Estimation of the marginal expected shortfall: the mean when a related variable is extreme

Juan-Juan Cai

Tilburg University

John H.J. Einmahl

Tilburg University

Laurens de Haan *

Erasmus University Rotterdam

University of Lisbon

Chen Zhou †

De Nederlandsche Bank

Erasmus University Rotterdam

August 2, 2012

Abstract. Denote the loss return on the equity of a financial institution as X and that of the entire market as Y. For a given very small value of $p > 0$, the marginal expected shortfall (MES) is defined as $E(X \mid Y > Q_Y(1-p))$, where $Q_Y(1-p)$ is the $(1-p)$-th quantile of the distribution of Y. The MES is an important factor when measuring the systemic risk of financial institutions. For a wide nonparametric class of bivariate distributions, we construct an estimator of the MES and establish the asymptotic normality of the estimator when $p \downarrow 0$, as the sample size $n \to \infty$. Since we are in particular interested in the case $p = O(1/n)$, we use extreme value techniques for deriving the estimator and its asymptotic behavior. The finite sample performance of the estimator and the adequacy of the limit theorem are shown in a detailed simulation study. We also apply our method to estimate the MES of three large U.S. investment banks.

---

1   Introduction

The financial crisis has led market participants, academics and regulators to further investigate the systemic risk in the financial industry. One key issue is to have a measure of systemic risk that is both economically sound and statistically assessable. Among others, Acharya et al. (2010) proposes a systemic risk measure under a general equilibrium model. An important factor in constructing this systemic risk measure is the contribution of a financial institution to a systemic crisis measured by the Marginal Expected Shortfall (MES). The MES of a financial institution is defined as the expected loss on its equity return conditional on the occurrence of an extreme loss in the aggregated return of the financial sector. In this paper, we provide an empirical methodology based on multivariate extreme value theory (EVT) to estimate the MES.

Denote the loss of the equity return of a financial firm and that of the entire market as X and Y, respectively. Then the MES is defined as $E(X \mid Y > t)$, where t is a high threshold such that $p = P(Y > t)$ is extremely small. In other words, the threshold t is the $(1-p)$-th quantile of the distribution of Y defined by $P(Y > Q_Y(1-p)) = p$. Thus, the MES at probability level p is defined as

$$MES(p) = E(X \mid Y > Q_Y(1-p)).$$

Notice that in applications the probability p is at an extremely low level that can be even lower than $1/n$, where n is the sample size of historical data that are used for estimating the MES.[1] Therefore, this is a typical extreme value problem.

We tackle this problem by a two-stage approach. Firstly, we consider the estimation of the MES at an intermediate probability level. More specifically, we consider an intermediate sequence $k = k(n)$ such that $k/n \to 0$, $k \to \infty$, as $n \to \infty$, and estimate $MES(k/n)$. At such an intermediate level, there are many observations (X, Y) such that $Y > Q_Y(1 - k/n)$. Thus the $MES(k/n)$ can be estimated non-parametrically by taking the average over the X components of those selected observations. Secondly, we follow the usual extrapolation method in extreme value statistics to obtain an estimator of $MES(p)$ for the intended low probability level p. Assuming that X follows a heavy-tailed distribution with finite mean, and the dependence structure of (X, Y) follows the bivariate EVT framework, we show that the MES can be extrapolated in a similar way as that for high quantiles of the distribution of X. More precisely, we estimate $MES(p)$ by multiplying the estimator of $MES(k/n)$ with an extrapolation factor, which is

---

[1]In Acharya et al. (2010), the probability of such an extreme tail event is specified as "that happen once or twice a decade (or less)", whereas the estimation is based on daily data from one year.

the same as that used in extrapolating the $(1 - k/n)$-quantile to the $(1 - p)$-th quantile of the distribution of X. Our main result establishes the asymptotic normality of the thus obtained estimator of $MES(p)$. This result makes statistical inference for the MES feasible for a large class of bivariate distributions. In addition, we show through a simulation study that the estimator performs well and that this limit theorem provides an adequate approximation for finite sample sizes.

The MES has been studied under the name "Conditional Tail Expectation" (CTE, or TCE) in statistics and actuarial science. The definition of CTE in a univariate context is the same as that of the tail value at risk. Mathematically, it is given by $E(X \,|\, X > Q_X(1-p))$ where $Q_X(1-p)$ is the $(1-p)$-th quantile of the distribution of X. In case X has a continuous distribution, this is also called the expected shortfall. Compared to the MES, it can be viewed as the special case that $Y = X$. The concept of CTE has been defined more generally in a multivariate setup. It is possible to have the conditioning event defined by another, related random variable Y exceeding its high quantile. In that case, the CTE coincides with the MES.

A few studies show how to calculate the CTE when the joint distribution of (X, Y) follows specific parametric models. For example, Landsman and Valdez (2003) and Kostadinov (2006) deal with elliptical distributions with heavy-tailed marginals. Cai and Li (2005) studies the CTE for multivariate phase-type distributions. Vernic (2006) considers skewed-normal distributions. Compared to these studies, our approach does not impose any parametric structure on (X, Y). A comparable result in the literature is the approach in Joe and Li (2011), where under multivariate regular variation, a formula for calculating the CTE is provided. The multivariate regularly varying distributions form a subclass of our model. Note that we do not make any assumption on the marginal distribution of Y. Furthermore, in contrast to these papers where only probabilistic calculations of the MES are provided, we focus on the statistical problem of estimating it and studying the performance of the estimator.

In Acharya et al. (2010), an estimation procedure for the MES *is* provided which is similar to our approach. Nevertheless, in order to obtain such an estimator, the probabilistic model is restricted to a specific linear relationship between X and Y. Such an assumption falls into the scope of our general framework. Thus their approach can be seen as a special case of our much broader setup. A similar parametric approach has been adopted in Brownlees and Engle (2011), where, as one step forward, a nonparametric kernel estimator of the MES is proposed. Such a kernel estimation method, however, is valid only if the threshold for defining a systemic crisis is not too high, that is, there are still several observations corresponding to the event $\{Y > Q_Y(1-p)\}$. In other words, the tail probability level p should be substantially larger

than $1/n$. Such a method cannot handle extreme events, that is $p < 1/n$, which is particularly required for systemic risk measures.

To summarize, compared to existing studies on MES or CTE, four advantages arise for our method. Firstly, we allow a wide, nonparametric class of dependence structures of (X, Y), which covers in particular most of the parametric models used in literature. Secondly, we do not impose any assumption on the marginal distribution of Y. This allows other potential applications of our estimation method on the MES measure. For example, in the context of measuring systemic risk, the condition $\{Y > Q_Y(1-p)\}$ is considered as the definition of a "systemic crisis". Our distribution-free feature on Y allows other potential indicators in defining "systemic crisis". Thirdly, we handle an extreme value statistics problem: we provide a novel estimator of $MES(p)$, corresponding to an extremely low probability level p, that can be lower than $1/n$. Lastly, the main contribution of this paper is the unraveling of the asymptotic behavior of this estimator.

The paper is organized as follows. Section 2 provides the main result: asymptotic normality of the estimator. In Section 3, a simulation study shows the good performance of the estimator. An application on estimating the MES for U.S. financial institutions is given in Section 4. The proofs are deferred to Section 5.

2 Main Results

Let (X, Y) be a random vector with a continuous distribution function F. Denote the marginal distribution functions as $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$ with corresponding tail quantile functions given by $U_j = \left(\frac{1}{1-F_j} \right)^{\leftarrow}$, $j = 1, 2$, where $\leftarrow$ denotes the left-continuous inverse. Then the MES at a probability level p can be written as

$$\theta_p := E(X \mid Y > U_2(1/p)).$$

The goal is to estimate θ_p based on independent and identically distributed (i.i.d.) observations, $(X_1, Y_1), \cdots, (X_n, Y_n)$ from F, where $p = p(n) \to 0$ as $n \to \infty$.

We adopt the bivariate EVT framework for modeling the tail dependence structure of (X, Y). Suppose for all $(x, y) \in [0, \infty]^2 \setminus \{(+\infty, +\infty)\}$, the following limit exists:

$$\lim_{t \to \infty} tP(1 - F_1(X) \leq x/t, 1 - F_2(Y) \leq y/t) =: R(x, y). \tag{1}$$

The function R completely determines the so-called stable tail dependence function l, as for all $x, y \geq 0$,

$$l(x, y) = x + y - R(x, y);$$

4

see Drees and Huang (1998); Beirlant et al. (2004, Chapter 8.2).

For the marginal distributions, we assume that only X follows a distribution with a heavy right tail: there exists $\gamma_1 > 0$ such that for $x > 0$,

$$\lim_{t \to \infty} \frac{U_1(tx)}{U_1(t)} = x^{\gamma_1}. \tag{2}$$

Then it follows that $1 - F_1$ is regularly varying with index $-\frac{1}{\gamma_1}$ and γ_1 is the extreme value index.

We first focus on X being positive, then we consider $X \in \mathbb{R}$. Throughout, there is no assumption, apart from continuity, on the marginal distribution of Y.

2.1 X Positive

Assume X takes values in $(0, \infty)$. The following limit result gives an approximation for θ_p.

Proposition 1. *Suppose that* (1) *and* (2) *hold with* $0 < \gamma_1 < 1$. *Then,*

$$\lim_{p \downarrow 0} \frac{\theta_p}{U_1(1/p)} = \int_0^\infty R\left(x^{-1/\gamma_1}, 1\right) dx.$$

In Joe and Li (2011, Theorem 2.4), this result is derived under the stronger assumption of multivariate regular variation.

Next, we construct an estimator of θ_p based on the limit given in Proposition 1. Let k be an intermediate sequence of integers, that is, $k \to \infty$, $k/n \to 0$, as $n \to \infty$. By Proposition 1 and a strengthening of (2) (see condition (b) below), we have that as $n \to \infty$,

$$\theta_p \sim \frac{U_1(1/p)}{U_1(n/k)} \theta_{\frac{k}{n}} \sim \left(\frac{k}{np}\right)^{\gamma_1} \theta_{\frac{k}{n}}. \tag{3}$$

For estimating θ_p, it thus suffices to estimate γ_1 and $\theta_{\frac{k}{n}}$.

We estimate γ_1 with the Hill (1975) estimator:

$$\hat{\gamma}_1 = \frac{1}{k_1} \sum_{i=1}^{k_1} \log X_{n-i+1,n} - \log X_{n-k_1,n}, \tag{4}$$

where k_1 is another intermediate sequence of integers and $X_{i,n}$, $i = 1, \ldots, n$ is the i-th order statistic of $X_1, \ldots, X_n$.

By regarding the $(n-k)$-th order statistic $Y_{n-k,n}$ of $Y_1, \ldots, Y_n$ as an estimator of $U_2(n/k)$, we construct a nonparametric estimator of $\theta_{\frac{k}{n}}$ which is the average of the selected X_i corresponding to the highest k values of Y:

$$\hat{\theta}_{\frac{k}{n}} = \frac{1}{k} \sum_{i=1}^{n} X_i I(Y_i > Y_{n-k,n}). \tag{5}$$

5

Combining (3), (4) and (5), we estimate θ_p by

$$\hat{\theta}_p = \left(\frac{k}{np}\right)^{\hat{\gamma}_1} \hat{\theta}_{\frac{k}{n}}. \tag{6}$$

We prove the asymptotic normality of $\hat{\theta}_p$ under the following conditions.

(a) There exist $\beta > \gamma_1$ and $\tau < 0$ such that, as $t \to \infty$,

$$\sup_{\substack{0 < x < \infty \\ 1/2 \leq y \leq 2}} \frac{|tP\left(1 - F_1(X) < x/t, 1 - F_2(Y) < y/t\right) - R(x,y)|}{x^\beta \wedge 1} = O(t^\tau).$$

(b) There exist $\rho_1 < 0$ and an eventually positive or negative function A_1 with $\lim_{t\to\infty} A_1(t) = 0$ such that

$$\lim_{t\to\infty} \frac{U_1(tx)/U_1(t) - x^{\gamma_1}}{A_1(t)} = x^{\gamma_1} \frac{x^{\rho_1} - 1}{\rho_1}.$$

As a consequence, $|A_1|$ is regularly varying with index ρ_1. Conditions (a) and (b) are natural second-order strengthenings of (1) and (2), respectively. We further require the following conditions on the intermediate sequences k_1 and k.

(c) As $n \to \infty$, $\sqrt{k_1}A_1(n/k_1) \to 0$.

(d) As $n \to \infty$, $k = O(n^\alpha)$ for some $\alpha < \min\left(\frac{-2\tau}{-2\tau+1}, \frac{2\gamma_1\rho_1}{2\gamma_1\rho_1+\rho_1-1}\right)$.

To characterize the limit distribution of $\hat{\theta}_p$, we define a mean zero Gaussian process W_R on $[0, \infty]^2 \backslash \{\infty, \infty\}$ with covariance structure

$$E(W_R(x_1, y_1)W_R(x_2, y_2)) = R(x_1 \wedge x_2, y_1 \wedge y_2),$$

i.e., W_R is a Wiener process. Set

$$\Theta = (\gamma_1 - 1)W_R(\infty, 1) + \left(\int_0^\infty R(s,1)ds^{-\gamma_1}\right)^{-1} \int_0^\infty W_R(s,1)ds^{-\gamma_1},$$

and

$$\Gamma = \gamma_1 \left(-W_R(1, \infty) + \int_0^1 s^{-1}W_R(s, \infty)ds\right).$$

It will be shown (see Proposition 3 and (24)) that $\hat{\theta}_{\frac{k}{n}}$ and $\hat{\gamma}_1$ are asymptotically normal with Θ and Γ as limit, respectively. The following theorem gives the asymptotic normality of $\hat{\theta}_p$.

6

Theorem 1. *Suppose conditions (a)–(d) hold and $\gamma_1 \in (0, 1/2)$. Assume $d_n := \frac{k}{np} \geq 1$ and $r := \lim_{n\to\infty} \frac{\sqrt{k} \log d_n}{\sqrt{k_1}} \in [0, \infty]$. If $\lim_{n\to\infty} \frac{\log d_n}{\sqrt{k_1}} = 0$, then, as $n \to \infty$,*

$$\min\left(\sqrt{k}, \frac{\sqrt{k_1}}{\log d_n}\right)\left(\frac{\hat{\theta}_p}{\theta_p} - 1\right) \xrightarrow{d} \begin{cases} \Theta + r\Gamma, & if \quad r \leq 1, \\ \frac{1}{r}\Theta + \Gamma, & if \quad r > 1, \end{cases}$$

where $\mathrm{Var}(\Theta) = (\gamma_1^2 - 1) - b^2 \int_0^\infty R(s, 1)ds^{-2\gamma_1}$, $\mathrm{Var}(\Gamma) = \gamma_1^2$ and
$\mathrm{Cov}(\Gamma, \Theta) = \gamma_1(1 - \gamma_1 + b)R(1, 1) - \gamma_1 \int_0^1 ((1 - \gamma_1) + bs^{-\gamma_1}(1 - \gamma_1 - \gamma_1 \ln s)) R(s, 1)s^{-1}ds$
with $b = \left(\int_0^\infty R(s, 1)ds^{-\gamma_1}\right)^{-1}$.

2.2 X Real

In this section, X takes values in $\mathbb{R}$, that is, we do not restrict X to be positive. Define $X^+ = \max(X, 0)$ and $X^- = X - X^+$. Besides the conditions of Theorem 1, we require two more conditions:

(e) $E|X^-|^{1/\gamma_1} < \infty$;

(f) As $n \to \infty$, $k = o\left(p^{2\tau(1-\gamma_1)}\right)$.

It can be shown that condition (e), together with (a), ensure that $\theta_p \sim E(X^+ | Y > U_2(1/p))$, as $p \downarrow 0$. Therefore, we estimate θ_p with

$$\hat{\theta}_p = \left(\frac{k}{np}\right)^{\hat{\gamma}_1} \frac{1}{k} \sum_{i=1}^n X_i I(X_i > 0, Y_i > Y_{n-k,n}), \tag{7}$$

with $\hat{\gamma}_1$ as in Section 2.1. Observe that when X is positive, this definition coincides with that in (6). As stated in the following theorem, the asymptotic behavior of the estimator remains the same as that for positive X.

Theorem 2. *Under the conditions of Theorem 1 and conditions (e) and (f), as $n \to \infty$,*

$$\min\left(\sqrt{k}, \frac{\sqrt{k_1}}{\log d_n}\right)\left(\frac{\hat{\theta}_p}{\theta_p} - 1\right) \xrightarrow{d} \begin{cases} \Theta + r\Gamma, & if \quad r \leq 1; \\ \frac{1}{r}\Theta + \Gamma, & if \quad r > 1, \end{cases}$$

where r, Θ and Γ are defined as in Theorem 1.

3 Simulation Study

In this section, a simulation and comparison study is implemented to investigate the finite sample performance of our estimator. We generate data from three bivariate distributions.

- A transformed Cauchy distribution on $(0,\infty)^2$ defined as

$$(X,Y) = \left(|Z_1|^{2/5}, |Z_2| \right),$$

where (Z_1, Z_2) is a standard Cauchy distribution on $\mathbb{R}^2$ with density $\frac{1}{2\pi}(1 + x^2 + y^2)^{-3/2}$. It follows that $\gamma_1 = 2/5$ and $R(x,y) = x + y - \sqrt{x^2 + y^2}$, $x, y \geq 0$. It can be shown that this distribution satisfies conditions (a) and (b) with $\tau = -1$, $\beta = 2$, and $\rho_1 = -2$. We shall refer to this distribution as "transformed Cauchy distribution (1)".

- Student-t_3 distribution on $(0,\infty)^2$ with density

$$f(x,y) = \frac{2}{\pi} \left(1 + \frac{x^2 + y^2}{3} \right)^{-5/2}, \quad x, y > 0.$$

We have $\gamma_1 = 1/3$, $R(x,y) = x+y - \frac{x^{4/3} + \frac{1}{2}x^{2/3}y^{2/3} + y^{4/3}}{\sqrt{x^{2/3} + y^{2/3}}}$, $\tau = -1/3$, $\beta = 4/3$ and $\rho_1 = -2/3$.

- A transformed Cauchy distribution on the whole $\mathbb{R}^2$ defined as

$$(X,Y) = \left(Z_1^{2/5} I(Z_1 \geq 0) + Z_1^{1/5} I(Z_1 < 0), Z_2 I(Z_1 \geq 0) + Z_2^{1/3} I(Z_1 < 0) \right).$$

We have $\gamma_1 = 2/5$, $R(x,y) = x/2 + y - \sqrt{x^2/4 + y}$, $\tau = -1$, $\beta = 2$, and $\rho_1 = -2$. We shall refer to this distribution as "transformed Cauchy distribution (2)".

We draw 500 samples from each distribution with sample sizes $n = 2{,}000$ and $n = 5{,}000$. Based on each sample, we estimate θ_p for $p = 1/500$, $1/5{,}000$ or $1/10{,}000$.

Besides the estimator given by (7), we construct two other estimators. Firstly, for $np \geq 1$, an empirical counterpart of θ_p, given by

$$\hat{\theta}_{emp} = \frac{1}{\lfloor np \rfloor} \sum_{i=1}^{n} X_i I(Y_i > Y_{n-\lfloor np \rfloor, n}), \tag{8}$$

is studied, where $\lfloor \cdot \rfloor$ denotes the integer part. Secondly, exploiting the relation in Proposition 1 and using the empirical estimator of R given by $\widehat{R}(x,y) = \frac{1}{k} \sum_{i=1}^{n} I(X_i > X_{n-\lfloor kx \rfloor, n}, Y_i > Y_{n-\lfloor ky \rfloor, n})$ and the Weissman (1978) estimator of $U_1(1/p)$ given by $\widehat{U}_1(1/p) = d_n^{\hat{\gamma}_1} X_{n-k,n}$, we define an alternative EVT estimator as

$$\begin{aligned}
\bar{\theta}_p &= -\widehat{U}_1(1/p) \int_0^{\infty} \widehat{R}(x,1) dx^{-\hat{\gamma}_1} \\
&= d_n^{\hat{\gamma}_1} X_{n-k,n} \frac{1}{k} \sum_{i=1}^{n} I(Y_i > Y_{n-k,n}) \left(\frac{n - \text{rank}(X_i) + 1}{k} \right)^{-\hat{\gamma}_1}.
\end{aligned} \tag{9}$$

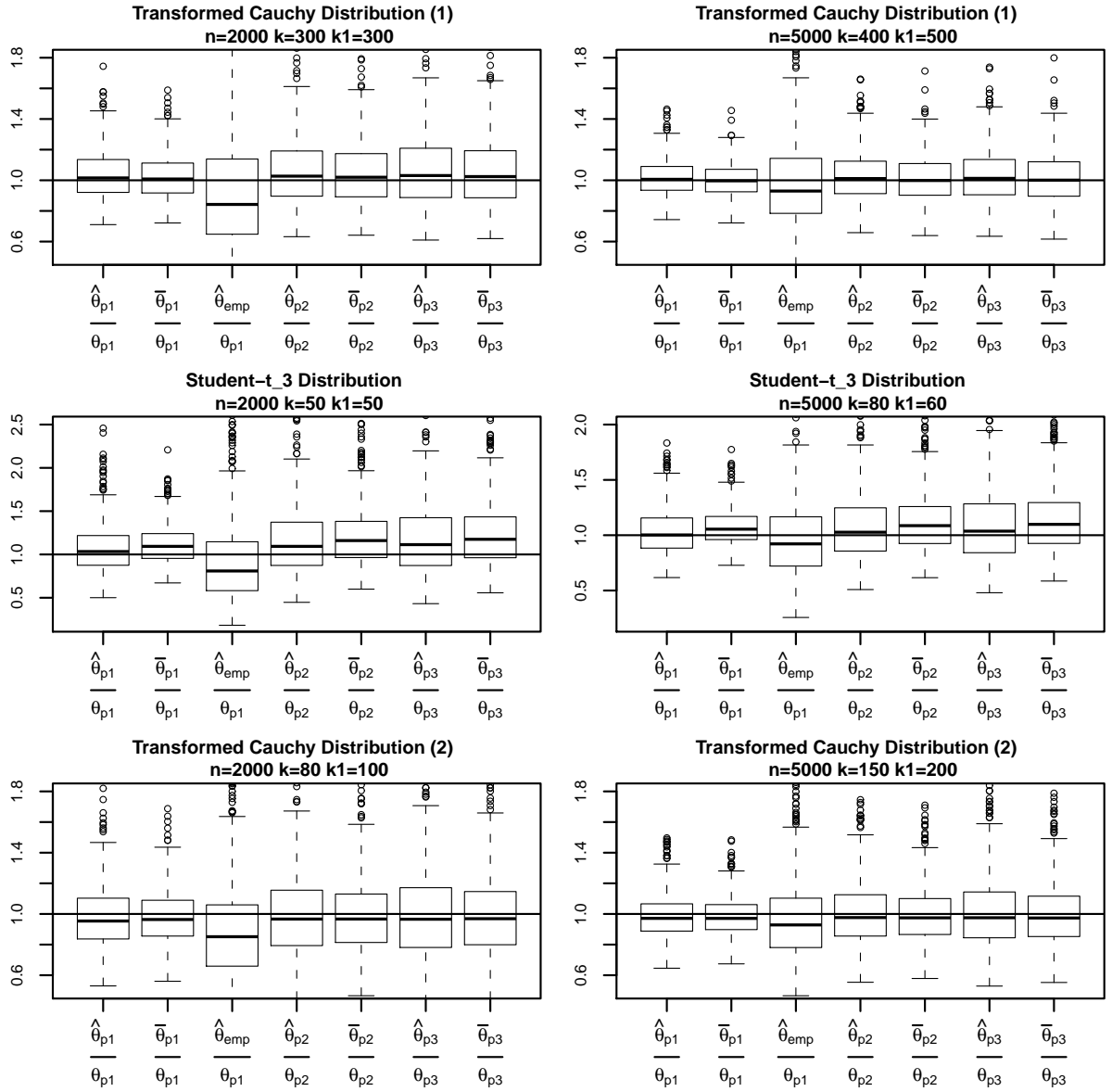Figure 1: Boxplots on ratios of estimates and true values. Each plot is based on 500 samples with sample size $n =$ 2,000 or 5,000 from the transformed Cauchy distributions (1), (2) or Student-t_3 distribution. The estimators are $\hat{\theta}_p$ of (7), $\bar{\theta}_p$ of (9) and $\hat{\theta}_{emp}$ of (8); $p = 1/500$ ($p1$), $1/5{,}000$ ($p2$) and $1/10{,}000$ ($p3$).

The comparison of the three estimators is shown in Figure 1, where we present boxplots of the ratios of the estimates and the true values. For all three distributions, the empirical estimator

underestimates the MES and is consistently outperformed by the EVT estimators. Additionally, it is not applicable for $p < 1/n$. The two EVT estimators, $\bar{\theta}_p$ and $\hat{\theta}_p$, both perform well. Their behavior is similar and remains stable when p changes from $1/500$ to $1/10{,}000$. The results for the transformed Cauchy distribution (1) are the best among the three distributions, as the medians of the ratios are closest to one and the variations are smallest.

Next, we investigate the normality of the estimator, $\hat{\theta}_p$, with $p = 1/n$. For $r < \infty$, the asymptotic normality of $\frac{\hat{\theta}_p}{\theta_p}$ in Theorem 1 can be expressed as $\sqrt{k}\left(\frac{\hat{\theta}_p}{\theta_p} - 1\right) \xrightarrow{d} \Theta + r\Gamma$, or equivalently,

$$\sqrt{k}\log\frac{\hat{\theta}_p}{\theta_p} \xrightarrow{d} \Theta + r\Gamma.$$

Notice that the limit distribution is a centered normal distribution. Write $\sigma_p^2 = \frac{1}{k}\operatorname{Var}(\Theta + r\Gamma)$ with $r = \frac{\sqrt{k}\log\frac{k}{np}}{\sqrt{k_1}}$. We compare the distribution of $\log\frac{\hat{\theta}_p}{\theta_p}$, with the limit distribution $N(0, \sigma_p^2)$. Table 1 reports the standardized mean of $\log\frac{\hat{\theta}_p}{\theta_p}$, i.e., the average value divided by σ_p, and between

Table 1: Standardized mean and standard deviation of $\log\frac{\hat{\theta}_p}{\theta_p}$

	$n = 2{,}000$	$n = 5{,}000$
	$p = 1/2{,}000$	$p = 1/5{,}000$
Transformed Cauchy distribution (1)	0.152 (1.027)	0.107 (1.054)
Student-t_3 distribution	0.232 (0.929)	0.148 (0.964)
Transformed Cauchy distribution (2)	-0.147 (1.002)	-0.070 (1.002)

The numbers are the standardized mean of $\log\frac{\hat{\theta}_p}{\theta_p}$ and between brackets, the ratio of the standard deviation and σ_p, based on 500 estimates with $n = 2{,}000$ or $5{,}000$ and $p = 1/n$.

brackets, the ratio of the "sample" standard deviation and σ_p. As indicated by the numbers, the mean and standard deviation of $\log\frac{\hat{\theta}_p}{\theta_p}$ are both close to that of the limit distribution.

After the numerical assessment on the parameters, we illustrate the normality of $\log\frac{\hat{\theta}_p}{\theta_p}$. Figure 2 shows the densities of the $N(0, \sigma_p^2)$-distribution and the histograms of $\log\frac{\hat{\theta}_p}{\theta_p}$, based on 500 estimates. The normality of the estimates is supported by the large overlap between the histograms and the areas under the density curves. Hence we conclude that the limit theorem provides an adequate approximation for finite sample sizes.
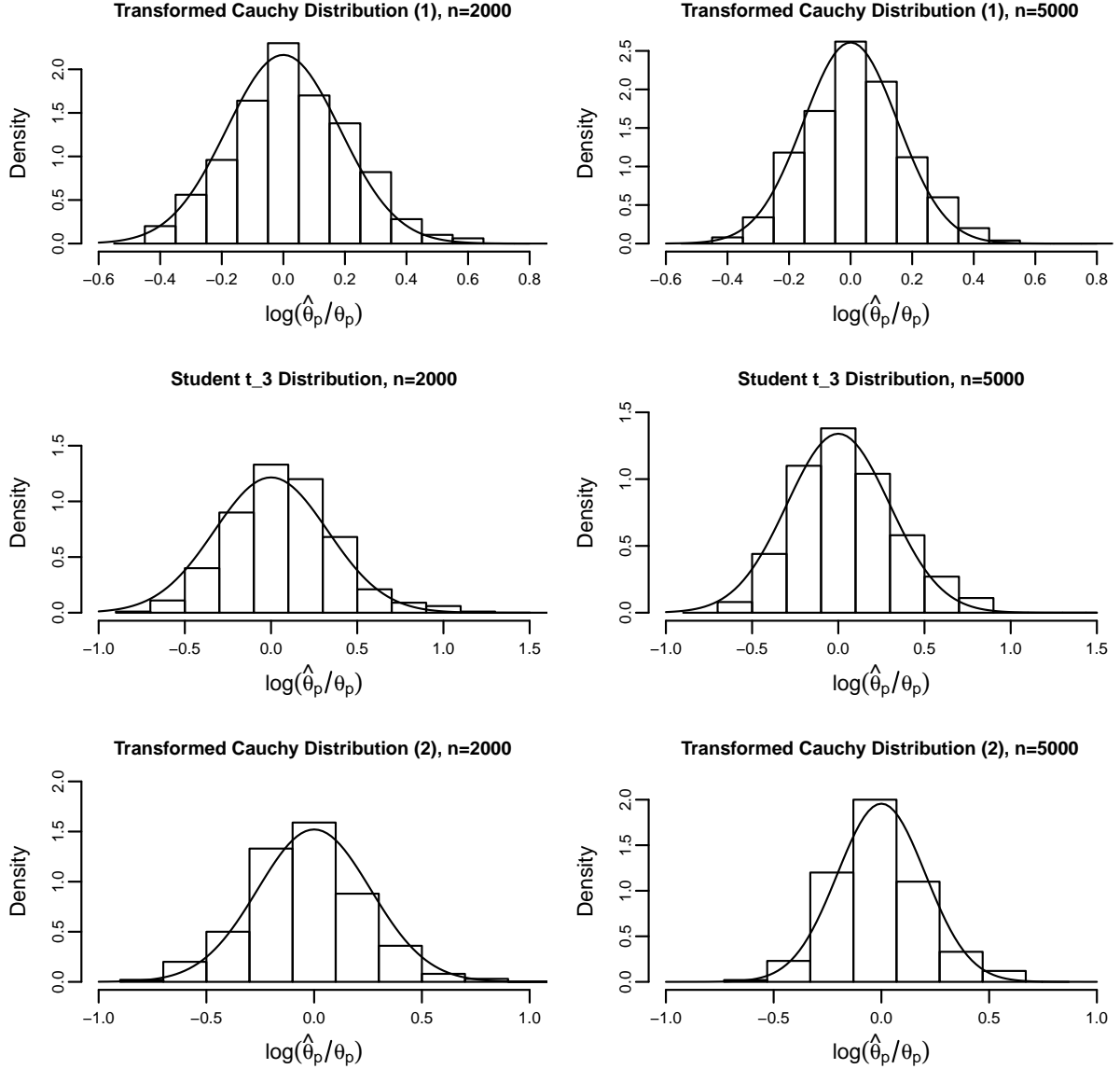
Figure 2: Histograms of $\log \frac{\hat{\theta}_p}{\theta_p}$ for $p = 1/n$, based on 500 samples with sample size $n = 2,000$ or 5,000 from the transformed Cauchy distributions (1), (2) or Student-t_3 distribution. The choices of k and k_1 are the same as in Figure 1. The curves are the densities of the $N(0, \sigma_p^2)$-distribution.

4 Application

In this section, we apply our estimation method to estimate the MES for some financial institutions. We consider three large investment banks in the U.S., namely, Goldman Sachs (GS),

11

Morgan Stanley (MS) and T. Rowe Price (TROW), all of which have a market capitalization greater than 5 billion USD as of the end of June 2007. The dataset consists of the loss returns (i.e., minus log returns) on their equity prices at a daily frequency from July 3, 2000 to June 30, 2010.[2] Moreover, for the same time period, we extract daily loss returns of a value weighted market index aggregating three markets: NYSE, AMES and Nasdaq. We use our method to estimate the MES, $E(X \mid Y > U_2(1/p))$, where X and Y refer to the daily loss returns of a bank equity and the market index, respectively and $p = 1/n = 1/2513$, that corresponds to a once per decade systemic event.
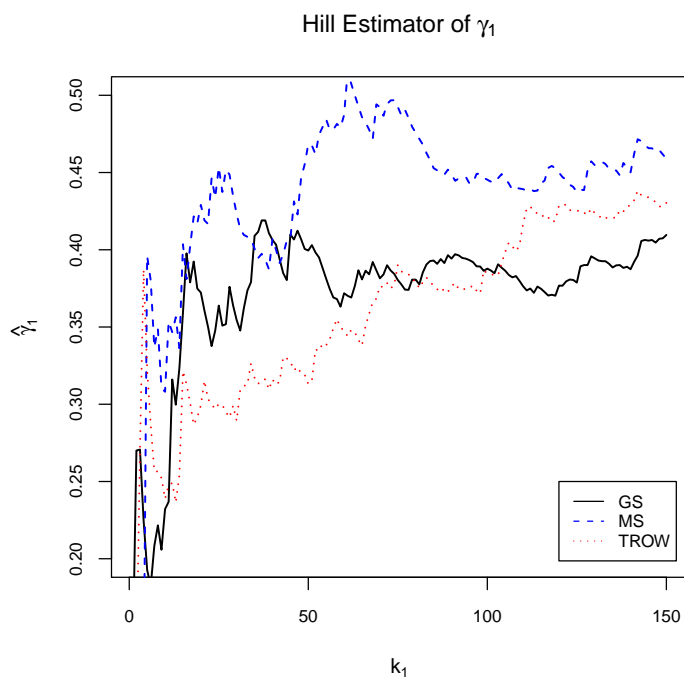

Figure 3: The Hill estimates of the extreme value indices of the daily loss returns on the three equities.

Since X may take negative values (i.e. positive returns of the equities of the banks), it is necessary to apply the estimator for the general case as defined in (7). For that purpose, we first verify two of the conditions required for the procedure. First of all, the assumption that $\gamma_1 < 1/2$

is confirmed by the plot of the Hill estimates in Figure 3. Secondly, since the estimation relies on the approximation of $\theta_p \sim E(X^+ \,|\, Y > U_2(1/p))$, it is important to check that high values of Y do not coincide with negative values of X, generally. Intuitive empirical evidence for this is presented in Figure 4. It plots the loss returns of the equity prices against the market index. The horizontal lines indicate the 50-th largest loss of the index. As one can see, from the upper parts of the plots, the largest 50 losses of the index are mostly associated with losses $(X > 0)$.
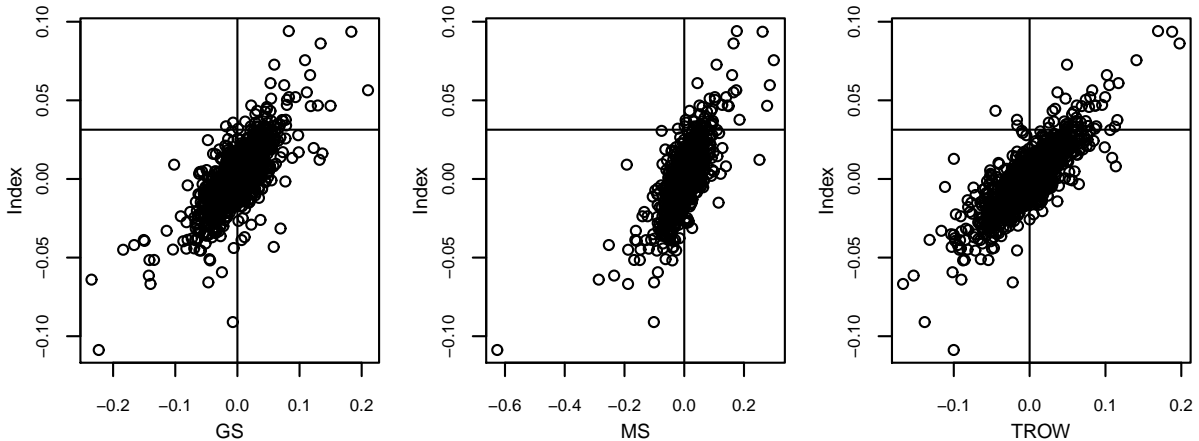

Figure 4: The points are the daily loss returns of the three equity prices and the market index. The horizontal lines indicate the 50-th largest loss of the market index. The vertical lines, at 0, distinguish the occurrence of losses and profits.

Hence, we can apply our method to obtain the estimates of γ_1 and $MES(p) = \theta_p$ for the

Table 2: MES of the three investment banks

Bank	$\hat{\gamma}_1$	$\hat{\theta}_p$
Goldman Sachs (GS)	0.386	0.301
Morgan Stanley (MS)	0.473	0.593
T. Rowe Price (TROW)	0.379	0.312

Here $\hat{\gamma}_1$ is computed by taking the average of the Hill estimates for $k_1 \in [70, 90]$. $\hat{\theta}_p$ is given in (7) with $n = 2513$, $k = 50$, and $p = 1/n = 1/2513$.

three banks, see Table 2. It follows that in case of a once per decade market crisis, we estimate

that on average the equity prices of Goldman Sachs and T. Rowe Price drop about 25% and Morgan Stanley falls even about 45% on that day.

5 Proofs

Proof of Proposition 1 Recall that for a non-negative random variable Z,

$$E(Z) = \int_0^\infty P(Z > x)dx.$$

Hence,

$$
\begin{aligned}
\frac{\theta_p}{U_1(1/p)} &= \int_0^\infty \frac{1}{p} P(X > x, Y > U_2(1/p)) \frac{dx}{U_1(1/p)} \\
&= \int_0^\infty \frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p))dx.
\end{aligned}
\tag{10}
$$

The limit relations (1) and (2) implies that

$$\lim_{p \downarrow 0} \frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p)) = R(x^{-1/\gamma_1}, 1).$$

Hence, we only have to prove that the integral in (10) and the limit procedure $p \to 0$ can be interchanged. This is ensured by the dominated convergence theorem as follows. Notice that for $x \geq 0$,

$$\frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p)) \leq \min\left(1, \frac{1}{p}\left(1 - F_1(U_1(1/p)x)\right)\right).$$

For $0 < \varepsilon < 1/\gamma_1 - 1$, there exists $p(\varepsilon)$ (see Proposition B.1.9.5 in de Haan and Ferreira (2006)) such that for all $p < p(\varepsilon)$ and $x > 1$,

$$\frac{1}{p}\left(1 - F_1(U_1(1/p)x)\right) \leq 2x^{-1/\gamma_1 + \varepsilon}.$$

Write

$$
h(x) = \begin{cases}
1, & 0 \leq x \leq 1; \\
2x^{-1/\gamma_1 + \varepsilon}, & x > 1.
\end{cases}
$$

Then h is integrable and $\frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p)) \leq h(x)$ on $[0, \infty)$ for $p < p(\varepsilon)$. Hence we can apply the dominated convergence theorem to complete the proof of the proposition. $\quad\square$

Next, we prove Theorem 1. The general idea of the proof is described as follows. It is clear that the asymptotic behavior of $\hat{\theta}_p$ results from that of $\hat{\gamma}_1$ and $\hat{\theta}_{\frac{k}{n}}$. The asymptotic normality of

14

$\hat{\gamma}_1$ is well-known, see, e.g., de Haan and Ferreira (2006). To prove the asymptotic normality of $\hat{\theta}_{\frac{k}{n}}$, write

$$\hat{\theta}_{\frac{k}{n}} = \frac{1}{k}\sum_{i=1}^{n} X_i I(Y_i > U_2(n/(ke_n))),$$

where $e_n = \frac{n}{k}(1 - F_2(Y_{n-k,n})) \xrightarrow{P} 1$, as $n \to \infty$. Hence, with denoting $\tilde{\theta}_{\frac{ky}{n}} := \frac{1}{ky}\sum_{i=1}^{n} X_i I(Y_i > U_2(n/(ky)))$, we first investigate the asymptotic behavior of $\tilde{\theta}_{\frac{ky}{n}}$ for $y \in [1/2, 2]$. Then, by applying the result for $y = e_n$ and considering the asymptotic behavior of e_n, we obtain the asymptotic normality of $\hat{\theta}_{\frac{k}{n}}$. Lastly, together with the asymptotic normality of $\hat{\gamma}_1$, we prove that of $\hat{\theta}_p$.

To obtain the asymptotic behavior of $\tilde{\theta}_{\frac{ky}{n}}$, we introduce some new notation and auxiliary lemmas. Write $R_n(x,y) := \frac{n}{k} P(1 - F_1(X) < kx/n, 1 - F_2(Y) < ky/n)$. A pseudo non-parametric estimator of R_n is given as $T_n(x,y) := \frac{1}{k}\sum_{i=1}^{n} I(1 - F_1(X_i) < kx/n, 1 - F_2(Y_i) < ky/n)$.[3] The following lemma shows the asymptotic behavior of the pseudo estimator. The limit process is characterized by the aforementioned W_R process. For convenient presentation, all the limit processes involved in the lemma are defined on the same probability space, via the Skorohod construction. However, they are only equal in distribution to the original ones. The proof of the lemma is analogous to that of Proposition 3.1 in Einmahl et al. (2006) and is thus omitted.

Lemma 1. *Suppose* (1) *holds. For any* $\eta \in [0, 1/2)$ *and* T *positive, with probability 1,*

$$\sup_{x,y\in(0,T]} \left| \frac{\sqrt{k}(T_n(x,y) - R_n(x,y)) - W_R(x,y)}{x^\eta} \right| \to 0,$$

$$\sup_{x\in(0,T]} \left| \frac{\sqrt{k}(T_n(x,\infty) - x) - W_R(x,\infty)}{x^\eta} \right| \to 0,$$

$$\sup_{y\in(0,T]} \left| \frac{\sqrt{k}(T_n(\infty,y) - y) - W_R(\infty,y)}{y^\eta} \right| \to 0.$$

The following lemma shows the boundedness of the W_R process with proper weighing function. It follows from, for instance, a modification of Example 1.8 in Alexander (1986) or that of Lemma 3.2 in Einmahl et al. (2006).

Lemma 2. *For any* $T > 0$ *and* $\eta \in [0, 1/2)$, *with probability 1,*

$$\sup_{0<x\leq T, 0<y<\infty} \frac{|W_R(x,y)|}{x^\eta} < \infty \quad \text{and} \quad \sup_{0<x<\infty, 0<y<T} \frac{|W_R(x,y)|}{y^\eta} < \infty.$$

---

[3]It is called "pseudo" estimator because the marginal distribution functions are unknown.

15

Next, denote $s_n(x) = \frac{n}{k}(1 - F_1(U_1(n/k)x^{-\gamma_1}))$ for $x > 0$. From the regular variation condition (2), we get that $s_n(x) \to x$ as $n \to \infty$. The following lemma shows that when handling proper integrals, $s_n(x)$ can be substituted by x in the limit.

Lemma 3. *Suppose (2) holds. Denote g as a bounded and continuous function on $[0, S_0) \times [a, b]$ with $0 < S_0 \leq \infty$ and $0 \leq a < b < \infty$. Moreover, suppose there exist $\eta_1 > \gamma_1$ and $m > 0$ such that*

$$\sup_{0 < x \leq S_0, \ a \leq y \leq b} \frac{|g(x, y)|}{x^{\eta_1}} \leq m.$$

If $S_0 < +\infty$, we further require that $0 < S < S_0$. Then,

$$\lim_{n \to \infty} \sup_{a \leq y \leq b} \left| \int_0^S g(s_n(x), y) - g(x, y) dx^{-\gamma_1} \right| = 0. \tag{11}$$

Furthermore, suppose $|g(x_1, y) - g(x_2, y)| \leq |x_1 - x_2|$ holds for all $0 \leq x_1, x_2 < S_0$ and $a \leq y \leq b$. Under conditions (b) and (d), we have that

$$\lim_{n \to \infty} \sup_{a \leq y \leq b} \sqrt{k} \left| \int_0^S g(s_n(x), y) - g(x, y) dx^{-\gamma_1} \right| = 0. \tag{12}$$

Proof of Lemma 3 We prove (11) and (12) for $S = S_0 = \infty$. The proof for $0 < S < S_0 < +\infty$ is similar but simpler. For any $0 < \varepsilon < 1$, denote $T(\varepsilon) = \varepsilon^{-1/\gamma_1}$. It follows from (2) and Proposition B.1.10 of de Haan and Ferreira (2006) that

$$\lim_{n \to \infty} \sup_{0 < x \leq 1} \frac{s_n(x)}{x^{\frac{\gamma_1 + \eta_1}{2\eta_1}}} = 1,$$

and

$$\lim_{n \to \infty} \sup_{0 < x \leq T(\varepsilon)} |s_n(x) - x| = 0.$$

With $\delta(\varepsilon) = \varepsilon^{1/(\eta_1 - \gamma_1)}$, we have that

$$\sup_{a \leq y \leq b} \left| \int_0^\infty (g(s_n(x), y) - g(x, y)) \, dx^{-\gamma_1} \right|$$

$$\leq \sup_{a \leq y \leq b} \left(\left| \int_0^\delta (g(s_n(x), y) - g(x, y)) \, dx^{-\gamma_1} \right| + \left| \int_\delta^T (g(s_n(x), y) - g(x, y)) \, dx^{-\gamma_1} \right| \right.$$

$$\left. + \left| \int_T^\infty (g(s_n(x), y) - g(x, y)) \, dx^{-\gamma_1} \right| \right)$$

$$\leq -m \int_0^\delta \left(x^{\frac{\gamma_1 + \eta_1}{2}} + x^{\eta_1} \right) dx^{-\gamma_1} + \delta^{-\gamma_1} \sup_{\substack{\delta \leq x \leq T \\ a \leq y \leq b}} |g(s_n(x), y) - g(x, y)| + 2\varepsilon \sup_{\substack{0 \leq x < \infty \\ a \leq y \leq b}} |g(x, y)|$$

16

$$\leq c_1 \varepsilon^{1/2} + \delta^{-\gamma_1} \sup_{\substack{\delta \leq x \leq T \\ a \leq y \leq b}} |g(s_n(x), y) - g(x, y)| + 2\varepsilon \sup_{\substack{0 \leq x < \infty \\ a \leq y \leq b}} |g(x, y)|,$$

where c_1 is a finite constant. Hence, (11) follows from the uniform continuity of g on $[\delta, T] \times [a, b]$ and the boundedness of g on $[0, +\infty) \times [a, b]$.

Next we prove (12). Denote $\widetilde{T}_n = |A_1(n/k)|^{\frac{1}{\rho_1 - 1}}$. By the Lipschitz property of g,

$$\sup_{a \leq y \leq b} \left| \int_0^\infty (g(s_n(x), y) - g(x, y)) \, dx^{-\gamma_1} \right|$$

$$\leq \int_0^{\widetilde{T}_n} |s_n(x) - x| \, dx^{-\gamma_1} + 2 \sup_{\substack{0 \leq x < \infty \\ a \leq y \leq b}} |g(x, y)| \widetilde{T}_n^{-\gamma_1}. \tag{13}$$

It is thus necessary to prove that both terms in the right hand side of (13) are $o(1/\sqrt{k})$. For the second term, condition (d) implies that $\frac{\alpha}{2(1-\alpha)} < \frac{\gamma_1 \rho_1}{\rho_1 - 1}$. Thus for any $\varepsilon_0 \in \left(0, \frac{\gamma_1 \rho_1}{\rho_1 - 1} - \frac{\alpha}{2(1-\alpha)}\right)$, as $n \to \infty$, we have that

$$\sqrt{k} \left(\frac{n}{k}\right)^{\frac{\gamma_1 \rho_1}{1 - \rho_1} + \varepsilon_0} = O(n^{\frac{\gamma_1 \rho_1}{1 - \rho_1} + \varepsilon_0 - \alpha\left(\frac{\gamma_1 \rho_1}{1 - \rho_1} + \varepsilon_0 - \frac{1}{2}\right)}) \to 0,$$

which leads to

$$\sqrt{k} \widetilde{T}_n^{-\gamma_1} = \sqrt{k} |A_1(n/k)|^{\frac{\gamma_1}{1 - \rho_1}} \to 0. \tag{14}$$

For the first term, notice that for $x \in (0, \widetilde{T}_n]$ and $0 < \varepsilon_1 < \frac{\gamma_1}{1 - \rho_1}$, when n is large enough,

$$U_1(n/k)x^{-\gamma_1} \geq U_1(n/k)\widetilde{T}_n^{-\gamma_1} = U_1(n/k) |A_1(n/k)|^{\frac{\gamma_1}{1 - \rho_1}} \geq \left(\frac{n}{k}\right)^{\frac{\gamma_1}{1 - \rho_1} - \varepsilon_1},$$

which implies that $U_1(n/k)x^{-\gamma_1} \to +\infty$ as $n \to \infty$. Hence we can apply Theorems 2.3.9 and B.3.10 in de Haan and Ferreira (2006) to condition (b) and obtain that for sufficiently large n,

$$\left| \frac{s_n(x) - x}{A_1(n/k)} - x \frac{x^{-\rho_1} - 1}{\gamma_1 \rho_1} \right| \leq x^{1 - \rho_1} \max(x^{\varepsilon_0}, x^{-\varepsilon_0}).$$

Thus, we get that

$$\sqrt{k} \int_0^{\widetilde{T}_n} |s_n(x) - x| \, dx^{-\gamma_1}$$

$$\leq \sqrt{k} |A_1(n/k)| \int_0^{\widetilde{T}_n} \left(x \left| \frac{x^{-\rho_1} - 1}{\gamma_1 \rho_1} \right| + x^{1 - \rho_1} \max(x^{\varepsilon_0}, x^{-\varepsilon_0}) \right) dx^{-\gamma_1}$$

$$\leq c_2 \sqrt{k} |A_1(n/k)| \widetilde{T}_n^{1 - \rho_1 - \gamma_1 + \varepsilon_0}$$

$$= c_2 \sqrt{k} |A_1(n/k)|^{\frac{\gamma_1 - \varepsilon_0}{1 - \rho_1}} \leq c_3 \sqrt{k} \left(\frac{n}{k}\right)^{\frac{\rho_1 \gamma_1}{1 - \rho_1} + \varepsilon_0}, \tag{15}$$

17

with c_2 and c_3 some positive constants. Again, by condition (d), as $n \to \infty$, $c_3\sqrt{k}\left(\frac{n}{k}\right)^{\frac{\rho_1\gamma_1}{1-\rho_1}+\varepsilon_0} \to 0$. Hence, (12) is proved by combining (13), (14) and (15). $\qquad\square$

With those auxiliary lemmas, we obtain the asymptotic behavior of $\tilde{\theta}_{\frac{ky}{n}}$ as follows.

Proposition 2. *Suppose* (1) *and* (2) *hold with* $0 < \gamma_1 < 1/2$. *Then,*

$$\sup_{1/2 \leq y \leq 2}\left|\frac{\sqrt{k}}{U_1(n/k)}\left(\tilde{\theta}_{\frac{ky}{n}} - \theta_{\frac{ky}{n}}\right) + \frac{1}{y}\int_0^\infty W_R(s,y)ds^{-\gamma_1}\right| \xrightarrow{P} 0.$$

Proof of Proposition 2 Recall $s_n(x) = \frac{n}{k}\left(1 - F_1(U_1(n/k)x^{-\gamma_1})\right)$, $x > 0$. Similar to (10),

$$
\begin{aligned}
y\theta_{\frac{ky}{n}} &= \int_0^\infty \frac{n}{k}P(X > s, Y > U_2(n/(ky)))ds \\
&= \int_0^\infty \frac{n}{k}P(1 - F_1(X) < 1 - F_1(s), 1 - F_2(Y) < ky/n)ds \\
&= \int_0^\infty R_n\left(\frac{n}{k}(1 - F_1(s)), y\right)ds \\
&= -U_1(n/k)\int_0^\infty R_n(s_n(x), y)dx^{-\gamma_1}.
\end{aligned}
\tag{16}
$$

Similarly, $y\tilde{\theta}_{\frac{ky}{n}} = -U_1(n/k)\int_0^\infty T_n(s_n(x), y)dx^{-\gamma_1}$. For any $T > 0$, we have

$$
\begin{aligned}
&\sup_{1/2 \leq y \leq 2}\left|\frac{\sqrt{k}}{U_1(n/k)}(y\tilde{\theta}_{\frac{ky}{n}} - y\theta_{\frac{ky}{n}}) + \int_0^\infty W_R(x,y)dx^{-\gamma_1}\right| \\
&= \sup_{1/2 \leq y \leq 2}\left|\int_0^\infty W_R(x,y)dx^{-\gamma_1} - \int_0^\infty \sqrt{k}\left(T_n(s_n(x),y) - R_n(s_n(x),y)\right)dx^{-\gamma_1}\right| \\
&\leq \sup_{1/2 \leq y \leq 2}\left|\int_T^\infty W_R(x,y)dx^{-\gamma_1}\right| + \sup_{1/2 \leq y \leq 2}\left|\int_T^\infty \sqrt{k}\left(T_n(s_n(x),y) - R_n(s_n(x),y)\right)dx^{-\gamma_1}\right| \\
&\quad + \sup_{1/2 \leq y \leq 2}\left|\int_0^T \sqrt{k}\left(T_n(s_n(x),y) - R_n(s_n(x),y)\right) - W_R(x,y)dx^{-\gamma_1}\right| \\
&=: I_1(T) + I_{2,n}(T) + I_{3,n}(T).
\end{aligned}
$$

It suffices to prove that for any $\varepsilon > 0$, there exists $T_0 = T_0(\varepsilon)$ such that

$$P\left(I_1(T_0) > \varepsilon\right) < \varepsilon, \tag{17}$$

and $n_0 = n_0(T_0)$ such that for any $n > n_0$,

$$P\left(I_{2,n}(T_0) > \varepsilon\right) < \varepsilon; \tag{18}$$

$$P\left(I_{3,n}(T_0) > \varepsilon\right) < \varepsilon. \tag{19}$$

Firstly, for the term $I_1(T)$, by Lemma 2 with $\eta = 0$, there exists $T_1 = T_1(\varepsilon)$ such that

$$P\left(\sup_{0<x<\infty, 0\leq y\leq 2} |W_R(x,y)| > T_1^{\gamma_1}\varepsilon\right) < \varepsilon.$$

Then for any $T > T_1$,

$$P(I_1(T) > \varepsilon) \leq P\left(\sup_{x>T_1, 1/2\leq y\leq 2} |W_R(x,y)| > T_1^{\gamma_1}\varepsilon\right) < \varepsilon.$$

Thus (17) holds provided that $T_0 > T_1$.

Next we deal with the term $I_{2,n}(T)$. Let $\widetilde{P}$ be the probability measure defined by $(1 - F_1(X), 1 - F_2(Y))$ and $\widetilde{P}_n$ the empirical probability measure defined by $(1 - F_1(X_i), 1 - F_2(Y_i))_{1\leq i\leq n}$. We have

$$P(I_{2,n}(T) > \varepsilon) = P\left(\sup_{1/2\leq y\leq 2} \left|\int_T^\infty \sqrt{k}\,(T_n(s_n(x),y) - R_n(s_n(x),y))\,dx^{-\gamma_1}\right| > \varepsilon\right)$$

$$\leq P\left(\sup_{x>T, 1/2\leq y\leq 2} \left|\sqrt{k}\,(T_n(s_n(x),y) - R_n(s_n(x),y))\right| > \varepsilon T^{\gamma_1}\right)$$

$$= P\left(\sup_{x>T, 1/2\leq y\leq 2} \left|\sqrt{n}(\widetilde{P}_n - \widetilde{P})\left\{\left(0, \frac{ks_n(x)}{n}\right) \times \left(0, \frac{ky}{n}\right)\right\}\right| > \varepsilon T^{\gamma_1}\sqrt{k/n}\right)$$

$$=: p_2.$$

Define $S_n = \{[0,1] \times (0, 2k/n)\}$, then $\widetilde{P}(S_n) = 2k/n$. Now by Inequality 2.5 in Einmahl (1987), there exists a constant c and a function ψ with $\lim_{t\to 0}\psi(t) = 1$, such that

$$p_2 \leq c\exp\left(-\frac{\left(\varepsilon T^{\gamma_1}\sqrt{k/n}\right)^2}{4\widetilde{P}(S_n)}\psi\left(\frac{\varepsilon T^{\gamma_1}\sqrt{k/n}}{\sqrt{n}\widetilde{P}(S_n)}\right)\right)$$

$$= c\exp\left(-\frac{\varepsilon^2 T^{\gamma_1}}{8}\psi\left(\frac{\varepsilon T^{\gamma_1/2}}{2\sqrt{k}}\right)\right).$$

Choose $T_2(\varepsilon)$ such that $c\exp\left(-\frac{\varepsilon^2 T_2^{\gamma_1}}{16}\right) \leq \varepsilon$. Then, for any $T > T_2$, $c\exp\left(-\frac{\varepsilon^2 T^{\gamma_1}}{16}\right) \leq \varepsilon$. Furthermore, we can choose $n_1 = n_1(T)$ such that for $n > n_1$, $\psi\left(\frac{\varepsilon T^{\gamma_1/2}}{2\sqrt{k}}\right) > 1/2$. Therefore, for $T > T_2(\varepsilon)$ and $n > n_1(T)$, we have $p_2 < \varepsilon$, which leads to (18) provided that $T_0 > T_2$ and $n_0 > n_1$.

Lastly, we deal with $I_{3,n}(T)$. We have that

$$P(I_{3,n}(T) > \varepsilon)$$

$$\leq P\left(\sup_{1/2\leq y\leq 2}\left|\int_0^T \sqrt{k}\,(T_n(s_n(x),y)-R_n(s_n(x),y))-W_R(s_n(x),y)dx^{-\gamma_1}\right|>\varepsilon/2\right)$$

$$+P\left(\sup_{1/2\leq y\leq 2}\left|\int_0^T W_R(s_n(x),y)-W_R(x,y)dx^{-\gamma_1}\right|>\varepsilon/2\right)$$

$$=:p_{31}+p_{32}.$$

We first consider p_{31}. Notice that for any T, there exists $n_2=n_2(T)$ such that for all $n>n_2, s_n(T)<T+1$. Hence, for $n>n_2$ and any $\eta_0\in(\gamma_1,1/2)$,

$$p_{31}\leq P\left(\sup_{\substack{0<s\leq T+1\\1/2\leq y\leq 2}}\left|\frac{\sqrt{k}\,(T_n(s,y)-R_n(s,y))-W_R(s,y)}{s^{\eta_0}}\right|\left|\int_0^T (s_n(x))^{\eta_0}dx^{-\gamma_1}\right|>\varepsilon/2\right)$$

Notice that by (11), as $n\to\infty$, $\left|\int_0^T (s_n(x))^{\eta_0}dx^{-\gamma_1}\right|\to\frac{\gamma_1}{\eta_0-\gamma_1}T^{\eta_0-\gamma_1}$. Together with Lemma 1, there exists $n_3(T)>n_2(T)$ such that for $n>n_3(T)$, $p_{31}<\varepsilon/2$.

Then, we consider p_{32}. Applying Lemma 2, with the aforementioned $\eta_0\in(\gamma_1,1/2)$, there exists $\lambda_0=\lambda(\eta_0,\varepsilon)$ such that

$$P\left(\sup_{0<x<\infty,1/2\leq y\leq 2}\frac{|W_R(x,y)|}{x^{\eta_0}}\geq\lambda_0\right)\leq\varepsilon/3. \tag{20}$$

Moreover, $W_R(x,y)$ is continuous on $(0,\infty)\times[1/2,2]$, see Corollary 1.11 in Adler (1990). Hence applying (20) and (11) with $g=W_R$, $S=T$ and $S_0=T+1$, we have that there exists a $n_4=n_4(T)$ such that for $n>n_4$, $p_{32}<\varepsilon/2$. Thus, (19) holds for any T_0 and $n_0>\max(n_3(T_0),n_4(T_0))$.

To summarize, choose $T_0=T_0(\varepsilon)>\max(T_1,T_2)$, and define $n_0(T_0)=\max_{1\leq j\leq 4}n_j(T_0)$. We get that for the chosen T_0 and any $n>n_0$, the three inequalities (17)–(19) hold, which completes the proof of the proposition. □

Next, we proceed with the second step: establishing the asymptotic normality of $\hat{\theta}_{\frac{k}{n}}$.

Proposition 3. *Under the condition of Theorem 1, we have*

$$\sqrt{k}\left(\frac{\hat{\theta}_{\frac{k}{n}}}{\theta_{\frac{k}{n}}}-1\right)\xrightarrow{d}\Theta.$$

Proof of Proposition 3 Observe that $\lim_{n\to\infty}\frac{\theta_{\frac{k}{n}}}{U_1(n/k)}\to\int_0^\infty R(s^{-1/\gamma_1},1)ds$. Therefore it is sufficient to show that

$$\frac{\sqrt{k}}{U_1(n/k)}\left(\hat{\theta}_{\frac{k}{n}}-\theta_{\frac{k}{n}}\right)\xrightarrow{P}\Theta\int_0^\infty R(s^{-1/\gamma_1},1)ds.$$

20

Recall $e_n = \frac{n}{k}(1 - F_2(Y_{n-k,n}))$. Hence, with probability 1, $\hat{\theta}_{\frac{k}{n}} = e_n \tilde{\theta}_{\frac{ke_n}{n}}$, we thus have that

$$\frac{\sqrt{k}}{U_1(n/k)}\left(e_n \tilde{\theta}_{\frac{ke_n}{n}} - \theta_{\frac{k}{n}}\right) - \Theta \int_0^\infty R(s^{-1/\gamma_1}, 1)ds$$

$$= \left(\frac{\sqrt{k}}{U_1(n/k)}\left(e_n \tilde{\theta}_{\frac{ke_n}{n}} - e_n \theta_{\frac{ke_n}{n}}\right) + \int_0^\infty W_R(s,1)ds^{-\gamma_1}\right)$$

$$+ \left(\frac{\sqrt{k}}{U_1(n/k)}\left(e_n \theta_{\frac{ke_n}{n}} - \theta_{\frac{k}{n}}\right) - W_R(\infty, 1)(\gamma_1 - 1)\int_0^\infty R(s^{-1/\gamma_1}, 1)ds\right)$$

$$=: J_1 + J_2.$$

We prove that both J_1 and J_2 converge to zero in probability as $n \to \infty$.

Firstly, we deal with J_1. By Lemma 1 and $T_n(\infty, e_n) = 1$, we get that

$$\sqrt{k}(e_n - 1) \xrightarrow{P} -W_R(\infty, 1), \tag{21}$$

which implies that

$$\lim_{n \to \infty} P(|e_n - 1| > k^{-1/4}) = 0.$$

Hence, with probability tending to 1,

$$|J_1| \leq \sup_{|y-1|<k^{-1/4}} \left|\frac{\sqrt{k}}{U_1(n/k)}\left(y\tilde{\theta}_{\frac{ky}{n}} - y\theta_{\frac{ky}{n}}\right) + \int_0^\infty W_R(s,y)ds^{-\gamma_1}\right|$$

$$+ \sup_{|y-1|<k^{-1/4}} \left|\int_0^\infty W_R(s,y) - W_R(s,1)ds^{-\gamma_1}\right|.$$

The first part converges to zero in probability by Proposition 2. For the second part, notice that for any $\varepsilon > 0$, $0 < \delta < 1$ and $\eta \in (\gamma_1, 1/2)$,

$$P\left(\sup_{|y-1|<k^{-1/4}} \left|\int_0^\infty W_R(s,y) - W_R(s,1)ds^{-\gamma_1}\right| > \varepsilon\right)$$

$$\leq P\left(\sup_{|y-1|<k^{-1/4}} \left|\int_0^\delta W_R(s,y) - W_R(s,1)ds^{-\gamma_1}\right| > \varepsilon/2\right)$$

$$+ P\left(\sup_{|y-1|<k^{-1/4}} \left|\int_\delta^\infty W_R(s,y) - W_R(s,1)ds^{-\gamma_1}\right| > \varepsilon/2\right)$$

$$\leq P\left(\sup_{0<s\leq 1, 1/2\leq y\leq 2} \frac{|W_R(s,y)|}{s^\eta} > \frac{\varepsilon(\eta - \gamma_1)}{4\gamma_1}\delta^{\gamma_1 - \eta}\right)$$

$$+ P\left(\sup_{s>0, |y-1|<k^{-1/4}} |W_R(s,y) - W_R(s,1)|\delta^{-\gamma_1} > \varepsilon/2\right).$$

$$=: p_{11} + p_{12}.$$

For any fixed ε, Lemma 2 ensures that there exists a positive $\delta(\varepsilon)$ such that for all $\delta < \delta(\varepsilon)$, we have that $p_{11} < \varepsilon$. Then, for any fixed δ, there must exists an positive integer $n(\delta)$ such that for $n > n(\delta)$ we can achieve that $p_{12} < \varepsilon$, because we have that as $n \to \infty$,

$$\sup_{s>0,|y-1|<k^{-1/4}} |W_R(s,y) - W_R(s,1)| \overset{a.s}{\to} 0,$$

see Corollary 1.11 in Adler (1990). Hence we proved that $J_1 \overset{P}{\to} 0$ as $n \to \infty$.

Next we deal with J_2. We first prove a non-stochastic limit relation: as $n \to \infty$,

$$\sup_{1/2 \leq y \leq 2} \sqrt{k} \left| \int_0^\infty R_n(s_n(x), y) - R(x, y)dx^{-\gamma_1} \right| \to 0. \tag{22}$$

Condition (a) implies that as $n \to \infty$,

$$\sup_{\substack{0<x<\infty \\ 1/2 \leq y \leq 2}} \frac{|R_n(x,y) - R(x,y)|}{x^\beta \wedge 1} = O\left(\left(\frac{n}{k} \right)^\tau \right).$$

Hence, as $n \to \infty$,

$$\sup_{1/2 \leq y \leq 2} \sqrt{k} \left| \int_0^\infty R_n(s_n(x), y) - R(s_n(x), y)dx^{-\gamma_1} \right|$$

$$\leq \sqrt{k} \sup_{\substack{0<x<\infty \\ 1/2 \leq y \leq 2}} \frac{|R_n(x,y) - R(x,y)|}{x^\beta \wedge 1} \left| \int_0^\infty (s_n(x))^\beta \wedge 1 dx^{-\gamma_1} \right|$$

$$= O\left(\sqrt{k} \left(\frac{n}{k} \right)^\tau \right) \left(-\int_0^{1/2} (s_n(x))^\beta dx^{-\gamma_1} - \int_{1/2}^\infty 1 dx^{-\gamma_1} \right) \to 0.$$

The last step follows from the following two facts. Firstly, condition (d) ensures that $k = O(n^\alpha)$ with $\alpha < \frac{2\tau}{2\tau-1}$. Secondly, we have that

$$\lim_{n \to \infty} -\int_0^{1/2} (s_n(x))^\beta dx^{-\gamma_1} = -\int_0^{1/2} x^\beta dx^{-\gamma_1} < \infty,$$

which is a consequence of (11).

To complete the proof of relation (22), it is still necessary to show that as $n \to \infty$,

$$\sup_{1/2 \leq y \leq 2} \sqrt{k} \left| \int_0^\infty R(s_n(x), y) - R(x, y)dx^{-\gamma_1} \right| \to 0.$$

This is achieved by applying (12) to the R function which satisfies the Lipschitz condition: $|R(x_1, y) - R(x_2, y)| \leq |x_1 - x_2|$, for $x_1, x_2, y \geq 0$. Hence, we proved the relation (22).

22

Combining (16) and (22), we obtain that

$$\frac{\theta_{\frac{k}{n}}}{U_1(n/k)} = -\int_0^\infty R(s_n(x),1)dx^{-\gamma_1} = -\int_0^\infty R(x,1)dx^{-\gamma_1} + o\left(\frac{1}{\sqrt{k}}\right), \tag{23}$$

and

$$\frac{e_n\theta_{\frac{ke_n}{n}}}{U_1(n/k)} = -\int_0^\infty R_n(s_n(x),e_n)dx^{-\gamma_1} = -\int_0^\infty R(x,e_n)dx^{-\gamma_1} + o_P\left(\frac{1}{\sqrt{k}}\right).$$

From the homogeneity of the R function, for $y > 0$, we have that

$$\int_0^\infty R(x,y)dx^{-\gamma_1} = y^{1-\gamma_1}\int_0^\infty R(x,1)dx^{-\gamma_1}.$$

Hence, we get that

$$e_n\theta_{\frac{ke_n}{n}} = e_n^{1-\gamma_1}\theta_{\frac{k}{n}} + o_P\left(\frac{U_1(n/k)}{\sqrt{k}}\right).$$

By applying (21), Proposition 1 and the Cramér's delta method, we get that, as $n \to \infty$,

$$\frac{\sqrt{k}}{U_1(n/k)}\left(e_n\theta_{\frac{ke_n}{n}} - \theta_{\frac{k}{n}}\right) = \sqrt{k}\left(e_n^{1-\gamma_1} - 1\right)\frac{\theta_{\frac{k}{n}}}{U_1(n/k)} + o_P(1)$$

$$\xrightarrow{P} (\gamma_1 - 1)W_R(\infty,1)\int_0^\infty R(s^{-1/\gamma_1},1)ds.$$

which implies that to $J_2 \xrightarrow{P} 0$. The proposition is thus proved. $\square$

Finally, we can combine the asymptotic relations on $\hat{\theta}_{\frac{k}{n}}$ and $\hat{\gamma}_1$ to obtain the proof of Theorem 1.

Proof of Theorem 1 Write

$$\frac{\hat{\theta}_p}{\theta_p} = \frac{d_n^{\hat{\gamma}_1}}{d_n^{\gamma_1}} \times \frac{\hat{\theta}_{\frac{k}{n}}}{\theta_{\frac{k}{n}}} \times \frac{d_n^{\gamma_1}\theta_{\frac{k}{n}}}{\theta_p} =: L_1 \times L_2 \times L_3.$$

We deal with the three factors separately.

Firstly, handling L_1 uses the asymptotic normality of the Hill estimator. Under conditions (b) and (c), we have that, as $n \to \infty$,

$$\sqrt{k_1}(\hat{\gamma}_1 - \gamma_1) \xrightarrow{P} \Gamma; \tag{24}$$

see, e.g., Example 5.1.5 in de Haan and Ferreira (2006). As in the proof of Theorem 4.3.8 of de Haan and Ferreira (2006), this leads to

$$\frac{\sqrt{k_1}}{\log d_n}(L_1 - 1) - \Gamma \xrightarrow{P} 0. \tag{25}$$

Secondly, the asymptotic behavior of the factor L_2 is given by Proposition 3.

Lastly, for L_3, by condition (b) and Theorem 2.3.9 in de Haan and Ferreira (2006), we have that

$$\frac{\frac{U_1(1/p)}{U_1(n/k)d_n^{\gamma_1}} - 1}{A_1(n/k)} \to -\frac{1}{\rho_1}.$$

Together with the fact that as $n \to \infty$, $\sqrt{k}\,|A_1(n/k)| \to 0$ (implied by condition (d)), we get that

$$\frac{U_1(1/p)}{U_1(n/k)d_n^{\gamma_1}} - 1 = o\left(\frac{1}{\sqrt{k}}\right) \tag{26}$$

Following the same reasoning of (23) for $p \le k/n$, we have $\frac{\theta_p}{U_1(1/p)} - \int_0^\infty R(s^{-1/\gamma_1}, 1)ds = o\left(\frac{1}{\sqrt{k}}\right)$. Combining this with (26), we have

$$L_3 = \frac{\theta_{\frac{k}{n}}/U_1(n/k)}{\theta_p/U_1(1/p)} \times \frac{U_1(n/k)d_n^{\gamma_1}}{U_1(1/p)} = 1 + o\left(\frac{1}{\sqrt{k}}\right). \tag{27}$$

Combining the asymptotic relations (25), (27) and Proposition 3, we get that

$$\frac{\hat{\theta}_p}{\theta_p} - 1$$

$$= L_1 \times L_2 \times L_3 - 1$$

$$= \left(1 + \frac{\log d_n}{\sqrt{k_1}}\Gamma + o_P\left(\frac{\log d_n}{\sqrt{k_1}}\right)\right)\left(1 + \frac{\Theta}{\sqrt{k}} + o_P\left(\frac{1}{\sqrt{k}}\right)\right)\left(1 + o\left(\frac{1}{\sqrt{k}}\right)\right) - 1$$

$$= \frac{\log d_n}{\sqrt{k_1}}\Gamma + \frac{\Theta}{\sqrt{k}} + o_P\left(\frac{\log d_n}{\sqrt{k_1}}\right) + o_P\left(\frac{1}{\sqrt{k}}\right).$$

The covariance matrix of (Θ, Γ) follows from the straightforward calculation. $\qquad\square$

Proof of Theorem 2 Write $\theta_p^+ := E(X^+ \mid Y > U_2(1/p))$. Then,

$$\frac{\hat{\theta}_p}{\theta_p} = \frac{\hat{\theta}_p}{\theta_p^+} \times \frac{\theta_p^+}{\theta_p}.$$

Hence, it suffices to prove that $\frac{\hat{\theta}_p}{\theta_p^+}$ follows the asymptotic normality stated in Theorem 1 and $\frac{\theta_p}{\theta_p^+} = 1 + o\left(\frac{1}{\sqrt{k}}\right)$.

We first show that (X^+, Y) satisfies conditions (a) and (b) of Section 2.1. Let $\widetilde{F}_1$ be the distribution function of X^+ and $\widetilde{U}_1 = \left(\frac{1}{1-\widetilde{F}_1}\right)^{\leftarrow}$. It is obvious that $U_1(t) = \widetilde{U}_1(t)$, for $t > \frac{1}{1-\widetilde{F}_1(0)}$. Hence X^+ satisfies condition (b).

Before checking condition (a) for (X^+, Y), we prove that, as $t \to \infty$,

$$tP(X < 0, 1 - F_2(Y) < 1/t) = O(t^\tau). \tag{28}$$

24

Observe that condition (a) implies that

$$\sup_{1/2 \le y \le 2} |y - R(t, y)| = O(t^\tau). \tag{29}$$

Because of the homogeneity of R, we have $1 - R(ct, 1) = O(t^\tau)$ for any $c \in (0, \infty)$. Hence, (28) is proved by

$$tP(X < 0, 1 - F_2(Y) < 1/t)$$
$$= 1 - tP(X > 0, 1 - F_2(Y) < 1/t)$$
$$= 1 - tP(1 - F_1(X) < 1 - F_1(0), 1 - F_2(Y) < 1/t)$$
$$\le 1 - R(t(1 - F_1(0)), 1) + |tP(1 - F_1(X) < 1 - F_1(0), 1 - F_2(Y) < 1/t) - R(t(1 - F_1(0)), 1)|$$
$$= O(t^\tau).$$

Now we show that (X^+, Y) satisfies condition (a), that is, as $t \to \infty$,

$$\sup_{\substack{0 < x < \infty \\ 1/2 \le y \le 2}} \frac{\left| tP(1 - \widetilde{F}_1(X^+) < x/t, 1 - F_2(Y) < y/t) - R(x, y) \right|}{x^\beta \wedge 1} = O(t^\tau). \tag{30}$$

Firstly, observe that for $0 < x \le t(1 - F_1(0))$,

$$\{1 - \widetilde{F}_1(X^+) < x/t\} = \{1 - F_1(X^+) < x/t\} = \{1 - F_1(X) < x/t\}.$$

Hence, the uniform convergence (in (30)) on $(0, t(1 - F_1(0))] \times [1/2, 2]$ follows from the fact that (X, Y) satisfies condition (a). Secondly, for $x > t(1 - F_1(0))$, we have $1 - \widetilde{F}_1(X^+) < x/t$. Therefore,

$$\sup_{\substack{t(1 - F_1(0)) < x < \infty \\ 1/2 \le y \le 2}} \left| tP(1 - \widetilde{F}_1(X^+) < x/t, 1 - F_2(Y) < y/t) - R(x, y) \right|$$
$$= \sup_{\substack{t(1 - F_1(0)) < x < \infty \\ 1/2 \le y \le 2}} (y - R(x, y))$$
$$\le \sup_{1/2 \le y \le 2} (y - R(t(1 - F_1(0)), y)) = O(t^\tau),$$

where the last relation follows from (29). This completes the verification of (30). As a result, Theorem 1 applies to $\frac{\hat{\theta}_p}{\theta_p^+}$.

Next we show that $\frac{\theta_p}{\theta_p^+} = 1 + o\left(\frac{1}{\sqrt{k}}\right)$. By Proposition 1, $\frac{\theta_p^+}{U_1(1/p)} \to \int_0^\infty R\left(x^{-1/\gamma_1}, 1\right) dx$. By Hölder's inequality, condition (e) and (28),

$$-E(X^- | Y > U_2(1/p)) = -\frac{1}{p} E(X^- I(X < 0, Y > U_2(1/p)))$$

25

$$\leq \frac{1}{p} \left(E\left|X^-\right|^{1/\gamma_1} \right)^{\gamma_1} \left(P(X < 0, Y > U_2(1/p)) \right)^{1-\gamma_1}$$

$$= O(p^{-1+(1-\tau)(1-\gamma_1)}).$$

Condition (b) can be written as:

$$\lim_{t \to \infty} \frac{U_1(tx)(tx)^{-\gamma_1} - U_1(t)t^{-\gamma_1}}{A_1(t)U_1(t)t^{-\gamma_1}} = \frac{x^{\rho_1} - 1}{\rho_1}.$$

It follows from Theorem B.2.2 in de Haan and Ferreira (2006) that $\frac{1}{U_1(1/p)} = O(p^{\gamma_1})$, as $p \downarrow 0$. Hence by condition (f),

$$\frac{\theta_p}{\theta_p^+} = 1 + \frac{E(X^- \mid Y > U_2(1/p))}{\theta_p^+} = 1 + O\left(\frac{p^{-1+(1-\tau)(1-\gamma_1)}}{U_1(1/p)} \right)$$

$$= 1 + O\left(p^{-\tau(1-\gamma_1)} \right) = 1 + o\left(\frac{1}{\sqrt{k}} \right). \qquad \square$$

References

V.V. Acharya, L.H. Pedersen, T. Philippon, and M. Richardson. Measuring systemic risk. *Preprint*, 2010.

R.J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics Lecture Notes-Monograph Series, 1990.

K.S. Alexander. Sample moduli for set-indexed Gaussian processes. *The Annals of Probability*, 14:598–611, 1986.

J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, 2004.

C. Brownlees and R. Engle. Volatility, correlation and tails for systemic risk measurement. *Preprint*, 2011.

J. Cai and H. Li. Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, 42:810–825, 2005.

L. de Haan and A. Ferreira. *Extreme Value Theory: an Introduction*. Springer Verlag, 2006.

H. Drees and X. Huang. Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64:25–47, 1998.

J.H.J. Einmahl. *Multivariate Empirical Processes.* CWI Tract 32, Mathematisch Centrum, Amsterdam, 1987.

J.H.J. Einmahl, L. de Haan, and D. Li. Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics*, 34: 1987–2014, 2006.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.

H. Joe and H. Li. Tail risk of multivariate regular variation. *Methodology and Computing in Applied Probability*, 13:671–693, 2011.

K. Kostadinov. Tail approximation for credit risk portfolios with heavy-tailed risk factors. *Journal of Risk*, 8:81–107, 2006.

Z.M. Landsman and E.A. Valdez. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7:55–71, 2003.

R. Vernic. Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, 38:413–426, 2006.

I. Weissman. Estimation of parameters and larger quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73:812–815, 1978.

Juan-Juan Cai	Laurens de Haan	Chen Zhou
j.cai@uvt.nl	ldehaan@ese.eur.nl	c.zhou@dnb.nl; zhou@ese.eur.nl
John H.J. Einmahl		
j.h.j.einmahl@uvt.nl		
Dept. of Econometrics & OR,	Department of Economics,	Economic and Research Division,
CentER, Tilburg University,	Erasmus University,	De Nederlandsche Bank,
P.O. Box 90153,	P.O. Box 1738,	P.O. Box 98,
5000 LE Tilburg,	3000 DR Rotterdam,	1000 AB Amsterdam,
The Netherlands	The Netherlands	The Netherlands