Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2015)
77, *Part* 2, *pp.* 417–442

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

Juan-Juan Cai,

Delft University of Technology, The Netherlands

John H. J. Einmahl,

Tilburg University, The Netherlands

Laurens de Haan

Erasmus University Rotterdam, The Netherlands, and University of Lisbon, Portugal

and Chen Zhou

De Nederlandsche Bank, Amsterdam, Erasmus University Rotterdam, and Tinbergen Institute, Amsterdam, The Netherlands

[Received March 2013. Revised February 2014]

Summary. 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 $E\{X|Y>Q_Y(1-p)\}$, 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 non-parametric 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 $n\rightarrow\infty$. 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 behaviour. The finite sample performance of the estimator and the relevance of the limit theorem are shown in a detailed simulation study. We also apply our method to estimate the MES of three large US investment banks.

Keywords: Asymptotic normality; Conditional tail expectation; Extreme values

1. Introduction

An important step in constructing a systemic risk measure for a financial institution is to measure the contribution of the institution to a systemic crisis. The total risk measured by the expected capital shortfall in the financial system during a systemic crisis can be decomposed into firm level contributions measured by the marginal expected shortfall (MES). Following such a decomposition, the MES of a financial institution is then defined as the expected loss on its equity return conditional on the occurrence of an extreme loss in the aggregated return of the financial market. Denote the loss of the equity return of a financial institution and that of the entire market as X and Y respectively. Then the MES is equal to $E(X|Y>t)$, where t is a high threshold such

Address for correspondence: John H. J. Einmahl, Department of Econometrics and Operations Research and Center for Economic Research, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands.
E-mail: j.h.j.einmahl@uvt.nl

1369–7412/15/77417

that $p=P(Y>t)$ is extremely small. In other words, the MES at probability level p is defined as

$$\text{MES}(p)=E\{X|Y>Q_Y(1-p)\},$$

where Q_Y is the quantile function of Y. Note 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. (Acharya *et al.* (2012) proposed to use the MES for constructing a systemic risk measure, in which an extreme tail event is specified as one 'that happens once or twice a decade (or less)'. The estimation procedure therein is based on daily data from only 1 year.)

It is the goal of this paper to establish a novel estimator of $\text{MES}(p)$ and to unravel its asymptotic behaviour. The main result establishes the asymptotic normality of our estimator for a large class of bivariate distributions, which makes statistical inference for the MES feasible. We also show through a simulation study that the estimator performs well and that the limit theorem provides an adequate approximation for finite sample sizes.

The MES has also been studied under the name 'conditional tail expectation' (CTE). The definition of the 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 is the quantile function of X. When X has a continuous distribution, this is also called the expected shortfall. As risk measures, the tail value at risk and expected shortfall have been widely studied and applied to quantifying risk exposure in finance and actuarial science. The concept of the CTE has been defined also in a multivariate set-up. 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. In particular, for a risk vector $(Z_1,\ldots,Z_d)$ the following multivariate CTEs are commonly used in the literature:

$$\begin{aligned} &E\{Z_j|Z_1+\ldots+Z_d>Q_{Z_1+\ldots+Z_d}(1-p)\},\\ &E[Z_j|\min\{Z_1,\ldots,Z_d\}>Q_{\min\{Z_1,\ldots,Z_d\}}(1-p)],\\ &E[Z_j|\max\{Z_1,\ldots,Z_d\}>Q_{\max\{Z_1,\ldots,Z_d\}}(1-p)], \qquad j=1,\ldots,d; \end{aligned}$$

see for instance Cai and Li (2005); other relevant conditioning events have been given in Cousin and Di Bernardino (2013). Note that these CTEs are covered by our general definition. The first (conditioning on the sum) has been calculated explicitly for a Farlie–Gumbel–Morgenstern family of copulas in Bargès *et al.* (2009), for elliptical distributions in Landsman and Valdez (2003) (see also Kostadinov (2006)), and for skewed normal distributions in Vernic (2006). Various CTEs for multivariate phase-type distributions were given in Cai and Li (2005).

Compared with 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 approximating the CTE was provided. The multivariate regularly varying distributions form a subclass of our model. Note that in the present paper we do not make any assumption on the marginal distribution of Y. Similar approximations of the CTE under related extreme-value-type conditions can be found in Asimit *et al.* (2011) and Hua and Joe (2011). It should be emphasized, however, that we focus on the *statistical* problem of estimating the MES and studying the performance of the estimator in contrast with these papers where (only) probabilistic properties of the MES were studied.

In Acharya *et al.* (2012) an estimator for the MES *was* provided assuming a specific *linear* relationship between X and Y. The estimation procedure there can be seen as a special case of the present procedure. A similar setting has been adopted in Brownlees and Engle (2012), where a non-parametric kernel estimator of the MES was proposed. Such a kernel estimation method, however, performs well only if the threshold for defining a systemic crisis is not too high: the tail probability level p should be substantially larger than $1/n$. That method cannot handle extreme events, i.e. $p<1/n$, which is particularly required for systemic risk measures. For a study of CTEs in a regression set-up using kernel estimation (and hence using not too high a threshold), see El Methni *et al.* (2014).

The paper is organized as follows. Section 2 provides the main result: asymptotic normality of the estimator. In Section 3, the finite sample performance of the estimator is studied through a simulation study. An application on estimating the MES for US financial institutions is given in Section 4. The proofs are deferred to Appendix A.

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

$$U_j=\left(\frac{1}{1-F_j}\right)^{\leftarrow}, \qquad j=1,2,$$

where '$\leftarrow$' denotes the left continuous inverse. Then the MES at a probability level p can be written as

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

The goal is to estimate θ_p on the basis of independent and identically distributed observations, $(X_1, Y_1),\ldots,(X_n, Y_n)$ from F, where $p=p(n)\to 0$ as $n\to\infty$.

We impose mild assumptions on the distribution function F motivated by applications. To deal with the tail event in the conditional expectation, we adopt conditions from extreme value theory (EVT). More specifically, since the MES is the expected loss on the equity return of one financial institution given the occurrence of an extreme loss in the system, we expect that high values of Y correspond to high values of X. Thus, we impose assumptions on the right-hand tail of X and on the right-hand upper tail dependence of (X, Y).

We begin with describing the right-hand upper tail dependence between X and Y, as in EVT. Suppose that, for all $(x, y)\in[0,\infty]^2\setminus\{(\infty,\infty)\}$, the following limit exists:

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

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

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

see Drees and Huang (1998) and Beirlant *et al.* (2004), section 8.2.

Next, for the marginal distributions, we assume as usual in financial or non-life insurance settings that X follows a distribution with a heavy right-hand tail: there exists $\gamma_1>0$ such that, for $x>0$,

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

Then it follows that $1-F_1$ is regularly varying with index $-1/\gamma_1$ and γ_1 is the extreme value index. Throughout, we do not need an assumption, apart from continuity, on the marginal distribution of Y because applying any continuous, increasing transformation on Y will not change the MES.

When considering the CTE (in actuarial settings) we typically have $X>0$. When considering loss returns we have $X\in\mathbb{R}$, but the MES is mainly determined by high, and hence positive, values of X. Therefore we consider first the case $X>0$. Subsequently we shall deal with the case $X\in\mathbb{R}$. In the latter case we shall make a mild assumption (condition (e) below) to guarantee that the left-hand tail of X is asymptotically negligible in the MES.

2.1. X positive

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

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

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

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

On the basis of the limit given in proposition 1, we construct an estimator of θ_p by a two-stage approach. Let $k=k(n)$ be an intermediate sequence of integers, i.e. $k\to\infty$ and $k/n\to 0$, as $n\to\infty$. Firstly, we consider the estimation of $\theta_{k/n}$, the MES at an intermediate probability level k/n. At such a level, there are sufficiently many (approximately k) observations (X_i,Y_i) such that $Y_i>U_2(n/k)$. Thus $\theta_{k/n}$ can be estimated non-parametrically by taking the average of the X_i of those selected observations. Secondly, we use an extrapolation method based on proposition 1 and a strengthening of condition (2) (see condition (b) below). More precisely, we have that, as $n\to\infty$,

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

For estimating θ_p, it thus remains to estimate γ_1. 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 $k_1=k_1(n)$ is another intermediate sequence of integers and $X_{i,n}$, $i=1,\ldots,n$, is the ith 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 the aforementioned non-parametric estimator of $\theta_{k/n}$ by

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

Combining expressions (3)–(5), we estimate θ_p by

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

To prove asymptotic normality of $\hat{\theta}_p$, we need to quantify the rates of convergence in conditions (1) and (2) (see, for example, de Haan and Ferreira (2006), conditions (7.2.8) and (3.2.4) respectively) as follows.

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

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

(b) There exist $\rho_1 < 0$ and an eventually positive or negative function A_1 such that, as $t \to \infty$, $A_1(tx)/A_1(t) \to x^{\rho_1}$ for all $x > 0$ and

$$\sup_{x>1} \left| x^{-\gamma_1} \frac{U_1(tx)}{U_1(t)} - 1 \right| = O\{A_1(t)\}.$$

Conditions (a) and (b) are second-order strengthenings of conditions (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 $\alpha < \min\{-2\tau/(-2\tau+1), 2\gamma_1\rho_1/(2\gamma_1\rho_1+\rho_1-1)\}$.

To characterize the limit distribution of $\hat{\theta}_p$, we define a mean 0 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

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

and

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

It will be shown (see proposition 3 and equation (29)) that $\hat{\theta}_{k/n}$ and $\hat{\gamma}_1$ are asymptotically normal with Θ and Γ as limit respectively. The following theorem gives the asymptotic normality of $\hat{\theta}_p$.

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

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

where

$$\mathrm{var}(\Theta) = \gamma_1^2 - 1 - b^2 \int_0^\infty R(s,1)\, \mathrm{d}s^{-2\gamma_1},$$

$$\mathrm{var}(\Gamma) = \gamma_1^2$$

and

$$\mathrm{cov}(\Gamma,\Theta) = \gamma_1(1-\gamma_1+b)\, R(1,1) - \gamma_1 \int_0^1 [(1-\gamma_1) + bs^{-\gamma_1}\{1-\gamma_1-\gamma_1 \ln(s)\}]\, R(s,1) s^{-1}\, \mathrm{d}s$$

with $b = \{\int_0^\infty R(s,1)\, \mathrm{d}s^{-\gamma_1}\}^{-1}$.

Remark 1. The assumption $\gamma_1 \in (0, \frac{1}{2})$ is necessary for theorem 1, i.e. the result does not hold true when $\gamma_1 = \frac{1}{2}$. For the consistency of $\hat{\theta}_p$ this assumption can be relaxed to $\gamma_1 \in (0,1)$, as in proposition 1; also several of the other assumptions are not required then. To be precise, when conditions (1) and (b) hold, $R(1,1) > 0$, $\lim_{n\to\infty} \log(d_n)/\sqrt{k_1} = 0$ and $\gamma_1 \in (0,1)$, then, as $n \to \infty$,

$$\hat{\theta}_p/\theta_p \xrightarrow{\mathrm{P}} 1. \tag{7}$$

The proof of result (7) is quite straightforward compared with that of theorem 1 and has therefore been omitted.

Remark 2. Observe that the assumptions in proposition 1 do allow the case $R \equiv 0$ on $[0,\infty)^2$, i.e. X and Y are asymptotically (or tail) independent. The uniformity requirement in assumption (a), however, excludes asymptotic independence (to see this, take $x=t$). In practice, financial and actuarial data on which the MES is based typically exhibit (quite strong) asymptotic *dependence*; see, for example, Section 4. We shall, nevertheless, address the case of asymptotically independent X and Y in the simulation study (Section 3).

2.2. X real

In this section, X takes values in $\mathbb{R}$, i.e. we do not restrict X to be positive. Define $X^+ = \max(X,0)$ and $X^- = X - X^+$. Besides the conditions of theorem 1 (which imply left-hand upper tail independence of (X,Y)), we require two more conditions.

(e) $E|X^-|^{1/\gamma_1} < \infty$.
(f) As $n \to \infty$, $k = o(p^{2\tau(1-\gamma_1)})$.

It can be shown that condition (e), together with condition (a), ensures that $E\{X^- | Y > U_2(1/p)\}$ is asymptotically negligible in the calculation of the MES. Hence $\theta_p \sim E\{X^+ | Y > U_2(1/p)\}$, as $p \downarrow 0$. Therefore, we estimate θ_p with

$$\hat{\theta}_p = \left(\frac{k}{np}\right)^{\hat{\gamma}_1} \frac{1}{k} \sum_{i=1}^{n} X_i \, I(X_i > 0, Y_i > Y_{n-k,n}), \tag{8}$$

with $\hat{\gamma}_1$ as in Section 2.1. Observe that, when X is positive, this definition coincides with that in equation (6). As stated in the following theorem, the asymptotic behaviour 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 $n \to \infty$,

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

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 first generate data from three bivariate distributions. Throughout this section, (Z_1, Z_2) denotes a standard Cauchy distribution on $\mathbb{R}^2$ with density $(1/2\pi)(1+x^2+y^2)^{-3/2}$.

(a) The first distribution is a transformed Cauchy distribution on $(0,\infty)^2$ defined as

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

It follows that $\gamma_1 = \frac{2}{5}$ and $R(x,y) = x + y - \sqrt{(x^2+y^2)}$, $x, y \geqslant 0$. 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'.

(b) The second distribution is a Student t_3-distribution on $(0,\infty)^2$ with density

$$f(x,y)=\frac{2}{\pi}\left(1+\frac{x^2+y^2}{3}\right)^{-5/2}, \qquad x,y>0. \tag{9}$$

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

(c) The third distribution is a transformed Cauchy distribution on the whole $\mathbb{R}^2$ defined as

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

We have $\gamma_1=\frac{2}{5}$, $R(x,y)=x/2+y-\surd(x^2/4+y)$, $\tau=-1$, $\beta=2$ 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 $n=500$ and $n=2000$. On the basis of each sample, we estimate θ_p for p equal to 1/500, 1/5000 or 1/10 000.

Besides the estimator given by equation (8), we construct two other estimators. Firstly, for $np\geqslant 1$, an empirical counterpart of θ_p, given by

$$\hat{\theta}_{\mathrm{emp}}=\frac{1}{\lfloor np\rfloor}\sum_{i=1}^{n}X_i\,I(Y_i>Y_{n-\lfloor np\rfloor,n}), \tag{10}$$

is studied, where $\lfloor\cdot\rfloor$ denotes the integer part. Clearly, this empirical estimator is not applicable when $np<1$ and is not expected to perform well if np is 'small'. Secondly, exploiting the relationship in proposition 1 and using the empirical estimator of R given by

$$\hat{R}(x,y)=\frac{1}{k}\sum_{i=1}^{n}I(X_i>X_{n-\lfloor kx\rfloor,n}, Y_i>Y_{n-\lfloor ky\rfloor,n}), \qquad x,y\geqslant 0, \tag{11}$$

and the Weissman (1978) estimator of $U_1(1/p)$ given by $\hat{U}_1(1/p)=d_n^{\hat{\gamma}_1}X_{n-k,n}$, we define an alternative EVT estimator as

$$\begin{aligned}\bar{\theta}_p&=-\hat{U}_1\left(\frac{1}{p}\right)\int_0^\infty \hat{R}(x,1)\,\mathrm{d}x^{-\hat{\gamma}_1}\\&=d_n^{\hat{\gamma}_1}X_{n-k,n}\,\frac{1}{k}\sum_{i=1}^{n}I(Y_i>Y_{n-k,n})\left\{\frac{n-\mathrm{rank}(X_i)+1}{k}\right\}^{-\hat{\gamma}_1}.\end{aligned} \tag{12}$$

A comparison of the three estimators is shown in Fig. 1, where we present boxplots of the ratios of the estimates and the true values. We choose specific values for k and k_1 for each distribution and sample size as indicated. These values are chosen according to a selection procedure which is similar to that described in the application; see Section 4. For all three distributions, the empirical estimator underestimates the MES and is consistently outperformed by the two EVT estimators. In addition, it is not applicable for $p<1/n$. The two EVT estimators, $\hat{\theta}_p$ and $\bar{\theta}_p$, both perform well. Their behaviour 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 1 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 $\hat{\theta}_p/\theta_p$ in theorem 1 can be expressed as $\surd k(\hat{\theta}_p/\theta_p-1)\to^{\mathrm{d}}\Theta+r\Gamma$ or, equivalently,

$$\surd k\,\log\left(\frac{\hat{\theta}_p}{\theta_p}\right)\xrightarrow{\mathrm{d}}\Theta+r\Gamma.$$

Note that the limit distribution is a centred normal distribution. Write $\sigma_p^2=(1/k)\,\mathrm{var}(\Theta+r\Gamma)$ with

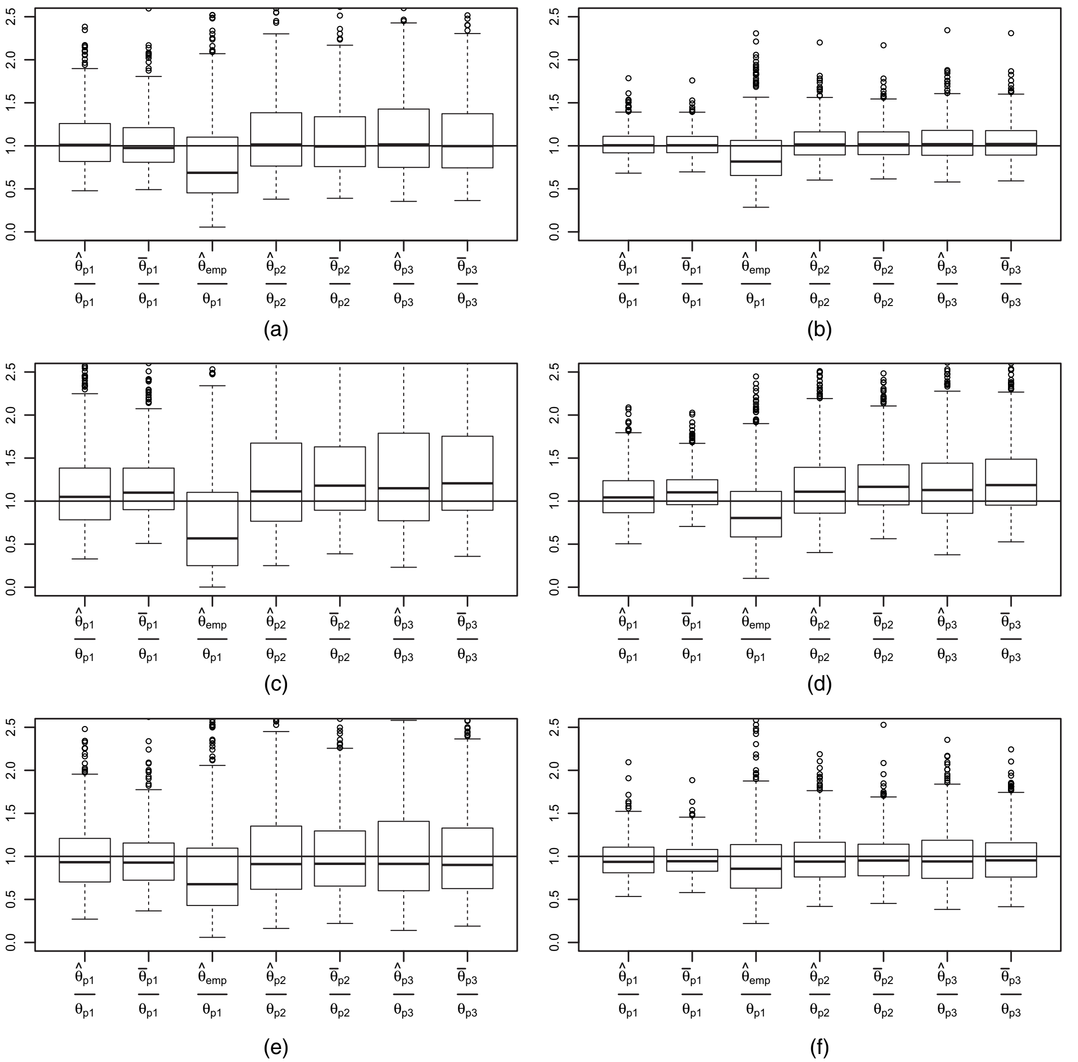

Fig. 1. Boxplots of ratios of estimates and true values (each plot is based on 500 samples with sample size $n=500$ or $n=2000$ from the transformed Cauchy distributions 1 or 2 or the Student t_3-distribution; the estimators are $\hat{\theta}_p$ of equation (8), $\bar{\theta}_p$ of equation (12) and $\hat{\theta}_{\text{emp}}$ of equation (10); p1 = 500; p2 = 5000 and p3 = 10000): (a) transformed Cauchy distribution 1, $n=500$, $k=75$, $k_1=75$; (b) transformed Cauchy distribution 1, $n=2000$, $k=300$, $k_1=300$; (c) Student t_3-distribution, $n=500$, $k=20$, $k_1=20$; (d) Student t_3-distribution, $n=2000$, $k=50$, $k_1=50$; (e) transformed Cauchy distribution 2, $n=500$, $k=20$, $k_1=25$; (f) transformed Cauchy distribution 2, $n=2000$, $k=80$, $k_1=100$

$$r=\frac{\sqrt{k}\,\log\{k/(np)\}}{\sqrt{k_1}}.$$

We compare the distribution of $(1/\sigma_p)\log(\hat{\theta}_p/\theta_p)$ with the limit distribution $N(0,1)$. The Q–Q-plots in Fig. 2 present the sample quantiles of $(1/\sigma_p)\log(\hat{\theta}_p/\theta_p)$, based on 500 estimates, *versus* the theoretical standard normal quantiles. We observe that the scatters line up on the line $y=x$ in each plot, which indicates that the sample quantiles coincide largely with the

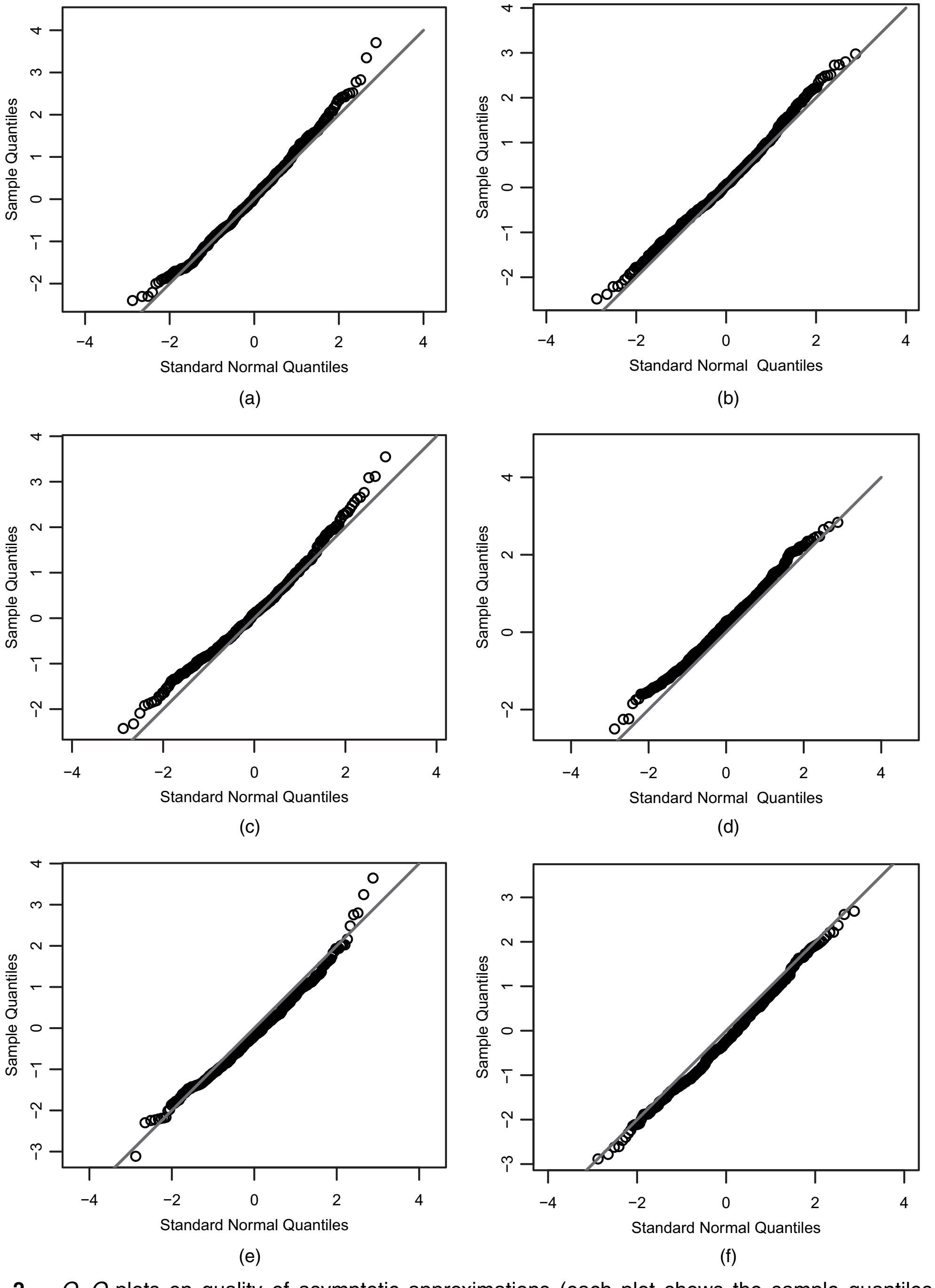

Fig. 2. *Q–Q*-plots on quality of asymptotic approximations (each plot shows the sample quantiles of $(1/\sigma_p)\log(\hat{\theta}_p/\theta_p)$ *versus* the theoretical standard normal quantiles, based on 500 samples with sample size $n = 500$ or $n = 2000$ and $p = 1/n$; data are simulated from the transformed Cauchy distribution 1 or 2 or the Student t_3-distribution; the choices of k and k_1 are the same as in Fig. 1; the line $y = x$ is also depicted): (a) transformed Cauchy distribution 1, $n = 500$; (b) transformed Cauchy distribution 1, $n = 2000$; (c) Student t_3-distribution, $n = 500$; (d) Student t_3-distribution, $n = 2000$; (e) transformed Cauchy distribution 2, $n = 500$; (f) transformed Cauchy distribution 2, $n = 2000$

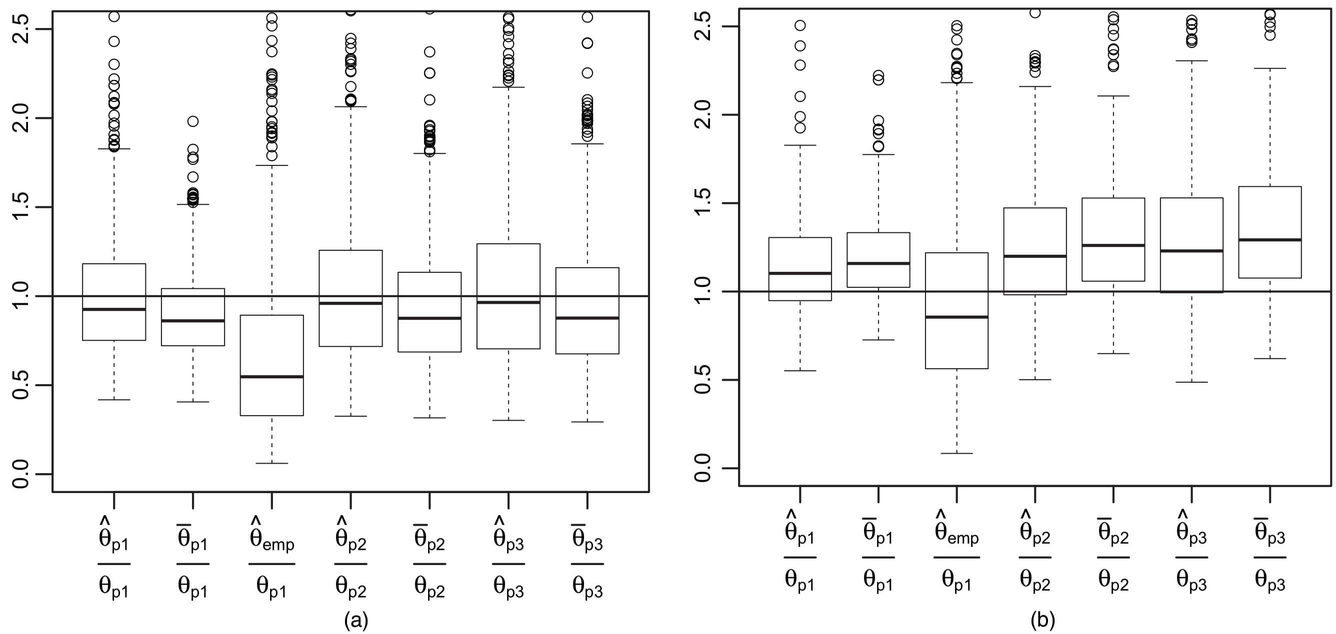

Fig. 3. Boxplots of ratios of estimates and true values when $\gamma_1 > \frac{1}{2}$ or $R \equiv 0$ (each plot is based on 500 samples with sample size $n = 2000$ from the transformed Cauchy distribution 3 or 'the' asymptotically independent distribution; the estimators are $\hat{\theta}_p$ of equation (8), $\bar{\theta}_p$ of equation (12) and $\hat{\theta}_{emp}$ of equation (10); p1 = 1/500; p2 = 1/5000; p3 = 1/10000): (a) transformed Cauchy distribution 3, $n = 2000$, $k = 300$, $k_1 = 300$; (b) asymptotically independent distribution, $n = 2000$, $k = 50$, $k_1 = 100$

theoretical quantiles from the asymptotic distribution. Consequently, we conclude that the limit theorem provides an adequate approximation for finite sample sizes.

We also investigate the performance of our estimator when our assumptions are partially violated. For that purpose, we generate data from the following two bivariate distributions.

(a) The transformed Cauchy distribution 3 is defined as

$$(X, Y) = (|Z_1|^{0.7}, |Z_2|).$$

The dependence structure between X and Y is the same as that of the transformed Cauchy distribution 1. Since $\gamma_1 = 0.7 > \frac{1}{2}$, this distribution is not permitted by the conditions of theorem 1. Nevertheless, the estimator $\hat{\theta}_p$ is still consistent; see remark 1.

(b) The second distribution is an asymptotically independent distribution defined as

$$(X, Y) = (V_1 + W_1, V_2 + W_2),$$

where (V_1, V_2) follows the Student t_3-distribution with density (9) and W_1 and W_2 are Pareto distributed with density $(25/2)(1+5x)^{-7/2}$, $x > 0$. Moreover (V_1, V_2), W_1 and W_2 are independent. X and Y are asymptotically independent, i.e. $R \equiv 0$ on $[0, \infty)^2$. This distribution does not satisfy condition (a); see remark 2.

We draw 500 samples from the two distributions with sample size $n = 2000$. On the basis of each sample, we estimate θ_p for $p = 1/500$, 1/5000 or 1/10 000. The comparison of the estimators is shown in Fig. 3, with boxplots of the ratios of the estimates and the true values. For both distributions, the empirical estimator is outperformed by the two EVT estimators. For the model with $\gamma_1 = 0.7$, between the two EVT estimators, $\hat{\theta}_p$ provides better, i.e. less biased, estimates. For the asymptotically independent model, both EVT estimators overestimate the actual MES slightly, but $\hat{\theta}_p$ performs better.

To conclude, the estimators $\hat{\theta}_p$ and $\bar{\theta}_p$ both perform well if our model assumptions are valid. The former is preferred with a lack of prior knowledge on the validity of the model assumptions.

4. Application

In this section, we apply our estimation method to estimate the MES for financial institutions. We consider three large investment banks in the USA, namely Goldman Sachs, Morgan Stanley and T. Rowe Price, all of which had a market capitalization greater than US \$5 billion at the end of June 2007. The data set consists of the loss returns (i.e. minus log-returns) on their equity prices at a daily frequency from July 3rd, 2000, to June 30th, 2010. (The choice of the banks, frequency of data and time horizon follows the same set-up as in Brownlees and Engle (2012). The sample size is close to the larger sample size in our simulation.) Moreover, for the same time period, we extract daily loss returns of a value-weighted market index aggregating three markets: the New York Stock Exchange, American Express stock exchange and the National Association of Securities Dealers Automated Quotation system. All return data are collected from the Center for Research in Security Prices database. We use our method to estimate the MES, $E\{X|Y > U_2(1/p)\}$, where X and Y refer to the daily loss returns of a bank equity and the market index respectively and $p = 1/n = 1/2513$, which corresponds to a once-per-decade systemic event.

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 equation (8). For that purpose,

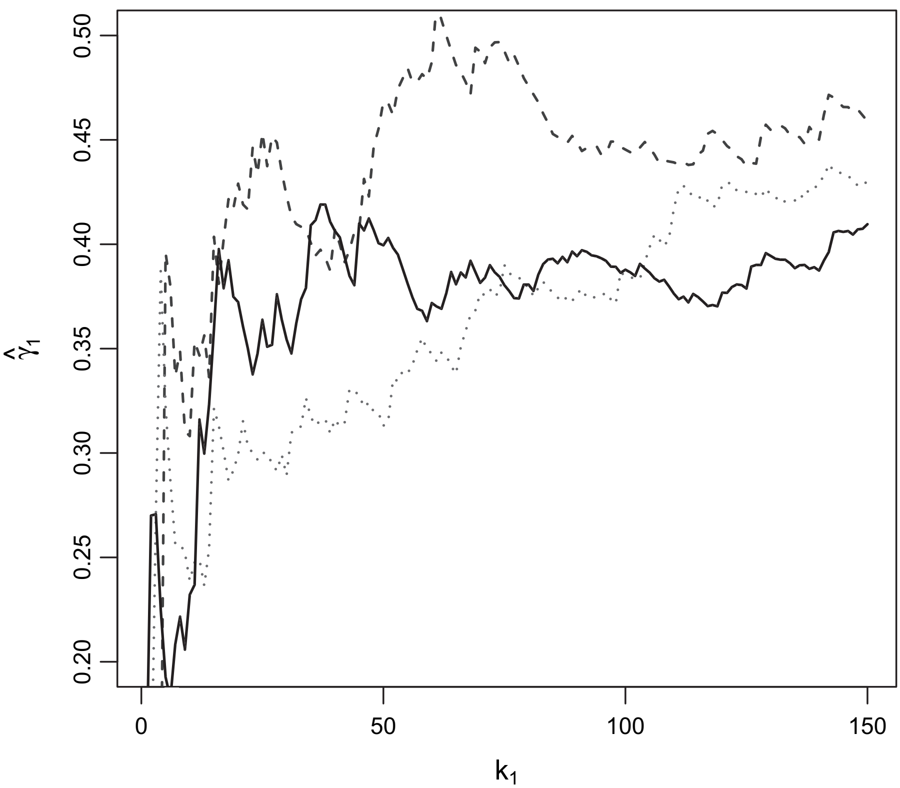

Fig. 4. Hill estimates based on daily loss returns of three investment banks: ———, Goldman Sachs; – – –, Morgan Stanley; · · · · · ·, T. Rowe Price

we first verify three of the conditions that are required for the procedure. First, the assumption that $\gamma_1 < \frac{1}{2}$ is confirmed by the plot of the Hill estimates in Fig. 4. Second, 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 Fig. 5, which plots the loss returns of the market index against those of the equity prices. The horizontal lines indicate the 50th largest loss of the index. The vertical lines, at 0, distinguish the occurrence of losses and profits. As we can see, from the upper parts of the plots, the largest 50 losses of the index are mostly associated with losses (i.e. $X > 0$). Finally we check the asymptotic dependence (see remark 2) by depicting in Fig. 6 the estimated tail dependence coefficient $\hat{R}(1,1)$ (see expression (11)) against k. Indeed all the plots show estimates that are at least $\frac{1}{2}$, strongly indicating that right-hand upper asymptotic dependence (i.e. $R(1,1) > 0$) is present.

Hence we can apply our method to obtain the estimates of $\text{MES}(p) = \theta_p$ for the three banks. In the estimation, two intermediate sequences k and k_1 are chosen in two steps. We first plot the Hill estimates of γ_1 against various values of k_1 as shown in Fig. 4. By balancing the potential estimation bias and variance, a usual practice is to choose k_1 from the first stable region of the plots. In this case, we choose $k_1 \in [70, 100]$. To gain stability in the estimates, we take the average of the estimates corresponding to those k_1-values and regard that as the estimate of γ_1. The results are reported in the second column in Table 1. Next, given this $\hat{\gamma}_1$, we estimate the MES with $\hat{\theta}_p$ and also plot the estimates against various values of k as in Fig. 7. Following the

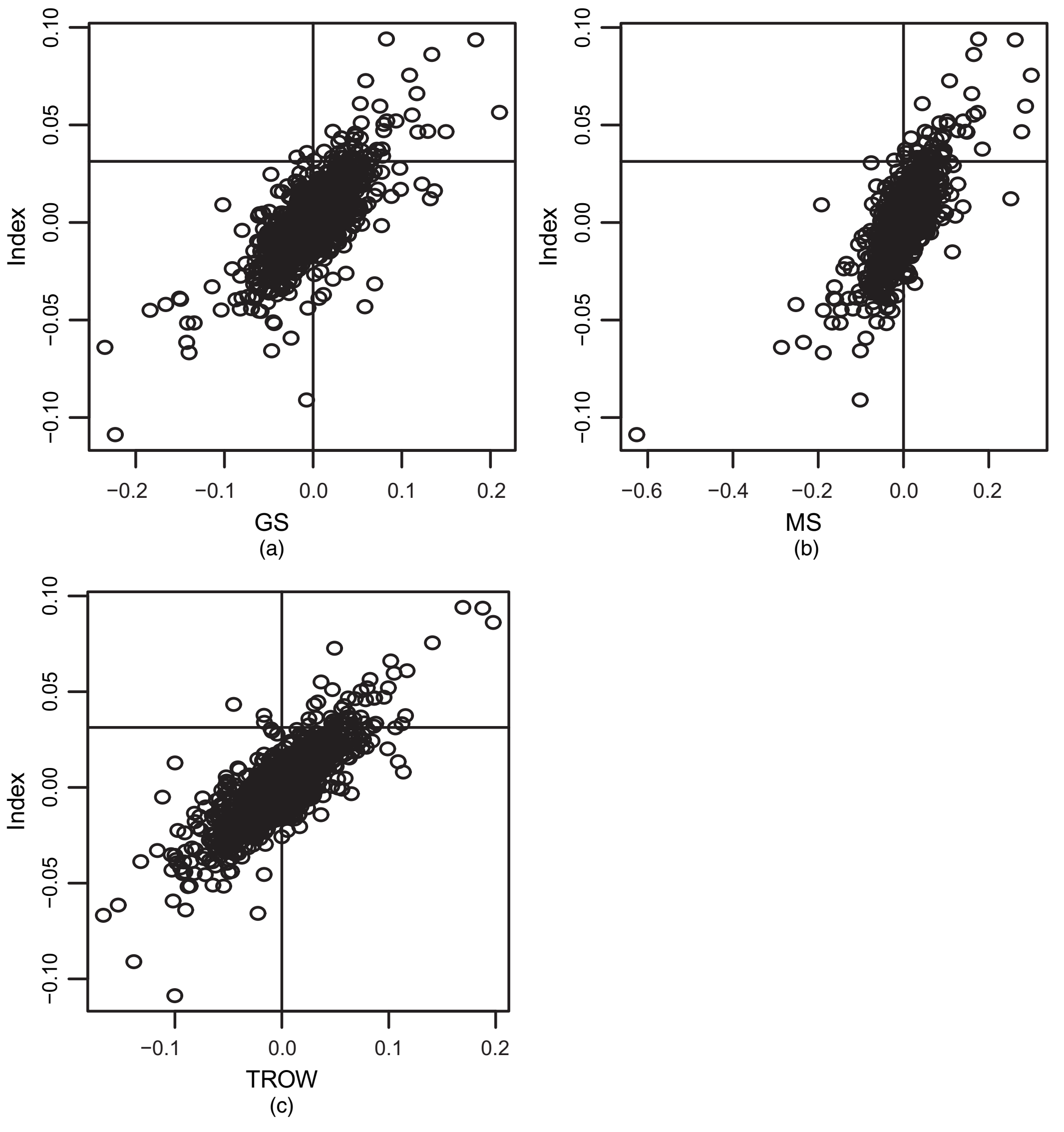

Fig. 5. Scatter plots of the loss returns of the individual equity and market index: (a) Goldman Sachs; (b) Morgan Stanley; (c) T. Rowe Price

same idea on balancing bias and variance, we choose $k \in [70, 100]$. The final estimates based on averaging the estimates from this region are reported in the third column of Table 1. These numbers represent the average daily loss return for a once-per-decade market crisis.

The theory for the MES estimator is derived for independent and identically distributed random vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$. For application to financial time series, the potential serial dependence may affect the estimation results. Similarly to extreme value analysis under mixing conditions in a univariate setting (see, for example, Drees (2000)), our method may work under serial dependence with an enlarged asymptotic variance. Theoretical results along these lines are left for future research. Here we consider a practical solution to reduce the po-

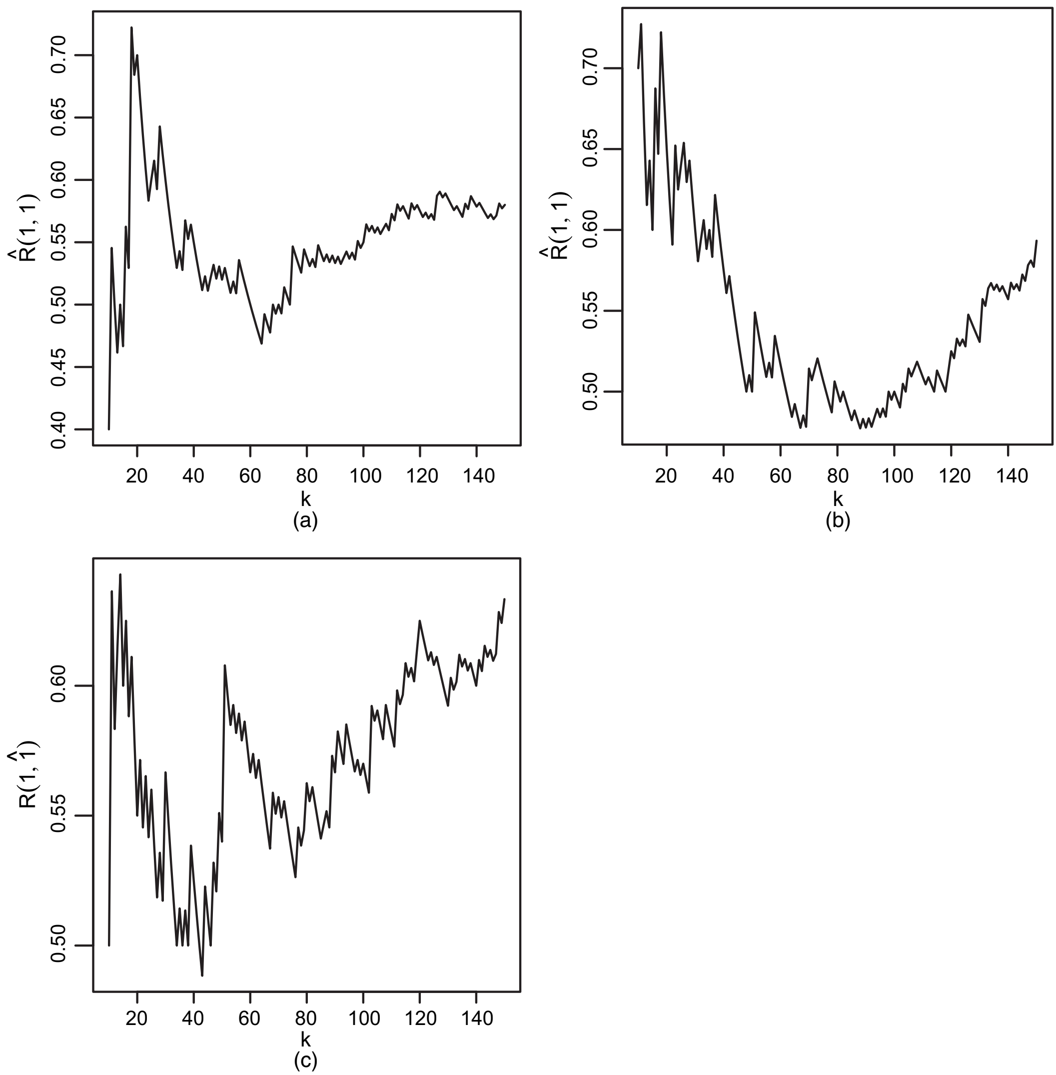

Fig. 6. $\hat{R}(1, 1)$ of the loss returns of individual equity and market index: (a) Goldman Sachs; (b) Morgan Stanley; (c) T. Rowe Price

tential serial dependence substantially: using lower frequency data. More specifically, we use *weekly* loss returns in the same sample period. That reduces the sample size to $n = 522$; a once-per-decade event occurs with probability $p = 1/n = 1/522$ accordingly. In the last two columns of Table 1, we report the estimation results on γ_1 and the MES based on weekly loss returns. The results are qualitatively robust to the change from daily to weekly data. Similarly to the results for the daily losses, the MES levels for Goldman Sachs and T. Rowe Price are almost equal, whereas for Morgan Stanley we find an MES that is twice as high. Comparing with the results based on daily losses, it seems that the extreme weekly losses of the three banks during a once-per-decade market crisis are mainly caused by the losses on the worst day during such a week.

Table 1. MES of the three investment banks†

Bank	*Daily loss*		*Weekly loss*	
	$\hat{\gamma}_1$	$\hat{\theta}_p$	$\hat{\gamma}_1$	$\hat{\theta}_p$
Goldman Sachs	0.388	0.308	0.417	0.346
Morgan Stanley	0.465	0.608	0.483	0.654
T. Rowe Price	0.378	0.316	0.347	0.339

†The second and third columns report the results based on *daily* loss returns ($n = 2513$ and $p = 1/n$). The estimates $\hat{\gamma}_1$ are computed by taking the average for $k_1 \in [70, 100]$. The estimates of the MES are based on these values of $\hat{\gamma}_1$. We report the average of the MES estimates $\hat{\theta}_p$ for $k \in [70, 100]$. The last two columns report the results based on *weekly* loss returns from the same sample period ($n = 522$ and $p = 1/n$), where both k_1 and k are from the interval [20, 30].

Acknowledgements

We are grateful to the Associate Editor and two referees for various insightful comments and questions that led to this improved version of the manuscript.

The research of Laurens de Haan is partially supported by the Fundação para a Ciência e a Tecnologia (Portugal), projects PEst-OE/MAT/UI0006/2014 and EXPL/MAT-STA/0622/2013.

Views of Chen Zhou expressed in the paper do not necessarily reflect the official position of De Nederlandsche Bank.

Appendix A: Proofs

A.1. Proof of proposition 1

Recall that, for a non-negative random variable Z,

$$E(Z) = \int_0^\infty P(Z > x)\,\mathrm{d}x.$$

Hence,

$$\begin{aligned} \frac{\theta_p}{U_1(1/p)} &= \int_0^\infty \frac{1}{p} P\{X > x, Y > U_2(1/p)\} \frac{\mathrm{d}x}{U_1(1/p)} \\ &= \int_0^\infty \frac{1}{p} P\{X > U_1(1/p)x, Y > U_2(1/p)\}\,\mathrm{d}x. \end{aligned} \tag{13}$$

The limit relationships (1) and (2) imply that

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

Hence, we must prove only that the integral in equation (13) and the limit procedure $p \to 0$ can be interchanged. This is ensured by the dominated convergence theorem as follows. For $x \geqslant 0$,

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

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 $x > 1$,

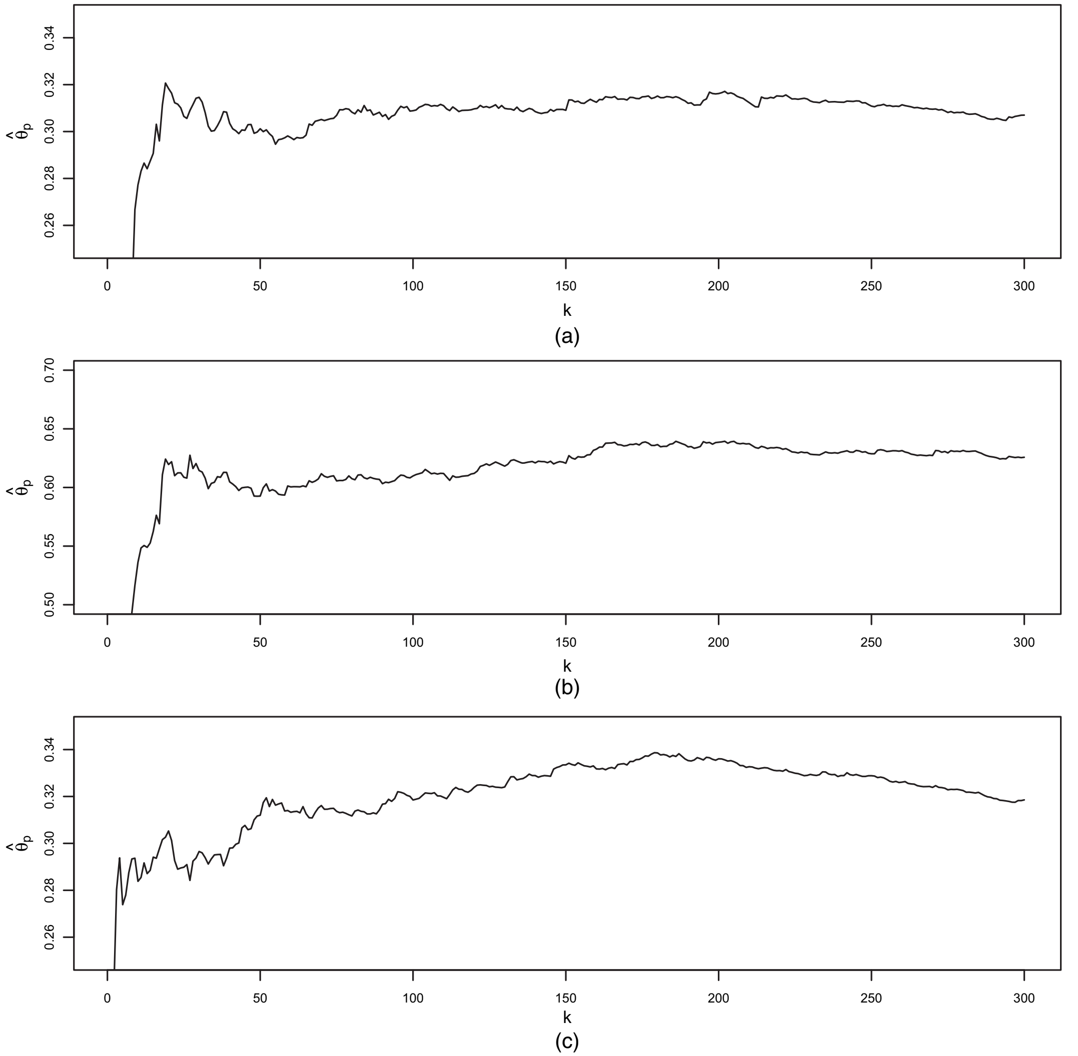

Fig. 7. Estimates of the MES for daily loss returns of three investment banks (estimates $\hat{\theta}_p$, with $\hat{\gamma}_1$ as in Table 1, from July 3rd, 2000, to June 30th, 2010): (a) Goldman Sachs; (b) Morgan Stanley; (c) T. Rowe Price

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

Write

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

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

Next, we prove theorem 1. The general idea of the proof is described as follows. It is clear that the asymptotic behaviour of $\hat{\theta}_p$ results from that of $\hat{\gamma}_1$ and $\hat{\theta}_{k/n}$. The asymptotic normality of $\hat{\gamma}_1$ is well known; see, for example, de Haan and Ferreira (2006), theorem 3.2.5. To prove the asymptotic normality of $\hat{\theta}_{k/n}$, write

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

where $e_n = (n/k)\{1 - F_2(Y_{n-k,n})\} \to^{\mathrm{P}} 1$, as $n \to \infty$, which follows from the fact that $1 - F_2(Y_{n-k,n})$ is the $(k+1)$th order statistic of a random sample of size n from the standard uniform distribution. Hence, with

$$\tilde{\theta}_{ky/n} := \frac{1}{ky}\sum_{i=1}^{n} X_i\, I\left\{Y_i > U_2\left(\frac{n}{ky}\right)\right\},$$

we first investigate the asymptotic behaviour of the random function $\tilde{\theta}_{k\cdot/n}$, on the interval $[\frac{1}{2}, 2]$. Then, by applying the result for $y = e_n$ and considering the asymptotic behaviour of e_n, we obtain the asymptotic normality of $\hat{\theta}_{k/n}$. Lastly, together with the asymptotic normality of $\hat{\gamma}_1$, we prove that of $\hat{\theta}_p$.

To obtain the asymptotic behaviour of the function $\tilde{\theta}_{ky/n}$, $y \in [\frac{1}{2}, 2]$, we introduce some new notation and auxiliary lemmas. Write $R_n(x, y) := (n/k)P\{1 - F_1(X) < kx/n, 1 - F_2(Y) < ky/n\}$. A pseudo-non-parametric estimator of R_n is given as

$$T_n(x, y) := \frac{1}{k}\sum_{i=1}^{n} I\left\{1 - F_1(X_i) < \frac{kx}{n}, 1 - F_2(Y_i) < \frac{ky}{n}\right\}.$$

(It is called a 'pseudo'-estimator because the marginal distribution functions are unknown.) The following lemma shows the asymptotic behaviour of the pseudo-estimator. The limit process is characterized by the aforementioned W_R-process. For convenient presentation, all the limit processes that are 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 processes. The proof of lemma 1 is analogous to that of proposition 3.1 in Einmahl *et al.* (2006) and has thus been omitted.

Lemma 1. Suppose that condition (1) holds. For any $\eta \in [0, \frac{1}{2})$ 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 weighting 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 $\eta \in [0, \frac{1}{2})$, with probability 1,

$$\sup_{0 < x \leqslant T, 0 < y < \infty} \frac{|W_R(x, y)|}{x^{\eta}} < \infty,$$

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

Next, denote $s_n(x) = (n/k)[1 - F_1\{U_1(n/k)x^{-\gamma_1}\}]$ for $x > 0$. From the regular variation condition (2), we obtain 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 that condition (2) holds. Denote g as a bounded and continuous function on $[0, S_0) \times [a, b]$ with $0 < S_0 \leqslant \infty$ and $0 \leqslant a < b < \infty$. Moreover, suppose that there exist $\eta_1 > \gamma_1$ and $m > 0$ such that

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

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

$$\lim_{n\to\infty} \sup_{a\leqslant y\leqslant b} \left| \int_0^S g\{s_n(x), y\} - g(x, y)\, dx^{-\gamma_1} \right| = 0. \tag{14}$$

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

$$\lim_{n\to\infty} \sup_{a\leqslant y\leqslant b} \sqrt{k} \left| \int_0^S g\{s_n(x), y\} - g(x, y)\, dx^{-\gamma_1} \right| = 0. \tag{15}$$

A.2. Proof of lemma 3

We prove results (14) and (15) 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 condition (2) and proposition B.1.10 of de Haan and Ferreira (2006) that

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

and

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

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

$$\begin{aligned}
\sup_{a\leqslant y\leqslant b} \left| \int_0^\infty [g\{s_n(x), y\} - g(x, y)]\, dx^{-\gamma_1} \right| &\leqslant \sup_{a\leqslant y\leqslant b} \left(\left| \int_0^\delta [g\{s_n(x), y\} - g(x, y)]\, dx^{-\gamma_1} \right| \right. \\
&\quad + \left| \int_\delta^T [g\{s_n(x), y\} - g(x, y)]\, dx^{-\gamma_1} \right| \\
&\quad \left. + \left| \int_T^\infty [g\{s_n(x), y\} - g(x, y)]\, dx^{-\gamma_1} \right| \right) \\
&\leqslant -m \int_0^\delta (x^{(\gamma_1+\eta_1)/2} + x^{\eta_1})\, dx^{-\gamma_1} + \delta^{-\gamma_1} \sup_{\substack{\delta\leqslant x\leqslant T\\ a\leqslant y\leqslant b}} |g\{s_n(x), y\} - g(x, y)| \\
&\quad + 2\varepsilon \sup_{\substack{0\leqslant x<\infty\\ a\leqslant y\leqslant b}} |g(x, y)| \leqslant c_1 \varepsilon^{1/2} + \delta^{-\gamma_1} \sup_{\substack{\delta\leqslant x\leqslant T\\ a\leqslant y\leqslant b}} |g\{s_n(x), y\} - g(x, y)| \\
&\quad + 2\varepsilon \sup_{\substack{0\leqslant x<\infty\\ a\leqslant y\leqslant b}} |g(x, y)|,
\end{aligned}$$

where c_1 is a finite constant. Hence, result (14) 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 result (15). Denote $\tilde{T}_n = |A_1(n/k)|^{1/(\rho_1-1)}$. By the Lipschitz property of g,

$$\sup_{a\leqslant y\leqslant b} \left| \int_0^\infty [g\{s_n(x), y\} - g(x, y)]\, dx^{-\gamma_1} \right| \leqslant \int_0^{\tilde{T}_n} |s_n(x) - x|\, dx^{-\gamma_1} + 2 \sup_{\substack{0\leqslant x<\infty\\ a\leqslant y\leqslant b}} |g(x, y)| \tilde{T}_n^{-\gamma_1}. \tag{16}$$

It is thus sufficient to prove that both terms on the right-hand side of inequality (16) are $o(1/\sqrt{k})$. For the second term, condition (d) implies that $\alpha/\{2(1-\alpha)\} < \gamma_1\rho_1/(\rho_1 - 1)$. Thus for any $\varepsilon_0 \in (0, \gamma_1\rho_1/(\rho_1 - 1) - \alpha/\{2(1-\alpha)\})$, as $n \to \infty$, we have that

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

which leads to

$$\sqrt{k}\tilde{T}_n^{-\gamma_1} = \sqrt{k}|A_1(n/k)|^{\gamma_1/(1-\rho_1)} \to 0. \tag{17}$$

To deal with the first term, we first prove that, for any ε_0, there exist $c_0 > 0$ and t_0 such that, for all $t > t_0$ and $tu > t_0$,

$$\left| \frac{u^{-\gamma_1} U_1(tu)/U_1(t) - 1}{A_1(t)} \right| < c_0 \max(1, u^{\rho_1 - \varepsilon_0}). \tag{18}$$

Condition (b) implies that there exist t_1 and c_2 such that, for $t > t_1$ and $u > 1$,

$$\left| \frac{u^{-\gamma_1} U_1(tu)/U_1(t) - 1}{A_1(t)} \right| < c_2. \tag{19}$$

Consider the case $tu > t_1$ and $u < 1$. Then we replace t and u by tu and $1/u$ in inequality (19) respectively and obtain that

$$\left| \frac{u^{\gamma_1} U_1(t)/U_1(tu) - 1}{A_1(tu)} \right| < c_2,$$

for $tu > t_1$ and $u < 1$. It implies that

$$\left| \frac{u^{-\gamma_1} U_1(tu)/U_1(t) - 1}{A_1(t)} \right| < 2c_2 \left| \frac{A_1(tu)}{A_1(t)} \right| < 4c_2 u^{\rho_1 - \varepsilon_0},$$

for all $tu > t_2$ and $u < 1$, where $t_2 > 0$ depends on ε_0. The last inequality follows from Potter's inequality; see de Haan and Ferreira (2006), proposition B.1.9. Inequality (18) is proved by taking $c_0 = 4c_2$ and $t_0 = \max(t_1, t_2)$.

For $x \in (0, \tilde{T}_n]$ and $0 < \varepsilon_1 < \gamma_1/(1 - \rho_1)$, when n is sufficiently large,

$$U_1(n/k) x^{-\gamma_1} \geqslant U_1(n/k) \tilde{T}_n^{-\gamma_1} = U_1(n/k) |A_1(n/k)|^{\gamma_1/(1-\rho_1)} \geqslant (n/k)^{\gamma_1/(1-\rho_1) - \varepsilon_1},$$

which implies that $U_1(n/k) x^{-\gamma_1} \to \infty$ as $n \to \infty$. Hence we can apply inequality (18) with $t = n/k$ and $u = 1/s_n(x)$ for all $x \in (0, \tilde{T}_n]$ and obtain that, for sufficiently large n,

$$\left| \frac{s_n(x) - x}{A_1(n/k)} \right| \leqslant c_3 \max(x, x^{1 - \rho_1 + \varepsilon_2}),$$

with $c_3 > 2^{-\rho_1 + \varepsilon_0} c_0/\gamma_1$ and $\varepsilon_2 > \varepsilon_0$ some positive constants. Here we use Potter's bound for the regularly varying function $1 - F_1$ to ensure that $s_n(x) \leqslant 2x^{(-\rho_1 + \varepsilon_2)/(-\rho_1 + \varepsilon_0)}$ for all $x > 1$ and sufficiently large n. Hence, we obtain that

$$\begin{aligned}
\sqrt{k} \int_0^{\tilde{T}_n} |s_n(x) - x| \, \mathrm{d}x^{-\gamma_1} &\leqslant \sqrt{k} |A_1(n/k)| c_3 \left(\int_0^1 x \, \mathrm{d}x^{-\gamma_1} + \int_1^{\tilde{T}_n} x^{1 - \rho_1 + \varepsilon_2} \, \mathrm{d}x^{-\gamma_1} \right) \\
&\leqslant c_4 \sqrt{k} |A_1(n/k)| \tilde{T}_n^{1 - \rho_1 - \gamma_1 + \varepsilon_2} \\
&= c_4 \sqrt{k} |A_1(n/k)|^{(\gamma_1 - \varepsilon_2)/(1 - \rho_1)} \leqslant c_5 \sqrt{k} (n/k)^{\rho_1 \gamma_1/(1 - \rho_1) + \varepsilon_2},
\end{aligned} \tag{20}$$

with c_4 and c_5 some positive constants. Again, by condition (d), as $n \to \infty$, $c_5 \sqrt{k} (n/k)^{\rho_1 \gamma_1/(1 - \rho_1) + \varepsilon_0} \to 0$. Hence, result (15) is proved by combining expressions (16), (17) and (20). □

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

Proposition 2. Suppose that conditions (1) and (2) hold with $0 < \gamma_1 < \frac{1}{2}$. Then,

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

A.3. Proof of proposition 2

Recall that $s_n(x) = (n/k)[1 - F_1\{U_1(n/k) x^{-\gamma_1}\}]$, $x > 0$. Similarly to expression (13),

$$y \theta_{ky/n} = \int_0^\infty \frac{n}{k} P\left\{ X > s, Y > U_2\left(\frac{n}{ky}\right) \right\} \mathrm{d}s$$

$$
\begin{aligned}
&= \int_0^\infty \frac{n}{k} P\left\{1 - F_1(X) < 1 - F_1(s), 1 - F_2(Y) < \frac{ky}{n}\right\} \mathrm{d}s \\
&= \int_0^\infty R_n\left[\frac{n}{k}\{1 - F_1(s)\}, y\right] \mathrm{d}s \\
&= -U_1\left(\frac{n}{k}\right) \int_0^\infty R_n\{s_n(x), y\} \, \mathrm{d}x^{-\gamma_1}. \qquad (21)
\end{aligned}
$$

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

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

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

$$P\{I_1(T_0) > \varepsilon\} < \varepsilon, \qquad (22)$$

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

$$P\{I_{2,n}(T_0) > \varepsilon\} < \varepsilon; \qquad (23)$$

$$P\{I_{3,n}(T_0) > \varepsilon\} < \varepsilon. \qquad (24)$$

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

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

Then, for any $T > T_1$,

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

Thus inequality (22) holds provided that $T_0 > T_1$.

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

$$
\begin{aligned}
P\{I_{2,n}(T) > \varepsilon\} &= P\left(\sup_{1/2 \leqslant y \leqslant 2} \left| \int_T^\infty \sqrt{k}[T_n\{s_n(x), y\} - R_n\{s_n(x), y\}] \, \mathrm{d}x^{-\gamma_1} \right| > \varepsilon \right) \\
&\leqslant P\left(\sup_{x > T, \, 1/2 \leqslant y \leqslant 2} |\sqrt{k}[T_n\{s_n(x), y\} - R_n\{s_n(x), y\}]| > \varepsilon T^{\gamma_1} \right) \\
&= P\left[\sup_{x > T, \, 1/2 \leqslant y \leqslant 2} \left| \sqrt{n}(\tilde{P}_n - \tilde{P}) \left\{ \left(0, \frac{ks_n(x)}{n}\right) \times \left(0, \frac{ky}{n}\right) \right\} \right| > \varepsilon T^{\gamma_1} \sqrt{(k/n)} \right] \\
&=: p_2.
\end{aligned}
$$

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

$$
\begin{aligned}
p_2 &\leqslant c \exp\left[-\frac{\{\varepsilon T^{\gamma_1} \sqrt{(k/n)}\}^2}{4 \tilde{P}(S_n)} \psi\left\{ \frac{\varepsilon T^{\gamma_1} \sqrt{(k/n)}}{\sqrt{n} \tilde{P}(S_n)} \right\} \right] \\
&= c \exp\left\{ -\frac{\varepsilon^2 T^{\gamma_1}}{8} \psi\left(\frac{\varepsilon T^{\gamma_1/2}}{2\sqrt{k}} \right) \right\}.
\end{aligned}
$$

Choose $T_2 = T_2(\varepsilon)$ such that $c \exp(-\varepsilon^2 T_2^{\gamma_1}/16) \leqslant \varepsilon$. Then, for any $T > T_2$, $c \exp(-\varepsilon^2 T^{\gamma_1}/16) \leqslant \varepsilon$. Furthermore, we can choose $n_1 = n_1(T)$ such that, for $n > n_1$, $\psi\{\varepsilon T^{\gamma_1/2}/(2\sqrt{k})\} > \frac{1}{2}$. Therefore, for $T > T_2(\varepsilon)$ and $n > n_1(T)$, we have $p_2 < \varepsilon$, which leads to inequality (23) provided that $T_0 > T_2$ and $n_0 > n_1$.

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

$$\begin{aligned} P\{I_{3,n}(T) > \varepsilon\} &\leqslant P\left(\sup_{1/2 \leqslant y \leqslant 2} \left| \int_0^T \sqrt{k}[T_n\{s_n(x), y\} - R_n\{s_n(x), y\}] - W_R\{s_n(x), y\} \, \mathrm{d}x^{-\gamma_1} \right| > \varepsilon/2 \right) \\ &\quad + P\left[\sup_{1/2 \leqslant y \leqslant 2} \left| \int_0^T W_R\{s_n(x), y\} - W_R(x, y) \, \mathrm{d}x^{-\gamma_1} \right| > \varepsilon/2 \right] \\ &=: p_{31} + p_{32}. \end{aligned}$$

We first consider p_{31}. 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 $\eta_0 \in (\gamma_1, \frac{1}{2})$,

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

By result (14), as $n \to \infty$,

$$\left| \int_0^T s_n(x)^{\eta_0} \, \mathrm{d}x^{-\gamma_1} \right| \to \frac{\gamma_1}{\eta_0 - \gamma_1} T^{\eta_0 - \gamma_1}.$$

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 $\eta_0 \in (\gamma_1, \frac{1}{2})$, there exists $\lambda_0 = \lambda(\eta_0, \varepsilon)$ such that

$$P\{ \sup_{0 < x < \infty, 1/2 \leqslant y \leqslant 2} |W_R(x, y)|/x^{\eta_0} \geqslant \lambda_0 \} \leqslant \varepsilon/3. \tag{25}$$

Moreover, $W_R(x, y)$ is continuous on $(0, \infty) \times [\frac{1}{2}, 2]$; see corollary 1.11 in Adler (1990). Hence applying expressions (25) and (14) with $g = W_R$, $S = T$ and $S_0 = T + 1$, we have that there is an $n_4 = n_4(T)$ such that, for $n > n_4$, $p_{32} < \varepsilon/2$. Thus, inequality (24) 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 $n_0(T_0) = \max_{1 \leqslant j \leqslant 4} n_j(T_0)$. We obtain that, for the chosen T_0 and any $n > n_0$, the three inequalities (22)–(24) hold, which completes the proof of the proposition. □

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

Proposition 3. Under the condition of theorem 1, we have

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

A.4. Proof of proposition 3

Observe that

$$\lim_{n \to \infty} \frac{\theta_{k/n}}{U_1(n/k)} \to \int_0^\infty R(s^{-1/\gamma_1}, 1) \, \mathrm{d}s.$$

Therefore it is sufficient to show that

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

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

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

$$+\left\{\frac{\sqrt{k}}{U_1(n/k)}(e_n\theta_{ke_n/n}-\theta_{k/n})-W_R(\infty,1)(\gamma_1-1)\int_0^\infty R(s^{-1/\gamma_1},1)\,\mathrm{d}s\right\}$$
$$=: J_1+J_2.$$

We prove that both J_1 and J_2 converge to 0 in probability as $n\to\infty$.

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

$$\sqrt{k}(e_n-1)\xrightarrow{\mathrm{P}}-W_R(\infty,1), \tag{26}$$

which implies that

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

Hence, with probability tending to 1,

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

The first part converges to 0 in probability by proposition 2. For the second part, note that, for any $\varepsilon>0$, $0<\delta<1$ and $\eta\in(\gamma_1,\frac{1}{2})$,

$$P\left\{\sup_{|y-1|<k^{-1/4}}\left|\int_0^\infty W_R(s,y)-W_R(s,1)\,\mathrm{d}s^{-\gamma_1}\right|>\varepsilon\right\}\leqslant P\left\{\sup_{|y-1|<k^{-1/4}}\left|\int_0^\delta W_R(s,y)-W_R(s,1)\,\mathrm{d}s^{-\gamma_1}\right|>\frac{\varepsilon}{2}\right\}$$
$$+P\left\{\sup_{|y-1|<k^{-1/4}}\left|\int_\delta^\infty W_R(s,y)-W_R(s,1)\,\mathrm{d}s^{-\gamma_1}\right|>\frac{\varepsilon}{2}\right\}$$
$$\leqslant P\left\{\sup_{0<s\leqslant 1,1/2\leqslant y\leqslant 2}\frac{|W_R(s,y)|}{s^\eta}>\frac{\varepsilon(\eta-\gamma_1)}{4\gamma_1}\delta^{\gamma_1-\eta}\right\}$$
$$+P\left\{\sup_{s>0,|y-1|<k^{-1/4}}|W_R(s,y)-W_R(s,1)|\delta^{-\gamma_1}>\frac{\varepsilon}{2}\right\}.$$
$$=: p_{11}+p_{12}.$$

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

$$\sup_{s>0,\,|y-1|<k^{-1/4}}|W_R(s,y)-W_R(s,1)|\to 0 \qquad \text{almost surely;}$$

see corollary 1.11 in Adler (1990). Hence we have proved that $J_1\to^{\mathrm{P}}0$ as $n\to\infty$.

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

$$\sup_{1/2\leqslant y\leqslant 2}\sqrt{k}\left|\int_0^\infty R_n\{s_n(x),y\}-R(x,y)\,\mathrm{d}x^{-\gamma_1}\right|\to 0. \tag{27}$$

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

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

Hence, as $n\to\infty$,

$$\sup_{1/2\leqslant y\leqslant 2}\sqrt{k}\left|\int_0^\infty R_n\{s_n(x),y\}-R\{s_n(x),y\}\,\mathrm{d}x^{-\gamma_1}\right|\leqslant\sqrt{k}\sup_{\substack{0<x<\infty\\1/2\leqslant y\leqslant 2}}\frac{|R_n(x,y)-R(x,y)|}{x^\beta\wedge 1}\left|\int_0^\infty s_n(x)^\beta\wedge 1\,\mathrm{d}x^{-\gamma_1}\right|$$
$$=O\left\{\sqrt{k}\left(\frac{n}{k}\right)^\tau\right\}\left\{-\int_0^{1/2}s_n(x)^\beta\,\mathrm{d}x^{-\gamma_1}-\int_{1/2}^\infty 1\,\mathrm{d}x^{-\gamma_1}\right\}$$
$$\to 0.$$

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

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

which is a consequence of result (14).

To complete the proof of relationship (27), it is still necessary to show that, as $n\to\infty$,

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

This is achieved by applying result (15) to the R-function which satisfies the Lipschitz condition: $|R(x_1, y) - R(x_2, y)| \leqslant |x_1 - x_2|$, for $x_1, x_2, y \geqslant 0$. Hence, we have proved relationship (27).

Combining expressions (21) and (27), we obtain that

$$\frac{\theta_{k/n}}{U_1(n/k)} = -\int_0^{\infty} R\{s_n(x), 1\}\,\mathrm{d}x^{-\gamma_1} = -\int_0^{\infty} R(x, 1)\,\mathrm{d}x^{-\gamma_1} + o\left(\frac{1}{\sqrt{k}}\right), \tag{28}$$

and

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

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

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

Hence, we obtain that

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

By applying result (26), proposition 1 and Cramér's delta method, we obtain that, as $n\to\infty$,

$$\begin{aligned}\frac{\sqrt{k}}{U_1(n/k)}(e_n\theta_{ke_n/n} - \theta_{k/n}) &= \sqrt{k}(e_n^{1-\gamma_1} - 1)\frac{\theta_{k/n}}{U_1(n/k)} + o_P(1)\\ &\xrightarrow{\mathrm{P}} (\gamma_1 - 1)\,W_R(\infty, 1)\int_0^{\infty} R(s^{-1/\gamma_1}, 1)\,\mathrm{d}s.\end{aligned}$$

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

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

A.5. Proof of theorem 1

Write

$$\frac{\hat{\theta}_p}{\theta_p} = \frac{d_n^{\hat{\gamma}_1}}{d_n^{\gamma_1}}\,\frac{\hat{\theta}_{k/n}}{\theta_{k/n}}\,\frac{d_n^{\gamma_1}\theta_{k/n}}{\theta_p} =: L_1L_2L_3.$$

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 $n\to\infty$,

$$\sqrt{k_1}(\hat{\gamma}_1 - \gamma_1) \xrightarrow{\mathrm{P}} \Gamma. \tag{29}$$

The proof follows lines similar to those in 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

$$\frac{\sqrt{k_1}}{\log(d_n)}(L_1 - 1) - \Gamma \xrightarrow{\mathrm{P}} 0. \tag{30}$$

Secondly, the asymptotic behaviour of the factor L_2 is given by proposition 3.

Lastly, for L_3, by condition (b), we have that, as $n\to\infty$,

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

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

$$\frac{U_1(1/p)}{U_1(n/k)d_n^{\gamma_1}}-1=o\left(\frac{1}{\sqrt{k}}\right). \tag{31}$$

Following the same reasoning as for equation (28) for $p\leqslant k/n$, we have

$$\frac{\theta_p}{U_1(1/p)}-\int_0^\infty R(s^{-1/\gamma_1},1)\,\mathrm{d}s=o\left(\frac{1}{\sqrt{k}}\right).$$

Combining this with equation (31), we have

$$L_3=\frac{\theta_{k/n}/U_1(n/k)}{\theta_p/U_1(1/p)}\,\frac{U_1(n/k)d_n^{\gamma_1}}{U_1(1/p)}=1+o\left(\frac{1}{\sqrt{k}}\right). \tag{32}$$

Combining the asymptotic relationships (30), (32) and proposition 3, we obtain that

$$\begin{aligned}\frac{\hat\theta_p}{\theta_p}-1&=L_1L_2L_3-1\\&=\left[1+\frac{\log(d_n)}{\sqrt{k_1}}\Gamma+o_P\left\{\frac{\log(d_n)}{\sqrt{k_1}}\right\}\right]\left\{1+\frac{\Theta}{\sqrt{k}}+o_P\left(\frac{1}{\sqrt{k}}\right)\right\}\left\{1+o\left(\frac{1}{\sqrt{k}}\right)\right\}-1\\&=\frac{\log(d_n)}{\sqrt{k_1}}\Gamma+\frac{\Theta}{\sqrt{k}}+o_P\left\{\frac{\log(d_n)}{\sqrt{k_1}}\right\}+o_P\left(\frac{1}{\sqrt{k}}\right).\end{aligned}$$

The covariance matrix of (Θ,Γ) follows from the straightforward calculation.

A.6. Proof of theorem 2

Write $\theta_p^+:=E\{X^+|Y>U_2(1/p)\}$. Then,

$$\frac{\hat\theta_p}{\theta_p}=\frac{\hat\theta_p}{\theta_p^+}\,\frac{\theta_p^+}{\theta_p}.$$

Hence, it suffices to prove that $\hat\theta_p/\theta_p^+$ follows the asymptotic normality stated in theorem 1 and $\theta_p/\theta_p^+=1+o(1/\sqrt{k})$.

We first show that (X^+,Y) satisfies conditions (a) and (b) of Section 2.1. Let $\tilde F_1$ be the distribution function of X^+ and

$$\tilde U_1=\left(\frac{1}{1-\tilde F_1}\right)^{\leftarrow}.$$

It is obvious that $U_1(t)=\tilde U_1(t)$, for $t>1/\{1-F_1(0)\}$. Hence X^+ satisfies condition (b).

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

$$t\,P\{X<0,1-F_2(Y)<1/t\}=O(t^\tau). \tag{33}$$

Observe that condition (a) implies that

$$\sup_{1/2\leqslant y\leqslant 2}|y-R(t,y)|=O(t^\tau) \tag{34}$$

and hence $1-R(ct,1)=O(t^\tau)$ for any $c\in(0,\infty)$. Now equation (33) follows:

$$\begin{aligned}t\,P\{X<0,1-F_2(Y)<1/t\}&=1-t\,P\{X>0,1-F_2(Y)<1/t\}\\&=1-t\,P\{1-F_1(X)<1-F_1(0),1-F_2(Y)<1/t\}\\&\leqslant 1-R[t\{1-F_1(0)\},1]+|t\,P\{1-F_1(X)<1-F_1(0),1-F_2(Y)<1/t\}\\&\quad-R[t\{1-F_1(0)\},1]|\\&=O(t^\tau).\end{aligned}$$

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

$$\sup_{\substack{0<x<\infty \\ 1/2 \leqslant y \leqslant 2}} \frac{|t\,P\{1-\tilde{F}_1(X^+)<x/t, 1-F_2(Y)<y/t\} - R(x,y)|}{x^\beta \wedge 1} = O(t^\tau). \tag{35}$$

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

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

Hence, the uniform convergence (in equation (35)) on $(0, t\{1-F_1(0)\}] \times [\frac{1}{2}, 2]$ follows from the fact that (X, Y) satisfies condition (a). Secondly, for $x > t\{1-F_1(0)\}$, we have $1-\tilde{F}_1(X^+) < x/t$. Therefore,

$$\begin{aligned}\sup_{\substack{t\{1-F_1(0)\}<x<\infty \\ 1/2 \leqslant y \leqslant 2}} |t\,P\{1-\tilde{F}_1(X^+)<x/t, 1-F_2(Y)<y/t\} - R(x,y)| &= \sup_{\substack{t\{1-F_1(0)\}<x<\infty \\ 1/2 \leqslant y \leqslant 2}} \{y - R(x,y)\} \\ &\leqslant \sup_{1/2 \leqslant y \leqslant 2} (y - R[t\{1-F_1(0)\}, y]) = O(t^\tau),\end{aligned}$$

where the last relationship follows from equation (34). This completes the verification of equation (35). As a result, theorem 1 applies to $\hat{\theta}_p/\theta_p^+$.

Next we show that $\theta_p/\theta_p^+ = 1 + o(1/\sqrt{k})$. By proposition 1,

$$\frac{\theta_p^+}{U_1(1/p)} \to \int_0^\infty R(x^{-1/\gamma_1}, 1)\,\mathrm{d}x.$$

By Hölder's inequality, condition (e) and equation (33),

$$\begin{aligned}-E\left\{X^- | Y > U_2\left(\frac{1}{p}\right)\right\} &= -\frac{1}{p} E\left[X^- I\left\{X<0, Y>U_2\left(\frac{1}{p}\right)\right\}\right] \\ &\leqslant \frac{1}{p}(E|X^-|^{1/\gamma_1})^{\gamma_1} P\left\{X<0, Y>U_2\left(\frac{1}{p}\right)\right\}^{1-\gamma_1} \\ &= O(p^{-1+(1-\tau)(1-\gamma_1)}).\end{aligned}$$

By applying inequality (19) for $t, s > t_1$, we obtain that

$$\begin{aligned}|t^{-\gamma_1} U_1(t) - s^{-\gamma_1} U_1(s)| &\leqslant |t^{-\gamma_1} U_1(t) - t_1^{-\gamma_1} U_1(t_1)| + |s^{-\gamma_1} U_1(s) - t_1^{-\gamma_1} U_1(t_1)| \\ &< 2c_2 t_1^{-\gamma_1} U_1(t_1)|A_1(t_1)| < \varepsilon_0,\end{aligned}$$

for sufficiently large t_1. Here we use the fact that $t^{-\gamma_1} U_1(t)|A_1(t)|$ is regularly varying with index $\rho_1 < 0$. Thus, as $t \to \infty$, the limit of $t^{-\gamma_1} U(t)$ is finite. In particular, $1/U_1(1/p) = O(p^{\gamma_1})$, as $p \downarrow 0$. Hence, by condition (f),

$$\begin{aligned}\frac{\theta_p}{\theta_p^+} &= 1 + \frac{E\{X^- | Y > U_2(1/p)\}}{\theta_p^+} = 1 + O\left\{\frac{p^{-1+(1-\tau)(1-\gamma_1)}}{U_1(1/p)}\right\} \\ &= 1 + O(p^{-\tau(1-\gamma_1)}) = 1 + o\left(\frac{1}{\sqrt{k}}\right).\end{aligned}$$

References

Acharya, V. V., Pedersen, L. H., Philippon, T. and Richardson, M. (2012) Measuring systemic risk. *Discussion Paper DP8824*. Centre for Economic Policy Research, London.

Adler, R. J. (1990) *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Hayward: Institute of Mathematical Statistics.

Alexander, K. S. (1986) Sample moduli for set-indexed Gaussian processes. *Ann. Probab.*, **14**, 598–611.

Asimit, A. V., Furman, E., Tang, Q. and Vernic, R. (2011) Asymptotics for risk capital allocations based on conditional tail expectation. *Insur. Math. Econ.*, **49**, 310–324.

Bargès, M., Cossette, H. and Marceau, É. (2009) TVaR-based capital allocation with copulas. *Insur. Math. Econ.*, **45**, 348–361.

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

Brownlees, C. and Engle, R. (2012) Volatility, correlation and tails for systemic risk measurement. *Preprint*. (Available from http://ssrn.com/abstract=1611229.)

Cai, J. and Li, H. (2005) Conditional tail expectations for multivariate phase-type distributions. *J. Appl. Probab.*, **42**, 810–825.

Cousin, A. and Di Bernardino, E. (2013) On multivariate extensions of value-at-risk. *J. Multiv. Anal.*, **119**, 32–46.

Drees, H. (2000) Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.*, **10**, 1274–1301.

Drees, H. and Huang, X. (1998) Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multiv. Anal.*, **64**, 25–47.

Einmahl, J. H. J. (1987) *Multivariate Empirical Processes*. Amsterdam: Mathematisch Centrum.

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

El Methni, J., Gardes, L. and Girard, S. (2014) Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scand. J. Statist.*, to be published.

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

Hill, B. M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.

Hua, L. and Joe, H. (2011) Second order regular variation and conditional tail expectation of multiple risks. *Insur. Math. Econ.*, **49**, 537–546.

Joe, H. and Li, H. (2011) Tail risk of multivariate regular variation. *Methodol. Comput. Appl. Probab.*, **13**, 671–693.

Kostadinov, K. (2006) Tail approximation for credit risk portfolios with heavy-tailed risk factors. *J. Risk*, **8**, 81–107.

Landsman, Z. M. and Valdez, E. A. (2003) Tail conditional expectations for elliptical distributions. *Nth Am. Act. J.*, **7**, 55–71.

Vernic, R. (2006) Multivariate skew-normal distributions with applications in insurance. *Insur. Math. Econ.*, **38**, 413–426.

Weissman, I. (1978) Estimation of parameters and larger quantiles based on the k largest observations. *J. Am. Statist. Ass.*, **73**, 812–815.