{1}{s} E \left[\frac{1}{2} - \frac{1}{2} \right] \\ &= 0. \end{aligned}$$

Now we just need to choose c so that the M-score $\psi_\nu(r)$ for degrees of freedom has zero expected value. First note that we require that

$$\begin{aligned} E[\psi_\nu(r)] &= \frac{1}{2} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right) - E \left[\frac{1}{2} \log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{\nu s^2} \right) \right] - c + \frac{s}{\nu} E[\psi_s(r)] \\ &= \frac{1}{2} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right) - E \left[\frac{1}{2} \log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{\nu s^2} \right) \right] - c \\ &= 0 \end{aligned}$$

which means that

$$c = \frac{1}{2} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right) - \frac{1}{2} E \left[\log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{\nu s^2} \right) \right]. \tag{4.10}$$

Noting that

$$\log \left(1 + \frac{(r-\mu)^2}{\nu s^2} \right) = \log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{\nu s^2} \right) + \log \left(1 + \frac{(r-\mu)^2 \cdot I_{(\mu, +\infty)}(r)}{\nu s^2} \right) \tag{4.11}$$

along with the symmetry of the t-distribution about μ shows that the expected value of the last two terms must be equal, and thus

$$E \left[\log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{\nu s^2} \right) \right] = \frac{1}{2} \cdot E \left[\log \left(1 + \frac{(r-\mu)^2}{\nu s^2} \right) \right] \tag{4.12}$$

Plugging (4.12) into (4.10) gives

$$\begin{aligned} c &= \frac{1}{2} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right) - \frac{1}{4} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right) \\ &= \frac{1}{4} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right). \end{aligned}$$

Thus, we have⁷

$$\begin{bmatrix} \tilde{\psi}_\mu(r) \\ \tilde{\psi}_s(r) \\ \tilde{\psi}_\nu(r) \end{bmatrix} = \begin{bmatrix} \psi_\mu(r) \\ \tilde{\psi}_s(r) \\ \frac{1}{4} \left(\Omega\left(\frac{\nu+1}{2}\right) - \Omega\left(\frac{\nu}{2}\right) \right) - \frac{1}{2} \log \left(1 + \frac{(r-\mu)^2 \cdot I_{(-\infty, \mu]}(r)}{\nu s^2} \right) + \frac{s}{2\nu} \tilde{\psi}_s(r) \end{bmatrix}. \tag{4.13}$$

Now to get the expression for the parameter estimator score function in (4.4), we just need to evaluate the M-matrix

⁷We note that this choice of M-score function is not unique. Another valid choice would be

$$\tilde{\psi}_{2s}(r) = \frac{\nu}{s} \left(\frac{(r-\mu)^2 - s^2}{\nu s^2 + (r-\mu)^2} \right) \cdot I_{(-\infty, \mu]}(r).$$

However, it is easily verified that this choice of scale score function is discontinuous at $r = \mu$ and it is a basic principle in robust statistics is that discontinuous influence functions are to be avoided.

$$M(\psi, F) = M(\mu, s, \nu) = \begin{pmatrix} E\left[\frac{\partial \psi_\mu}{\partial \mu}\right] & E\left[\frac{\partial \psi_\mu}{\partial s}\right] & E\left[\frac{\partial \psi_\mu}{\partial \nu}\right] \\ E\left[\frac{\partial \tilde{\psi}_s}{\partial \mu}\right] & E\left[\frac{\partial \tilde{\psi}_s}{\partial s}\right] & E\left[\frac{\partial \tilde{\psi}_s}{\partial \nu}\right] \\ E\left[\frac{\partial \tilde{\psi}_\nu}{\partial \mu}\right] & E\left[\frac{\partial \tilde{\psi}_\nu}{\partial s}\right] & E\left[\frac{\partial \tilde{\psi}_\nu}{\partial \nu}\right] \end{pmatrix} \quad (4.14)$$

where the expectations are taken with $F = F_\theta$. Straightforward but tedious calculations (see Appendix A in an early draft version of this paper available at SSRN, <https://ssrn.com/abstract=4605604>) show that

$$M = \begin{pmatrix} \frac{1}{s^2} \left[1 - \frac{2}{\nu+3} \right] & 0 & 0 \\ \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2) \cdot \sqrt{\nu\pi}} \cdot \frac{-2(\nu+1)}{(\nu+3) \cdot s^2} & \frac{1}{s^2} \cdot \frac{\nu}{\nu+3} & \frac{1}{2s} \left[\frac{1}{\nu+3} - \frac{1}{\nu+1} \right] \\ \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2) \cdot \sqrt{\nu\pi}} \cdot \frac{\nu-1}{(\nu+3) \cdot (\nu+1) \cdot \nu \cdot s} & \frac{1}{2s} \left(\frac{1}{\nu+3} - \frac{1}{\nu+1} \right) & \frac{1}{8} \left[\Omega' \left(\frac{\nu}{2} \right) - \Omega' \left(\frac{\nu+1}{2} \right) \right] - \frac{\nu+5}{4\nu(\nu+1)(\nu+3)} \end{pmatrix}$$

which has the form

$$M = \begin{pmatrix} A & 0 & 0 \\ B & C & D \\ E & D & F \end{pmatrix} \quad (4.15)$$

with inverse

$$M^{-1} = \frac{1}{A(CF - D^2)} \begin{pmatrix} CF - D^2 & 0 & 0 \\ ED - BF & AF & -AD \\ BD - CE & -AD & AC \end{pmatrix}. \quad (4.16)$$

The left-hand plot in **Figure 2** displays $\gamma = 5\%$ tail probability parametric t-Distribution ES estimator influence function for with Degrees of Freedom = 10, mean = 0.12 and standard deviation = 0.24. Except for the curious negative bump for small positive values of the Return r , this influence function is essentially symmetric about the $\mu = 0.12$, as we pointed out in the last paragraph of Section 3. Thus, the parametric t-distribution ES influence function suffers from the lack of monotonicity in a manner similar to that of the parametric normal distribution ES estimator influence function in the right-hand plot of **Figure 1**.

On the other hand, the right-hand plot in **Figure 2** displays the parametric t-distribution semi-scale M-estimator influence function, with the same μ, s, ν parameters as for the influence function in the left-hand plot. But now, except for the curious negative bump for small negative values of the Return r , this influence function has the desirable monotonic decreasing character similar to that of the nonparametric ES influence function in the left-hand plot of **Figure 1**.

Figure 3 gives a good feeling for how the shape of the parametric t-distribution ES M-estimator influence function changes with changes in the t-distribution degrees of freedom and ES tail probability, for the four degrees of freedom

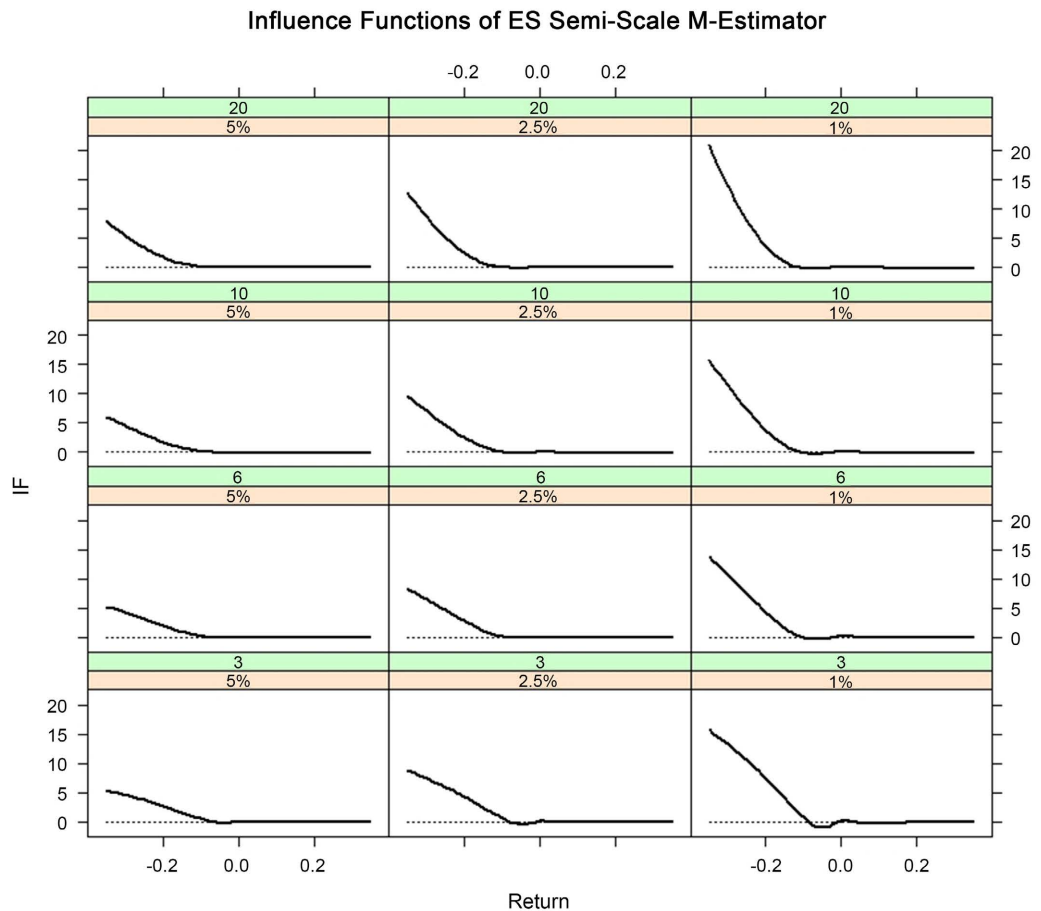


Figure 3. Influence functions of ES semi-scale m-estimators with monthly $\mu = 1\%$, $s = 7\%$, and annual SR = 0.5

parameter ν values 20, 10, 6, 3, and the three tail probability parameter γ values 1%, 2.5%, 5%. These influence functions all have the desirable shape that there is essentially zero influence for positive values of return r , and positive influence that increases rapidly for decreasing negative return values. As one expects, the positive values of the influence function increase for each fixed negative r as the tail probability γ decreases. The behavior of the shape of the influence functions for negative returns close to 0 as the degrees of freedom decrease from 20 to 3 is more subtle, with the shape being slightly convex for $\nu = 20$ and slightly concave for $\nu = 3$. The latter is related to the fact, demonstrated in MZ2019 [6], that the t-distribution maximum-likelihood ES influence function is logarithmically unbounded.

5. Implementation of the Parametric Semi-Scale ES Estimator

Now we propose to construct an ES semi-scale M-estimator for risk monitoring, with the following straightforward steps.

- (1) First compute the t-distribution MLE estimates $(\hat{\nu}, \hat{s}_0, \hat{\mu})$, for example using the Azzalini SN package [13] available on CRAN

(<https://cran.r-project.org/web/packages/sn/index.html>).

(2) Compute the semi-scale parameter estimate as follows. Plug the $\hat{\mu}$ and $\hat{\nu}$ MLE's from step one into the semi-scale score function $\tilde{\psi}_s(r; \mu, s, \nu)$. Then the semi-scale parameter estimate \hat{s}_{semi} can be computed by solving the equation:

$$\sum_{i=1}^n \tilde{\psi}_s(r_i; \hat{\mu}, s, \hat{\nu}) = 0.$$

Note that above summation has the form

$$\begin{aligned} \sum_{i=1}^n \tilde{\psi}_s(r_i; \hat{\mu}, s, \hat{\nu}) &= \sum_{i=1}^n \left[\frac{(\hat{\nu}+1)(r_i - \hat{\mu})^2 \cdot I_{(-\infty, \hat{\mu}]}(r_i)}{\hat{\nu}s^3 + s \cdot (r_i - \mu)^2} - \frac{1}{2s} \right] \\ &= \frac{1}{s} \cdot \sum_{i=1}^n \left[\frac{(\hat{\nu}+1)(r_i - \hat{\mu})^2 \cdot I_{(-\infty, \hat{\mu}]}(r_i)}{\hat{\nu}s^2 + (r_i - \mu)^2} - \frac{1}{2} \right] \\ &= \frac{\hat{\nu}+1}{s} \cdot \left[\sum_{i=1}^n \frac{(r_i - \hat{\mu})^2 \cdot I_{(-\infty, \hat{\mu}]}(r_i)}{\hat{\nu}s^2 + (r_i - \mu)^2} - \frac{n}{2(\hat{\nu}+1)} \right] \end{aligned}$$

and to compute \hat{s}_{semi} , one just needs to solve the following equation whose left-hand side is strictly monotonic in s :

$$\sum_{i=1}^n \frac{(r_i - \hat{\mu})^2 \cdot I_{(-\infty, \hat{\mu}]}(r_i)}{\hat{\nu}s^2 + (r_i - \hat{\mu})^2} - \frac{n}{2(\hat{\nu}+1)} = 0.$$

Any simple search algorithm will suffice, for example using the Newton-Raphson method package `rootSolve` available on CRAN

(<https://CRAN.R-project.org/package=rootSolve>).

(3) Finally, plug $(\hat{\nu}, \hat{s}_{semi}, \hat{\mu})$ into the parametric t-distribution ES expression (2.5) to obtain the ES semi-scale M-estimator:

$$ES_{\gamma}(\hat{\mu}, \hat{s}_{semi}, \hat{\nu}) = -\hat{\mu} + \frac{g_{\gamma, \hat{\nu}}}{\gamma} \cdot \hat{s}_{semi}.$$

It remains to carry out some empirical studies of the performance of this risk estimator.

6. ES Semi-Scale M-Estimator Asymptotic Variance

The asymptotic variance of a consistent M-estimator $\hat{\theta}_n = \theta(F_n)$ of θ has the form

$$V(\theta(F), F) = M^{-1}(\psi, F) \cdot Q(\psi, F) \cdot M^{-1}(\psi, F)' \tag{6.1}$$

where

$$\begin{aligned} Q(\psi, F) &= \int \psi(r; \theta(F)) \psi'(r; \theta(F)) dF(r) \\ &= \begin{pmatrix} E[\tilde{\psi}_{\mu} \tilde{\psi}_{\mu}] & E[\tilde{\psi}_{\mu} \tilde{\psi}_s] & E[\tilde{\psi}_{\mu} \tilde{\psi}_{\nu}] \\ E[\tilde{\psi}_s \tilde{\psi}_{\mu}] & E[\tilde{\psi}_s \tilde{\psi}_s] & E[\tilde{\psi}_s \tilde{\psi}_{\nu}] \\ E[\tilde{\psi}_{\nu} \tilde{\psi}_{\mu}] & E[\tilde{\psi}_{\nu} \tilde{\psi}_s] & E[\tilde{\psi}_{\nu} \tilde{\psi}_{\nu}] \end{pmatrix} \end{aligned} \tag{6.2}$$

See for example Hampel *et al.* (1986) [10]. In the special case of MLE estima-

tors both \mathbf{M} and \mathbf{Q} reduce to the information matrix $\mathbf{I}(\boldsymbol{\theta})$ and the expression (6.1) reduces to $\mathbf{I}(\boldsymbol{\theta})^{-1}$ as expected.

Straightforward but tedious derivations (see Appendix B in an early draft version of this paper available at SSRN, <https://ssrn.com/abstract=4605604>) give the following expressions:

$$E[\psi_v \psi_v] = \frac{1}{16} \left[\Omega\left(\frac{v+1}{2}\right) - \Omega\left(\frac{v}{2}\right) - \frac{1}{v} \right]^2 + \frac{1}{8} \left[\Omega'\left(\frac{v}{2}\right) - \Omega'\left(\frac{v+1}{2}\right) \right] - \frac{1}{2v} \left[\frac{1}{v+1} - \frac{1}{2(v+3)} \right] \tag{6.3}$$

$$E[\psi_s \psi_s] = \frac{1}{s^2} \left[\frac{v}{v+3} + \frac{1}{4} \right] \tag{6.4}$$

$$E[\psi_\mu \psi_\mu] = \frac{1}{s^2} \left[1 - \frac{2}{v+3} \right] \tag{6.5}$$

$$E[\psi_v \psi_s] = \frac{1}{2s} \left[\frac{1}{v+3} - \frac{1}{v+1} \right] - \frac{1}{8s} \left[\Omega\left(\frac{v+1}{2}\right) - \Omega\left(\frac{v}{2}\right) - \frac{1}{v} \right] \tag{6.6}$$

$$E[\psi_v \psi_\mu] = \frac{\Gamma((v+1)/2)}{\Gamma(v/2) \cdot \sqrt{v\pi} \cdot s} \cdot \frac{v-1}{(v+3)(v+1)v} \tag{6.7}$$

$$E[\psi_s \psi_\mu] = \frac{\Gamma((v+1)/2)}{\Gamma(v/2) \cdot \sqrt{v\pi} \cdot s^2} \cdot \frac{-2(v+1)}{v+3} \tag{6.8}$$

Since our t-distribution ES Semi-scale M-estimator is a small modification of the t-distribution ES MLE, one expects that the increase in the asymptotic variance of this estimator, relative to a t-distribution MLE, will not be very great. **Figure 4** and **Figure 5**, which are based on standard errors (SE's) obtained as the square root of the asymptotic variances of the ES semi-scale estimator using 6 degrees of freedom, indicate that indeed the increase in variance is not very great.

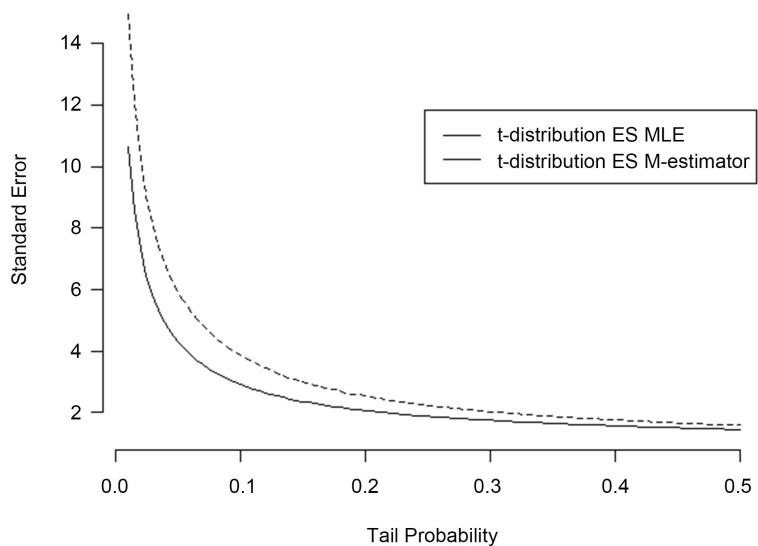


Figure 4. Asymptotic standard error of t-distribution ES semi-scale m-estimator and the asymptotic standard error of the ES MLE for a t-distribution with $\mu = 0, s = 1, \nu = 5$.

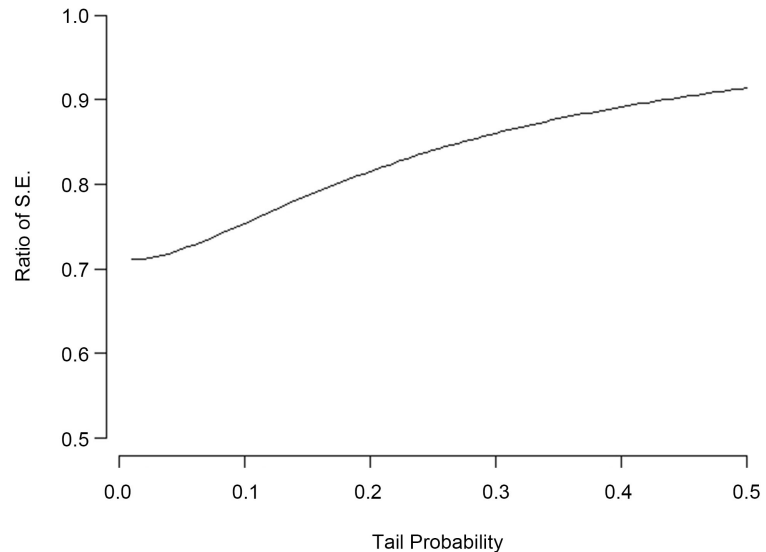


Figure 5. Ratio of standard error of t-distribution ES estimators: semi-scale m-estimator versus maximum likelihood ES estimator for 5 degrees of freedom.

7. Concluding Comments

We have introduced a new ES Semi-Scale M-Estimator by replacing the scale estimator of an ES t-distribution joint MLE of the location, scale, and degrees of freedom, with a semi-scale estimator, and we derived the new estimator's influence function and asymptotic variance formula. The mathematical form of the new estimator's estimating equation and influence function show that the estimator avoids the unsatisfactory behavior of the t-distribution ES MLE that (large) positive returns indicate (large) risk.

Since an ES semi-scale M-estimator influence function is not exactly monotonic decreasing as returns increase, we cannot assert that this ES is a coherent risk measure, as is a mean semi-deviation estimator. However, the fact that such influence functions are nearly monotonic decreasing suggests that ES semi-scale M-estimators will have good properties for risk reporting. What is needed now is an in-depth empirical study of the relative performance of the ES semi-scale M-estimators and the mean semi-deviation coherent risk measures, using both simulated and real returns whose distributions range from approximately normal to moderately fat-tailed (e.g., t-distributions with 8 to 12 degrees of freedom), and to very fat-tailed distributions (e.g., t-distributions with 3 to 6 degrees of freedom).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] JPMorgan/Reuters (1996) RiskMetrics—Technical Document. 4th Edition.

- <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95a>
- [2] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) Coherent Measures of Risk. *Mathematical Finance*, **9**, 203-228. <https://doi.org/10.1111/1467-9965.00068>
- [3] McNeil, A.J., Frey, R. and Embrechts, P. (2015) Quantitative Risk Management. Princeton University Press, Princeton.
- [4] Rockafellar, R.T. and Uryasev, S. (2000) Optimization of Conditional Value-at-Risk. *Journal of Risk*, **2**, 21-41. <https://doi.org/10.21314/JOR.2000.038>
- [5] Fischer, T. (2003) Risk Capital Allocation by Coherent Risk Measures Based on One-Sided Moments. *Insurance: Mathematics and Economics*, **32**, 135-146. [https://doi.org/10.1016/S0167-6687\(02\)00209-3](https://doi.org/10.1016/S0167-6687(02)00209-3)
- [6] Martin, R.D. and Zhang, S. (2019) Non-Parametric versus Parametric Expected Shortfall. *Journal of Risk*, **21**, 1-41. <https://doi.org/10.21314/JOR.2019.416>
- [7] Jorion, P. (2007) Value-at-Risk. 3rd Edition, McGraw-Hill, New York.
- [8] Hampel, F.R. (1974) The Influence Curve and Its Role in Robust Estimation. *Journal of American Statistical Association*, **69**, 383-393. <https://doi.org/10.1080/01621459.1974.10482962>
- [9] Zhang, S. (2016) Two Equivalent Parametric Expected Shortfall Formulas for T-Distributions. <https://ssrn.com/abstract=2883935>
- [10] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986) Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons, Inc., New York.
- [11] Zhang, S., Martin, R.D. and Chritidis, A.A. (2021) Influence Functions for Risk and Performance Estimators. *Journal of Mathematical Finance*, **1**, 1-33. <https://doi.org/10.4236/jmf.2021.111002>
- [12] Lucas, A. (1997) Robustness of the Student-t Based M-Estimator. *Communications in Statistics-Theory and Methods*, **26**, 1165-1182. <https://doi.org/10.1080/03610929708831974>
- [13] Azzalini, A. (2023). <http://azzalini.stat.unipd.it/SN/>