



Improved estimation of the extreme value index using related variables

Hanan Ahmed¹ · John H. J. Einmahl¹ 

Received: 24 August 2018 / Accepted: 16 July 2019 /
Published online: 1 August 2019
© The Author(s) 2019

Abstract

Heavy tailed phenomena are naturally analyzed by extreme value statistics. A crucial step in such an analysis is the estimation of the extreme value index, which describes the tail heaviness of the underlying probability distribution. We consider the situation where we have next to the n observations of interest another $n+m$ observations of one or more related variables, like, e.g., financial losses due to earthquakes and the related amounts of energy released, for a longer period than that of the losses. Based on such a data set, we present an adapted version of the Hill estimator. For this adaptation the tail dependence between the variable of interest and the related variable(s) plays an important role. We establish the asymptotic normality of this new estimator. It shows greatly improved behavior relative to the Hill estimator, in particular the asymptotic variance is substantially reduced, whereas we can keep the asymptotic bias the same. A simulation study confirms the substantially improved performance of our adapted estimator. We also present an application to the aforementioned earthquake losses.

Keywords Asymptotic normality · Heavy tail · Hill estimator · Tail dependence · Variance reduction

AMS 2000 Subject Classifications Primary—62G32 · 62G05 · 62G20 · 62P05;
Secondary—60F05 · 60G70

1 Introduction

Consider univariate extreme value theory for heavy tails, that is, the case where the extreme value index γ is positive. This index describes the tail heaviness of the underlying probability distribution, the larger γ , the heavier the tail. See de Haan

✉ John H. J. Einmahl
J.H.J.Einmahl@uvt.nl

¹ Department of Econometrics & OR and CentER, Tilburg University, Tilburg, Netherlands

and Ferreira (2006) or Beirlant et al. (2004) for a comprehensive introduction to univariate and multivariate extreme value theory and Gomes and Guillou (2015) for a more recent review of the univariate case. Given a random sample, we often estimate γ with the well-known (Hill 1975) estimator. Such an estimate of γ is the crucial ingredient for estimating important tail functionals of the distribution, like very high quantiles, very small tail probabilities, but also the expected shortfall or an excess-of-loss reinsurance premium.

In this paper we consider the situation where we have a bivariate (or multivariate) data set, where the first component is the variable of interest with extreme value index γ_1 and the second component is a heavy-tailed related variable, with extreme value index γ_2 , that should help to improve the estimation of γ_1 . We assume that we have a random sample of size n of these pairs and in addition another m observations of the second component; these m observations are independent of the pairs and mutually independent. Hence we have a larger sample of the related variable than that of the variable of interest. Such a situation occurs in, e.g., an insurance setting when we have recorded both variables for a certain period of time (2008–2017, say), but in addition have data for the second variable only, for an earlier period (1980–2007, say). We can think of financial losses as the variable of interest and some physical quantity (like wind speed, air pressure, earthquake magnitude, water height) as the related variable (see Section 5). The independence assumption between the n pairs and m earlier observations is then naturally fulfilled. Specifically, the situation with hurricane losses as variable of interest and (transformed) air pressures as related variable was brought to our attention by a reinsurance company. A related situation where our setup can occur is when in a certain period the related variable is measured more frequently than the variable of interest. Also in a cross-sectional context our setup can be relevant. E.g., in a medical setting it often happens that for a group of n patients a specific variable is measured together with one (or more) other variable(s), whereas for a (larger) group of m patients, due to cost constraints, the specific variable is not measured, only the related variable(s), see Chakraborty and Cai (2018). We will also, as suggested in the medical setting, consider the situation where there is *more* than one related variable, the multivariate case, but in this introductory section we will focus on the bivariate case.

We can estimate γ_1 with the Hill estimator $\hat{\gamma}_1$ and γ_2 with the Hill estimators $\hat{\gamma}_2$, based on the n data, and $\hat{\gamma}_{2+}$, based on all $n + m$ data. The latter estimator is better than $\hat{\gamma}_2$, “hence” their difference can be used to update and improve $\hat{\gamma}_1$. For this updating the strength of the tail dependence between both variables is important and should be estimated. A detailed derivation of our adapted Hill estimator is presented in the next section. We will show that our estimator improves greatly on the Hill estimator, in particular the asymptotic variance is substantially reduced, whereas we can keep the asymptotic bias the same. To the best of our knowledge this approach is novel and there are no results of this type in the literature.

The remainder of this paper is organized as follows. In Section 2, for the clearness of the exposition, the bivariate case is treated as indicated above and the asymptotic normality of the adapted estimator is established and in Section 3 the corresponding results for the multivariate case are presented. In Section 4, the finite sample performance of our estimator is studied through a simulation study, which confirms

the improved performance of the adapted Hill estimator. In Section 5, we present an application to earthquake damage amounts with the “amount of energy released” as related variable. The proofs of the results in Section 3 are deferred to Section 6. Since Section 3 generalizes Section 2, the proofs of Section 2 can be obtained by specializing those of Section 3, and are hence omitted.

2 Main results: the bivariate case

Let F be a bivariate distribution function with marginals F_1 and F_2 . Assume that F is in the bivariate max-domain of attraction (i.e., $F \in D(G)$) with both extreme value indices γ_1 and γ_2 positive, see Chapter 6 in de Haan and Ferreira (2006). Let $U_j(\cdot) = F_j^{-1}(1 - 1/\cdot)$ be the tail quantile corresponding to $F_j, j = 1, 2$. Then $F \in D(G)$ with positive extreme value indices implies that U_j is regularly varying with index $\gamma_j, j = 1, 2$, i.e., $\lim_{t \rightarrow \infty} U_j(tx)/U_j(t) = x^{\gamma_j}, x > 0$. Let (X_1, Y_1) have distribution function F . Then $F \in D(G)$ also implies the existence of the tail copula R defined by

$$R(x, y) = \lim_{t \downarrow 0} \frac{1}{t} \mathbb{P}(1 - F_1(X_1) \leq tx, 1 - F_2(Y_1) \leq ty), \quad (x, y) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}. \tag{1}$$

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a bivariate random sample from F , and let Y_{n+1}, \dots, Y_{n+m} be a univariate random sample from F_2 , independent from the n pairs. Denote the order statistics of the $X_i, i = 1, \dots, n$, with $X_{1,n} \leq \dots \leq X_{n,n}$ and use similar notation for the order statistics of the $Y_i, i = 1, \dots, n$, and also for the order statistics of all the $Y_i, i = 1, \dots, n + m$. For $k \in \{1, \dots, n - 1\}$ define the (Hill 1975) estimator of γ_1 by

$$\hat{\gamma}_1 = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}. \tag{2}$$

Define, using the same k , similarly the Hill estimator $\hat{\gamma}_2$ based on the $Y_i, i = 1, \dots, n$:

$$\hat{\gamma}_2 = \frac{1}{k} \sum_{i=0}^{k-1} \log Y_{n-i,n} - \log Y_{n-k,n}. \tag{3}$$

Also, let $\hat{\gamma}_{2+}$ be the Hill estimator of all $Y_i, i = 1, \dots, n + m$, with k replaced by $k_+ \in \{k + 1, \dots, n + m - 1\}$:

$$\hat{\gamma}_{2+} = \frac{1}{k_+} \sum_{i=0}^{k_+-1} \log Y_{n+m-i,n+m} - \log Y_{n+m-k_+,n+m}. \tag{4}$$

Throughout for the asymptotical theory we will assume that $m = m(n)$ and that

$$k \rightarrow \infty, \quad \frac{k}{n} \rightarrow 0, \quad \sqrt{\frac{k}{k_+}} \rightarrow v \in (0, 1), \quad \frac{n}{n+m} \frac{k_+}{k} \rightarrow \beta \in (0, 1], \quad \text{as } n \rightarrow \infty. \tag{5}$$

Observe that we now also have $k_+ \rightarrow \infty, m \rightarrow \infty,$ and $k_+/(n + m) \rightarrow 0$; actually $n/(n + m) \rightarrow \beta v^2 \in (0, 1)$.

First we consider the joint asymptotic normality of the three Hill estimators $\hat{\gamma}_1 = \hat{\gamma}_1(k), \hat{\gamma}_2 = \hat{\gamma}_2(k),$ and $\hat{\gamma}_{2+} = \hat{\gamma}_{2+}(k_+)$. For this, we need the usual second order conditions, on F_1 and F_2 : there exist positive or negative functions $A_j, j = 1, 2,$ with $\lim_{t \rightarrow \infty} A_j(t) = 0,$ such that for $x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{U_j(tx)}{U_j(t)} - x^{\rho_j}}{A_j(t)} = x^{\rho_j} \frac{x^{\rho_j} - 1}{\rho_j}, \quad \text{for some } \rho_j \leq 0, j = 1, 2. \tag{6}$$

Proposition 2.1 *If $F \in D(G),$ conditions Eqs. 5 and 6 hold, and $\sqrt{k}A_j(\frac{n}{k}) \rightarrow \lambda_j \in \mathbb{R}, j = 1, 2,$ as $n \rightarrow \infty,$ then*

$$\left(\sqrt{k}(\hat{\gamma}_1 - \gamma_1), \sqrt{k}(\hat{\gamma}_2 - \gamma_2), \sqrt{k_+}(\hat{\gamma}_{2+} - \gamma_2)\right) \xrightarrow{d} N\left(\left(\frac{\lambda_1}{1 - \rho_1}, \frac{\lambda_2}{1 - \rho_2}, \frac{\lambda_2 \beta^{-\rho_2}}{v(1 - \rho_2)}\right), \check{\Sigma}\right), \tag{7}$$

with, see Eq. 1,

$$\check{\Sigma} = \begin{bmatrix} \gamma_1^2 & R(1, 1)\gamma_1\gamma_2 & vR(1, \beta)\gamma_1\gamma_2 \\ R(1, 1)\gamma_1\gamma_2 & \gamma_2^2 & v\beta\gamma_2^2 \\ vR(1, \beta)\gamma_1\gamma_2 & v\beta\gamma_2^2 & \gamma_2^2 \end{bmatrix}.$$

Corollary 2.1 *Under the conditions of Proposition 2.1, as $n \rightarrow \infty,$*

$$\begin{aligned} &\left(\sqrt{k}(\hat{\gamma}_1 - \gamma_1), \sqrt{k}(\hat{\gamma}_{2+} - \hat{\gamma}_2)\right) \\ &\xrightarrow{d} N\left(\left(\frac{\lambda_1}{1 - \rho_1}, \frac{\lambda_2(\beta^{-\rho_2} - 1)}{1 - \rho_2}\right), \right. \\ &\left. \begin{bmatrix} \gamma_1^2 & (v^2R(1, \beta) - R(1, 1))\gamma_1\gamma_2 \\ (v^2R(1, \beta) - R(1, 1))\gamma_1\gamma_2 & (1 + v^2 - 2v^2\beta)\gamma_2^2 \end{bmatrix}\right). \end{aligned}$$

Corollary 2.1 is the basis for deriving our adapted Hill estimator. For this derivation only, take $\lambda_1 = \lambda_2 = 0.$ The tail copula R is estimated as usual, cf. Drees and Huang (1998), by

$$\hat{R}(x, y) = \frac{1}{k} \sum_{i=1}^n 1_{[X_i \geq X_{n-[kx]+1, n}, Y_i \geq Y_{n-[ky]+1, n}]}, \quad x, y \geq 0. \tag{8}$$

Now consider $(\hat{\gamma}_1, \hat{\gamma}_{2+} - \hat{\gamma}_2)$ and its approximate bivariate normal distribution according to Corollary 2.1, with estimated covariance matrix:

$$N\left((\gamma_1, 0), \frac{1}{k} \begin{bmatrix} \hat{\gamma}_1^2 & \left(\frac{k}{k_+} \hat{R}\left(1, \frac{k_+}{k} \frac{n}{n+m}\right) - \hat{R}(1, 1)\right) \hat{\gamma}_1 \hat{\gamma}_{2+} \\ \left(\frac{k}{k_+} \hat{R}\left(1, \frac{k_+}{k} \frac{n}{n+m}\right) - \hat{R}(1, 1)\right) \hat{\gamma}_1 \hat{\gamma}_{2+} & \left(1 + \frac{k}{k_+} - 2\frac{n}{n+m}\right) \hat{\gamma}_{2+}^2 \end{bmatrix}\right). \tag{9}$$

Maximizing this approximate likelihood of the single observation $(\hat{\gamma}_1, \hat{\gamma}_{2+} - \hat{\gamma}_2)$ with respect to γ_1 , we obtain our adapted estimator for γ_1 :

$$\hat{\gamma}_{1,2} = \hat{\gamma}_1 + \frac{\hat{\gamma}_1}{\hat{\gamma}_{2+}} \left(\frac{\hat{R}(1, 1) - \frac{k}{k_+} \hat{R}(1, \frac{k_+}{k} \frac{n}{n+m})}{1 + \frac{k}{k_+} - 2 \frac{n}{n+m}} \right) (\hat{\gamma}_{2+} - \hat{\gamma}_2). \tag{10}$$

The main result of this section, the asymptotic normality of this estimator, shows that it improves substantially on the Hill estimator.

Theorem 2.1 *Under the conditions of Proposition 2.1, as $n \rightarrow \infty$,*

$$\sqrt{k}(\hat{\gamma}_{1,2} - \gamma_1) \xrightarrow{d} N \left(\frac{\lambda_1}{1 - \rho_1} + \frac{\gamma_1}{\gamma_2} \cdot \frac{R(1, 1) - v^2 R(1, \beta)}{1 + v^2 - 2v^2 \beta} \cdot \frac{\lambda_2(\beta^{-\rho_2} - 1)}{1 - \rho_2}, \right. \\ \left. \gamma_1^2 \left[1 - \frac{(R(1, 1) - v^2 R(1, \beta))^2}{1 + v^2 - 2v^2 \beta} \right] \right).$$

Remark 1 Note that in case $\rho_1 \neq \rho_2$, we have, since $|A_j|$ is regularly varying at ∞ with index ρ_j , $j = 1, 2$, that $\lambda_1 = 0$ or $\lambda_2 = 0$. Hence in this case the expression for the asymptotic bias is simplified. In case $\lambda_2 = 0$ (which is implied by $\rho_1 > \rho_2$) or $\beta = 1$ or $\rho_2 = 0$, the Hill estimator and the adapted estimator have the same asymptotic bias $\lambda_1/(1 - \rho_1)$.

We highlight the natural choice $\beta = 1$ in the following corollary.

Corollary 2.2 *Under the conditions of Proposition 2.1 with $\beta = 1$, as $n \rightarrow \infty$,*

$$\sqrt{k}(\hat{\gamma}_{1,2} - \gamma_1) \xrightarrow{d} N \left(\frac{\lambda_1}{1 - \rho_1}, \gamma_1^2 \left[1 - (1 - v^2)R^2(1, 1) \right] \right).$$

Remark 2 Since the asymptotic biases of both estimators are the same now, we can in the comparison focus on the asymptotic variances. Clearly the asymptotic variance of the adapted Hill estimator never exceeds the γ_1^2 of the classical Hill estimator. The (relative) variance reduction is equal to $(1 - v^2)R^2(1, 1)$, which is positive in case of tail dependence, i.e., $R(1, 1) > 0$. When, e.g., $m = n$ and $k_+ = 2k$, this becomes $\frac{1}{2}R^2(1, 1)$. Then, depending on the value of $R(1, 1) \in [0, 1]$, the variance reduction can be as large as 50%. In case of tail independence ($R(1, 1) = 0$), the estimators have the same asymptotic variances. In such a case a “better” related variable should be looked for.

Remark 3 It is well-known that choosing a good k is a difficult problem in extreme value theory. We will not address this problem here, but compare for many values of k our adapted estimator and the Hill estimator, see Remark 2, Remark 4, and the simulation section. On the other hand, there are many methods for choosing the k of

the Hill estimator, see, e.g., Caeiro and Gomes (2015). If one of these methods for the Hill estimator itself is used, we can choose the same k for our adapted estimator and obtain the discussed improvements.

3 Main results: the multivariate case

Now we consider a d -variate distribution function F , with marginals F_1, \dots, F_d and corresponding tail quantile functions $U_j, j = 1, \dots, d$; write F_- for the distribution function of the last $d - 1$ components of a random vector with distribution function F . We assume that F is in the multivariate max-domain of attraction, that is $F \in D(G)$, with all extreme value indices $\gamma_1, \dots, \gamma_d$ positive. Let R_{ij} be the tail copula of the i -th and the j -th component, $1 \leq i, j \leq d, i \neq j$, see Eq. 1.

Let $(X_1, Y_{1,2}, \dots, Y_{1,d}), \dots, (X_n, Y_{n,2}, \dots, Y_{n,d})$, be a d -variate random sample from F and let $(Y_{n+1,2}, \dots, Y_{n+1,d}), \dots, (Y_{n+m,2}, \dots, Y_{n+m,d})$ be a $(d - 1)$ -variate random sample from F_- , independent of the d -variate random sample of size n . Let $\hat{\gamma}_1, \hat{\gamma}_j$, and $\hat{\gamma}_{j+}$ be the Hill estimators based on $X_1, \dots, X_n, Y_{1,j}, \dots, Y_{n,j}$, and $Y_{1,j}, \dots, Y_{n+m,j}, j = 2, \dots, d$, respectively, cf. Eqs. 2, 3, and 4; here again we replace k with k_+ for $\hat{\gamma}_{j+}, j = 2, \dots, d$. First we consider the joint asymptotic normality of all the $2d - 1$ Hill estimators.

Proposition 3.1 *If $F \in D(G)$, condition Eq. 5 holds, condition Eq. 6 holds for $j = 1, \dots, d$, and $\sqrt{k}A_j(\frac{n}{k}) \rightarrow \lambda_j \in \mathbb{R}, j = 1, \dots, d$, as $n \rightarrow \infty$, then*

$$\left(\sqrt{k}(\hat{\gamma}_1 - \gamma_1), \sqrt{k}(\hat{\gamma}_2 - \gamma_2), \sqrt{k_+}(\hat{\gamma}_{2+} - \gamma_2), \dots, \sqrt{k}(\hat{\gamma}_d - \gamma_d), \sqrt{k_+}(\hat{\gamma}_{d+} - \gamma_d)\right) \xrightarrow{d} N(\check{\mu}_d, \check{\Sigma}_d), \tag{11}$$

where

$$\check{\mu}_d = \left(\frac{\lambda_1}{1 - \rho_1}, \frac{\lambda_2}{1 - \rho_2}, \frac{\lambda_2 \beta^{-\rho_2}}{v(1 - \rho_2)}, \dots, \frac{\lambda_d}{1 - \rho_d}, \frac{\lambda_d \beta^{-\rho_d}}{v(1 - \rho_d)} \right),$$

$$\check{\Sigma}_d = \begin{bmatrix} \gamma_1^2 & R_{12}(1, 1)\gamma_1\gamma_2 & vR_{12}(1, \beta)\gamma_1\gamma_2 & \dots & R_{1d}(1, 1)\gamma_1\gamma_d & vR_{1d}(1, \beta)\gamma_1\gamma_d \\ R_{12}(1, 1)\gamma_1\gamma_2 & \gamma_2^2 & v\beta\gamma_2^2 & \dots & R_{2d}(1, 1)\gamma_2\gamma_d & vR_{2d}(1, \beta)\gamma_2\gamma_d \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ R_{1d}(1, 1)\gamma_1\gamma_d & R_{2d}(1, 1)\gamma_2\gamma_d & vR_{2d}(1, \beta)\gamma_2\gamma_d & \dots & \gamma_d^2 & v\beta\gamma_d^2 \\ vR_{1d}(1, \beta)\gamma_1\gamma_d & vR_{2d}(1, \beta)\gamma_2\gamma_d & R_{2d}(1, 1)\gamma_2\gamma_d & \dots & v\beta\gamma_d^2 & \gamma_d^2 \end{bmatrix}.$$

Corollary 3.1 *Under the conditions of Proposition 3.1, as $n \rightarrow \infty$,*

$$\left(\sqrt{k}(\hat{\gamma}_1 - \gamma_1), \sqrt{k}(\hat{\gamma}_{2+} - \gamma_2), \dots, \sqrt{k}(\hat{\gamma}_{d+} - \gamma_d)\right) \xrightarrow{d} N(\mu_d, \Sigma_d), \tag{12}$$

where $\mu_d = \left(\frac{\lambda_1}{1-\rho_1}, \frac{\lambda_2(\beta^{-\rho_2}-1)}{1-\rho_2}, \dots, \frac{\lambda_d(\beta^{-\rho_d}-1)}{1-\rho_d} \right)$, $\Sigma_d = \Gamma\Gamma^T \circ H$ ("o" denotes the Hadamard or entrywise product), with

$$H = \begin{bmatrix} 1 & h_{12} & \dots & h_{1d} \\ h_{12} & h & \dots & h_{2d} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ h_{1d} & h_{2d} & \dots & h \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \gamma_d \end{bmatrix},$$

$h = 1 + v^2 - 2v^2\beta$, $h_{1i} = v^2R_{1i}(1, \beta) - R_{1i}(1, 1)$, and $h_{ij} = (1 + v^2)R_{ij}(1, 1) - v^2(R_{ij}(1, \beta) + R_{ij}(\beta, 1))$, $i = 2, \dots, d$, $j = i + 1, \dots, d$.

Very similar to the the bivariate case we approximate, for $\lambda_j = 0$, $j = 1, \dots, d$, the d -variate normal limiting distribution of $(\hat{\gamma}_1, \hat{\gamma}_{2+} - \hat{\gamma}_2, \dots, \hat{\gamma}_{d+} - \hat{\gamma}_d)$, with mean vector $(\gamma_1, 0, \dots, 0)$, and estimate the approximated $\frac{1}{k}\Sigma_d$, where for the estimation of R_{ij} , \hat{R}_{ij} is defined similarly as \hat{R} in Eq. 8. The thus obtained approximated and estimated version of $\frac{1}{k}\Sigma_d$ is denoted by $\frac{1}{k}\hat{\Sigma}_d$. In this normal distribution the only unknown parameter is the first component of the mean: γ_1 , cf. Eq. 9. Maximizing this approximate likelihood of the single observation $(\hat{\gamma}_1, \hat{\gamma}_{2+} - \hat{\gamma}_2, \dots, \hat{\gamma}_{d+} - \hat{\gamma}_d)$ with respect to γ_1 , we obtain our adapted estimator for γ_1 :

$$\hat{\gamma}_{1,d} = \hat{\gamma}_1 + \sum_{j=2}^d \frac{\hat{\Sigma}_{1j}^{-1}}{\hat{\Sigma}_{11}^{-1}} (\hat{\gamma}_{j+} - \hat{\gamma}_j),$$

where A_{ij}^{-1} denotes the entry in the i^{th} row and j^{th} column of the inverse of the matrix A . Using, in the obvious notation, $\hat{\Sigma}_d = \hat{\Gamma}\hat{\Gamma}^T \circ \hat{H}$ (see above), we can rewrite our adapted estimator as

$$\hat{\gamma}_{1,d} = \hat{\gamma}_1 + \sum_{j=2}^d \frac{\hat{\gamma}_1}{\hat{\gamma}_{j+}} \frac{\hat{H}_{1j}^{-1}}{\hat{H}_{11}^{-1}} (\hat{\gamma}_{j+} - \hat{\gamma}_j). \tag{13}$$

Theorem 3.1 Assume H is invertible. Then under the conditions of Proposition 3.1, as $n \rightarrow \infty$,

$$\sqrt{k}(\hat{\gamma}_{1,d} - \gamma_1) \xrightarrow{d} N \left(\frac{\lambda_1}{1 - \rho_1} + \sum_{j=2}^d \frac{\gamma_1}{\gamma_j} \frac{H_{1j}^{-1}}{H_{11}^{-1}} \frac{\lambda_j(\beta^{-\rho_j} - 1)}{1 - \rho_j}, \sigma^2 \right), \tag{14}$$

where

$$\sigma^2 = \gamma_1^2 \left(1 - \frac{1}{(H_{11}^{-1})^2} \left[2H_{11}^{-1} \sum_{j=2}^d [R_{1j}(1, 1) - v^2 R_{1j}(1, \beta)] H_{1j}^{-1} - [1 + v^2 - 2v^2 \beta] \sum_{j=2}^d (H_{1j}^{-1})^2 - 2 \sum_{i=2}^d \sum_{j>i}^d [(1 + v^2) R_{ij}(1, 1) - v^2 (R_{ij}(1, \beta) + R_{ij}(\beta, 1))] H_{1i}^{-1} H_{1j}^{-1} \right] \right).$$

Corollary 3.2 Under the conditions of Theorem 3.1 with $\beta = 1$, as $n \rightarrow \infty$,

$$\sqrt{k}(\hat{\gamma}_{1,d} - \gamma_1) \xrightarrow{d} N \left(\frac{\lambda_1}{1 - \rho_1}, \sigma^2 \right),$$

where the asymptotic variance now simplifies to

$$\sigma^2 = \gamma_1^2 \left(1 - \frac{1 - v^2}{(H_{11}^{-1})^2} \left[2H_{11}^{-1} \sum_{j=2}^d R_{1j}(1, 1) H_{1j}^{-1} - \sum_{j=2}^d (H_{1j}^{-1})^2 - 2 \sum_{i=2}^d \sum_{j>i}^d R_{ij}(1, 1) H_{1i}^{-1} H_{1j}^{-1} \right] \right),$$

where $\beta = 1$ also yields simplified entries for the matrix H .

Remark 4 We have seen that in the bivariate case for $\beta = 1$ and $v^2 = \frac{1}{2}$ the reduction in asymptotic variance is equal to $\frac{1}{2} R^2(1, 1)$. For, e.g., $R(1, 1) = 0.8$, this becomes 0.320. Now consider the trivariate case with the same values for β and v^2 and with (also) $R_{12}(1, 1) = R_{13}(1, 1) = 0.8$, but $R_{23}(1, 1) = 0.4$. Then the reduction in asymptotic variance, see the next section, becomes much larger: 0.457. In other words, adding a third variable that has the same (as the second variable) tail copula value at (1,1) with the variable of interest and does not have a high tail dependence with the second variable reduces the asymptotic variance much more than when using only one related variable.

4 Simulation study

In this section we will perform a simulation study in order to compare the finite sample behavior of the adapted estimator and the Hill estimator. We will consider 6 bivariate distributions and 8 trivariate distributions and 3 different pairs (n, m) . Every setting is replicated 10,000 times.

To be precise, we consider the Cauchy distribution restricted to the first quadrant/octant in dimensions $d = 2$ and $d = 3$. This Cauchy density is proportional to

$$(1 + xS^{-1}x^T)^{-(1+d)/2},$$

where the 2×2 or 3×3 scale matrix S has 1 as diagonal elements and s as off-diagonal elements, but when $d = 3$ we take $S_{23} = S_{32} = r$. For s we take the values 0, 0.5, and 0.8, respectively. When $d = 3$ we take $r = s$, but for $s = 0.5$ and $s = 0.8$

Table 1 $R(1, 1)$ -values for the Cauchy distribution

$d = 2$			$d = 3$				
$s = 0$	$s = 0.5$	$s = 0.8$	$s = 0$	$s = 0.5$	$s = 0.5$	$s = 0.8$	$s = 0.8$
			$r = 0$	$r = 0.5$	$r = 0$	$r = 0.8$	$r = 0.3$
0.59	0.67	0.76	0.59	0.68	0.69	0.77	0.81
					0.59		0.63

we also take $r = 0$ and $r = 0.3$, respectively. Approximated $R(1, 1)$ -values are given in Table 1. In the case $r < s$, two values are given; the lower one is $R_{23}(1, 1)$.

We will also consider the bi- and trivariate logistic distribution function with standard Fréchet marginals:

$$F(x_1, \dots, x_d) = \exp \left\{ - \left(x_1^{-1/\theta} + \dots + x_d^{-1/\theta} \right)^\theta \right\}, \quad x_1 > 0, \dots, x_d > 0; \quad d = 2 \text{ or } d = 3.$$

For θ we take the values 0.1, 0.3, and 0.5, respectively. The corresponding $R(1, 1)$ -values are 0.93, 0.77, and 0.59. All γ -values in the simulations are equal to 1.

We use the following values for n, m , and k :

- $n = 1000, m = 500$, and $k = 100$,
- $n = 1000, m = 1000$, and $k = 100$,
- $n = 500, m = 1000$, and $k = 50$.

Then we choose k_+ according to

$$\frac{k}{k_+} = \frac{n}{n + m}. \tag{15}$$

In case $d = 2$, using Eq. 15, our adapted estimator in Eq. 10 specializes to

$$\hat{\gamma}_{1,2} = \hat{\gamma}_1 + \frac{\hat{\gamma}_1}{\hat{\gamma}_{2+}} \hat{R}(1, 1)(\hat{\gamma}_{2+} - \hat{\gamma}_2),$$

Table 2 Empirical variance reduction for the Cauchy distribution

	$d = 2$			$d = 3$				
	$s = 0$	$s = 0.5$	$s = 0.8$	$s = 0$	$s = 0.5$	$s = 0.5$	$s = 0.8$	$s = 0.8$
				$r = 0$	$r = 0.5$	$r = 0$	$r = 0.8$	$r = 0.3$
$n = 1000, m = 500$	10.5%	12.4%	17.3%	12.4%	17.9%	18.9%	21.2%	26.5%
$n = 1000, m = 1000$	15.5%	20.1%	28.9%	21.0%	27.3%	30.5%	31.8%	40.3%
$n = 500, m = 1000$	20.6%	27.7%	38.3%	27.6%	36.4%	38.9%	41.3%	51.4%

Table 3 Empirical variance reduction for the logistic distribution

	$d = 2$			$d = 3$		
	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$
$n = 1000, m = 500$	26.8%	17.4%	8.8%	28.8%	20.7%	13.4%
$n = 1000, m = 1000$	41.1%	27.3%	14.4%	42.0%	31.4%	20.6%
$n = 500, m = 1000$	54.5%	37.4%	21.4%	58.9%	43.1%	27.8%

and the asymptotic variance in Theorem 2.1 becomes $\gamma_1^2 (1 - (1 - v^2)R^2(1, 1))$. When $d = 3$, using Eq. 15, our adapted estimator in Eq. 13 can be rewritten as

$$\hat{\gamma}_{1,3} = \hat{\gamma}_1 + \frac{\hat{\gamma}_1}{\hat{\gamma}_{2+}} \frac{\hat{R}_{12}(1, 1) - \hat{R}_{13}(1, 1)\hat{R}_{23}(1, 1)}{1 - \hat{R}_{23}^2(1, 1)} (\hat{\gamma}_{2+} - \hat{\gamma}_2) + \frac{\hat{\gamma}_1}{\hat{\gamma}_{3+}} \frac{\hat{R}_{13}(1, 1) - \hat{R}_{12}(1, 1)\hat{R}_{23}(1, 1)}{1 - \hat{R}_{23}^2(1, 1)} (\hat{\gamma}_{3+} - \hat{\gamma}_3),$$

and the asymptotic variance in Theorem 3.1 specializes to

$$\sigma^2 = \gamma_1^2 \left(1 - (1 - v^2) \left(\frac{R_{12}^2(1, 1) + R_{13}^2(1, 1) - 2R_{12}(1, 1)R_{13}(1, 1)R_{23}(1, 1)}{1 - R_{23}^2(1, 1)} \right) \right).$$

Tables 2 and 3 show the (empirical percentages of) variance reduction as discussed below Theorems 2.1 and 3.1, and above, based on the 10,000 estimates. We see that the variance reduction ranges from about 10% to more than 50%, that is, our adapted estimator yields much better results than the Hill estimator. A stronger tail dependence between the variable of interest and the related variable(s) yields a larger variance reduction. In case $d = 3$, due to the exchangeability of the components of the logistic distribution, a stronger tail dependence between the variable of interest and the related variables, yields also a stronger tail dependence between the two related variables and hence increasing the dimension from 2 to 3 does not help that much, but in case of the Cauchy distribution with $r < s$ we see a large improvement when adding the third variable. Comparing the numbers in the table with the

Table 4 Empirical MSE reduction for the Cauchy distribution

	$d = 2$			$d = 3$				
	$s = 0$	$s = 0.5$	$s = 0.8$					
				$r = 0$	$r = 0.5$	$r = 0.5$	$r = 0.8$	$r = 0.8$
$n = 1000, m = 500$	10.6%	12.5%	17.3%	12.5%	18.1%	19.0%	21.3%	26.6%
$n = 1000, m = 1000$	15.7%	20.2%	28.9%	21.1%	27.6%	30.6%	31.9%	40.3%
$n = 500, m = 1000$	20.6%	27.8%	38.3%	27.8%	36.4%	38.9%	41.3%	51.3%

Table 5 Empirical MSE reduction for the logistic distribution

	$d = 2$			$d = 3$		
	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$
$n = 1000, m = 500$	26.3%	17.5%	9.1%	28.4%	20.6%	13.4%
$n = 1000, m = 1000$	40.7%	27.0%	14.7%	41.3%	31.0%	20.6%
$n = 500, m = 1000$	55.1%	38.0%	22.6%	59.4%	43.9%	28.7%

(not presented) theoretical asymptotic reductions shows that the empirical numbers are about the same but slightly smaller, partly due to the variability of the tail copula estimators, which does not show up in the asymptotic variance.

Although the asymptotic biases are the same (see Corollaries 2.2 and 3.2), we also present, in order to show the full behavior of the estimator, similar tables for the reduction in mean squared error (MSE). We see in Tables 4 and 5 that considering the MSE instead of the variance yields approximately the same reduction percentages. This shows that indeed our adapted estimator substantially outperforms the Hill estimator.

For every simulation setting we have only taken one value of k . It is of interest to investigate the sensitivity to the choice of k of the variance reduction in Tables 2 and 3. For the two tables below we doubled the value of k compared to the initial settings in Tables 2 and 3. For the choice of k_+ the formula in Eq. 15 is still used, i.e., k_+ is also doubled. Tables 6 and 7 show that this large change in k leads to about the same percentages of variance reduction, in other words, when comparing the adapted Hill estimator and the Hill estimator the choice of k is not so important.

5 Application

We consider financial losses (in US\$) due to earthquakes as variable of interest with the corresponding energy released as related variable. The aim of this application is to assess the tail heaviness of the loss distribution and also to estimate a very high quantile of the losses. We make use of the adapted Hill estimator, since the losses are

Table 6 Empirical variance reduction for the Cauchy distribution

	$d = 2$			$d = 3$				
	$s = 0$	$s = 0.5$	$s = 0.8$	$s = 0$	$s = 0.5$	$s = 0.5$	$s = 0.8$	$s = 0.8$
				$r = 0$	$r = 0.5$	$r = 0$	$r = 0.8$	$r = 0.3$
$n = 1000, m = 500, k = 200$	10.1%	12.4%	17.6%	13.5%	17.3%	18.4%	21.4%	25.0%
$n = 1000, m = 1000, k = 200$	15.2%	20.2%	26.0%	19.4%	25.7%	28.1%	33.5%	38.4%
$n = 500, m = 1000, k = 100$	19.9%	27.3%	36.1%	27.6%	35.4%	36.0%	44.1%	52.7%

Table 7 Empirical variance reduction for the logistic distribution

	$d = 2$			$d = 3$		
	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.1$	$\theta = 0.3$	$\theta = 0.5$
$n = 1000, m = 500, k = 200$	27.2%	19.4%	10.1%	28.5%	21.0%	14.6%
$n = 1000, m = 1000, k = 200$	39.6%	27.5%	14.5%	44.6%	30.8%	20.3%
$n = 500, m = 1000, k = 100$	55.3%	38.0%	21.8%	57.1%	43.6%	26.0%

influenced by the amounts of energy and hence these variables are expected to be tail dependent.

The earthquakes concern 29 countries.¹ The data are provided by the National Oceanic and Atmospheric Administration (NOAA). Ignoring tsunami losses, we consider the financial losses of categories at least “moderate” for the time period from 1993 through 2017. We used linear regression analysis per country for imputation of missing loss values, with “number of deaths due to the earthquake” and “severity of the financial loss” (a categorical variable) as independent variables. We also corrected the financial losses for inflation. The highest loss in the data set is US\$ 36×10^9 . We obtained the related Richter scale magnitude M of the earthquakes for the much longer period 1940 through 2017. (Note that also for the earlier period 1940-1992 the financial loss *categories* are available and again we used only magnitudes with losses at least “moderate”, as for the period 1993-2017.) The energy E released by earthquakes (in megajoules) is given by $E = 2 \times 10^{1.5(M-1)}$; Lay and Wallace (1995).

We have $n = 330$ and $m = 512$. Figure 1 shows a plot of the adapted Hill estimator and the Hill estimator against k , with k_+ based on Eq. 15. We take the average values over the region $k = 40, \dots, 60$ of both estimators, respectively. This yields the average Hill estimate $\hat{\gamma}_1 = 1.504$ and the final average adapted Hill estimate of γ_1 , which is somewhat lower than the average Hill estimate:

$$\hat{\gamma}_{1,2} = 1.465.$$

Both estimates indicate that the loss distribution has a very heavy right tail.

We also estimate the high quantile $F_1^{-1}(1 - p)$ of the loss distribution for $p = \frac{1}{n} = \frac{1}{330}$. This high quantile is estimated as usual (see, e.g., page 138 of de Haan and Ferreira 2006) with

$$X_{n-k,n} \left(\frac{k}{np} \right)^{\hat{\gamma}}$$

where $\hat{\gamma}$ is the Hill estimator or the adapted Hill estimator (and $k = 40, \dots, 60$). This yields for the average high quantile estimate US\$ 130×10^9 when we use the Hill estimates and US\$ 113×10^9 when we use our estimates of γ_1 , which is a reduction of 17 billion dollars. This shows that, from an insurer’s perspective, improved (that

¹Algeria, Burma, Chile, China, Ecuador, El Salvador, Germany, Greece, Haiti, Iceland, India, Indonesia, Iran, Italy, Japan, Mexico, Morocco, Nepal, New Zealand, Nicaragua, Pakistan, Philippines, Russia, Taiwan, Tajikistan, Tanzania, Thailand, Turkey, United States.

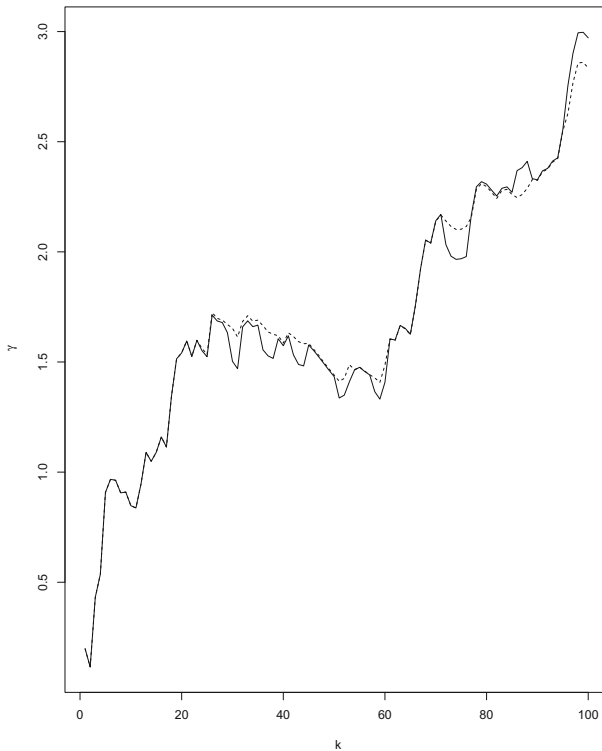


Fig. 1 Adapted Hill estimator (solid line) and Hill estimator (dashed line) of the financial losses of the earthquakes

is, less variable) estimation of the extreme value index can lead to huge changes in high quantiles, here the 25 year return level. The lower estimate we obtain indicates less risk for (re)insurers.

6 Proofs

Proof of Proposition 3.1 Let C be a copula corresponding to the distribution function of $(-X_1, -Y_{1,2}, \dots, -Y_{1,d})$ and let C_- be the distribution function of the last $d - 1$ components of a random vector with distribution function C . Let $(V_{1,1}, V_{1,2}, \dots, V_{1,d}), \dots, (V_{n,1}, V_{n,2}, \dots, V_{n,d})$ be a random sample of size n from C and let $(V_{n+1,2}, \dots, V_{n+1,d}), \dots, (V_{n+m,2}, \dots, V_{n+m,d})$ be a random sample of size m from C_- , independent of the random sample from C . Clearly all the $V_{i,j}$ have a uniform-(0,1) distribution. Write $X_i = F_1^{-1}(1 - V_{i,1})$, $i = 1, \dots, n$, and $Y_{l,j} = F_j^{-1}(1 - V_{l,j})$, $l = 1, \dots, n + m$, $j = 2, \dots, d$. Then $(X_1, Y_{1,2}, \dots, Y_{1,d}), \dots, (X_n, Y_{n,2}, \dots, Y_{n,d})$, and $(Y_{n+1,2}, \dots, Y_{n+1,d}), \dots, (Y_{n+m,2}, \dots, Y_{n+m,d})$ have the distributions as specified in the beginning of Section 3.

Consider the univariate empirical distribution functions $\Gamma_{n,j}(s) = \frac{1}{n} \sum_{i=1}^n 1_{[0,s]}(V_{i,j})$, $0 \leq s \leq 1$, $j = 1, 2, \dots, d$, and $\Gamma_{n+m,j}(t) = \frac{1}{n+m} \sum_{l=1}^{n+m} 1_{[0,t]}(V_{l,j})$, $0 \leq t \leq 1$, $j = 2, \dots, d$, and the corresponding uniform tail empirical processes

$$w_{n,j}(s) = \frac{n}{\sqrt{k}} \left[\Gamma_{n,j} \left(\frac{k}{n} s \right) - \frac{k}{n} s \right], \quad 0 \leq s \leq 1,$$

$$w_{n+m,j}(t) = \frac{n+m}{\sqrt{k_+}} \left[\Gamma_{n+m,j} \left(\frac{k_+}{n+m} t \right) - \frac{k_+}{n+m} t \right], \quad 0 \leq t \leq 1.$$

Now define the Gaussian vector of processes (W_1, \dots, W_{2d-1}) , where W_j , $j = 1, \dots, 2d - 1$, is a standard Wiener process on $[0, 1]$, and the covariances are as follows:

$$\begin{aligned} \text{Cov}(W_i(s), W_j(t)) &= R_{ij}(s, t), \quad 0 \leq s, t \leq 1, \quad 1 \leq i < j \leq d, \\ \text{Cov}(W_i(s), W_j(t)) &= \nu R_{i,j-d+1}(s, \beta t), \quad 0 \leq s, t \leq 1, \quad 1 \leq i \leq d, \quad d+1 \leq j \\ &\leq 2d-1, \quad j \neq i+d-1, \\ \text{Cov}(W_i(s), W_{i+d-1}(t)) &= \nu(s \wedge \beta t), \quad 0 \leq s, t \leq 1, \quad 2 \leq i \leq d. \\ \text{Cov}(W_i(s), W_j(t)) &= R_{i-d+1,j-d+1}(s, t), \quad 0 \leq s, t \leq 1, \quad d+1 \leq i < j \leq 2d-1. \end{aligned} \tag{16}$$

Let I denote the identity function on $[0, 1]$. Then we have on $(D[0, 1])^{2d-1}$, for $0 \leq \delta < \frac{1}{2}$, as $n \rightarrow \infty$,

$$\left(\frac{w_{n,1}}{I^\delta}, \dots, \frac{w_{n,d}}{I^\delta}, \frac{w_{n+m,2}}{I^\delta}, \dots, \frac{w_{n+m,d}}{I^\delta} \right) \xrightarrow{d} \left(\frac{W_1}{I^\delta}, \dots, \frac{W_d}{I^\delta}, \frac{W_{d+1}}{I^\delta}, \dots, \frac{W_{2d-1}}{I^\delta} \right). \tag{17}$$

For the proof of this statement, note that the convergence and tightness of every component is well-known, see Corollary 4.2.1 in Csörgő et al. (1986) or Theorem 3 in Einmahl (1992). This also yields the tightness of the entire vector on the left-hand side. It remains to prove the convergence of the finite-dimensional distributions (without the I^δ), which follows from the (general) multivariate central limit theorem. It suffices to compute the limits of the covariances: we perform this computation for the second formula in Eq. 16; the other three formulas there are essentially special cases of that one. We have

$$\begin{aligned} \text{Cov}(w_{n,i}(s), w_{n+m,j-d+1}(t)) &= \text{Cov} \left(\frac{1}{\sqrt{k}} \sum_{l=1}^n 1_{[0, \frac{k}{n}s]}(V_{l,i}), \frac{1}{\sqrt{k_+}} \sum_{l=1}^{n+m} 1_{[0, \frac{k_+}{n+m}t]}(V_{l,j-d+1}) \right) \\ &= \text{Cov} \left(\frac{1}{\sqrt{k}} \sum_{l=1}^n 1_{[0, \frac{k}{n}s]}(V_{l,i}), \frac{1}{\sqrt{k_+}} \sum_{l=1}^n 1_{[0, \frac{k_+}{n+m}t]}(V_{l,j-d+1}) \right) \\ &= \frac{n}{\sqrt{kk_+}} \text{Cov}(1_{[0, \frac{k}{n}s]}(V_{1,i}), 1_{[0, \frac{k_+}{n+m}t]}(V_{1,j-d+1})) \\ &= \frac{n}{\sqrt{kk_+}} \left[\mathbb{P} \left(V_{1,i} \leq \frac{k}{n}s, V_{1,j-d+1} \leq \frac{k_+}{n+m}t \right) - \frac{kk_+}{n(n+m)}st \right] \\ &= \sqrt{\frac{k}{k_+}} \left[\frac{n}{k} \mathbb{P} \left(V_{1,i} \leq \frac{k}{n}s, V_{1,j-d+1} \leq \frac{k}{n} \frac{k_+}{n+m}t \right) - \frac{k_+}{n+m}st \right] \\ &\rightarrow \nu R_{i,j-d+1}(s, \beta t) = \text{Cov}(W_i(s), W_j(t)). \end{aligned}$$

Hence Eq. 17 is established.

According to de Haan and Ferreira (2006), Chapter 5 and Theorem 2.3.9, we have, as $n \rightarrow \infty$,

$$\sqrt{k}(\hat{\gamma}_j - \gamma_j) = -\gamma_j \left(w_{n,j}(1) - \int_0^1 \frac{w_{n,j}(u)}{u} du \right) + \frac{\lambda_j}{1 - \rho_j} + o_p(1), \quad j = 1, \dots, d.$$

Using that $|A_j|$ is regularly varying at ∞ with index ρ_j , we get similarly

$$\sqrt{k_+}(\hat{\gamma}_{j+} - \gamma_j) = -\gamma_j \left(w_{n+m,j}(1) - \int_0^1 \frac{w_{n+m,j}(u)}{u} du \right) + \frac{\lambda_j \beta^{-\rho_j}}{v(1 - \rho_j)} + o_p(1), \quad j = 2, \dots, d.$$

Combining all these with Eq. 17 we obtain

$$\begin{aligned} & (\sqrt{k}(\hat{\gamma}_1 - \gamma_1), \dots, \sqrt{k}(\hat{\gamma}_d - \gamma_d), \sqrt{k_+}(\hat{\gamma}_{2+} - \gamma_2), \dots, \sqrt{k_+}(\hat{\gamma}_{d+} - \gamma_d)) \\ & \xrightarrow{d} \left(-\gamma_1 \left(W_1(1) - \int_0^1 \frac{W_1(u)}{u} du \right) + \frac{\lambda_1}{1 - \rho_1}, \dots, -\gamma_d \left(W_d(1) - \int_0^1 \frac{W_d(u)}{u} du \right) + \frac{\lambda_d}{1 - \rho_d}, \right. \\ & \left. -\gamma_2 \left(W_{d+1}(1) - \int_0^1 \frac{W_{d+1}(u)}{u} du \right) + \frac{\lambda_2 \beta^{-\rho_2}}{v(1 - \rho_2)}, \dots, -\gamma_d \left(W_{2d-1}(1) - \int_0^1 \frac{W_{2d-1}(u)}{u} du \right) \right. \\ & \left. + \frac{\lambda_d \beta^{-\rho_d}}{v(1 - \rho_d)} \right). \end{aligned}$$

It is immediate and well-known that this yields the mean vector and the variances as in the proposition. (Note that the components of the left-hand side there are listed in a different order.) It remains to derive the covariances. Again we only consider the case where $1 \leq i \leq d, d + 1 \leq j \leq 2d - 1, j \neq i + d - 1$. The other cases are easier and essentially special cases of this one. We have

$$\begin{aligned} & Cov \left(-\gamma_i \left(W_i(1) - \int_0^1 \frac{W_i(u)}{u} du \right) + \frac{\lambda_i}{1 - \rho_i}, -\gamma_{j-d+1} \left(W_j(1) - \int_0^1 \frac{W_j(v)}{v} dv \right) \right. \\ & \left. + \frac{\lambda_{j-d+1} \beta^{-\rho_{j-d+1}}}{v(1 - \rho_{j-d+1})} \right) = \gamma_i \gamma_{j-d+1} \left[E(W_i(1)W_j(1)) + \int_0^1 \int_0^1 \frac{E(W_i(u)W_j(v))}{uv} dudv \right. \\ & \left. - \int_0^1 \frac{E(W_i(u)W_j(1))}{u} du - \int_0^1 \frac{E(W_i(1)W_j(v))}{v} dv \right] = v\gamma_i \gamma_{j-d+1} [R_{i,j-d+1}(1, \beta) \\ & \left. + \int_0^1 \int_0^1 \frac{R_{i,j-d+1}(u, \beta v)}{uv} dudv - \int_0^1 \frac{R_{i,j-d+1}(u, \beta)}{u} du - \int_0^1 \frac{R_{i,j-d+1}(1, \beta v)}{v} dv \right]. \end{aligned}$$

Observe that by two changes of variables and the homogeneity of order 1 of $R_{i,j-d+1}$:

$$\begin{aligned} \int_0^1 \int_0^1 \frac{R_{i,j-d+1}(u, \beta v)}{uv} dudv &= \int_0^1 \int_0^v \frac{R_{i,j-d+1}(u, \beta v)}{uv} dudv + \int_0^1 \int_0^u \frac{R_{i,j-d+1}(u, \beta v)}{uv} dvdu \\ &= \int_0^1 \int_0^1 \frac{R_{i,j-d+1}(vu, \beta v)}{uv} dudv + \int_0^1 \int_0^1 \frac{R_{i,j-d+1}(u, \beta vu)}{uv} dvdu \\ &= \int_0^1 \frac{R_{i,j-d+1}(u, \beta)}{u} du + \int_0^1 \frac{R_{i,j-d+1}(1, \beta v)}{v} dv. \end{aligned}$$

Hence the covariance is equal to $v\gamma_i\gamma_{j-d+1}R_{i,j-d+1}(1, \beta)$. □

Proof of Theorem 3.1 From the uniform consistency of the tail copula estimators and the continuity of the tail copulas we have $\hat{H}_{1j}^{-1} \xrightarrow{P} H_{1j}^{-1}$, $j = 1, \dots, d$. This in combination with Eq. 13 and Corollary 3.1 yields

$$\sqrt{k}(\hat{\gamma}_{1,d} - \gamma_1) = \sqrt{k}(\hat{\gamma}_1 - \gamma_1) + \sum_{j=2}^d \frac{\gamma_1}{\gamma_j} \frac{H_{1j}^{-1}}{H_{11}^{-1}} \sqrt{k}(\hat{\gamma}_{j+} - \hat{\gamma}_j) + o_p(1). \tag{18}$$

Now Corollary 3.1 and the continuous mapping theorem yield Eq. 14. □

Remark 5 In the bivariate case in Section 2, the determinant of the matrix H is always positive and hence the additional invertibility *assumption* on H is not needed there.

Acknowledgments We thank an associate editor and three referees for many insightful comments, which helped to improve this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J.: Statistics of extremes: theory and applications. Wiley, New York (2004)

Caeiro, F., Gomes, M.I.: Threshold selection in extreme value analysis. In: Dey, D., Yan, J. (eds.) Extreme Value Modeling and Risk Analysis: Methods and Applications, pp. 69–87. Chapman and Hall/CRC (2015)

Chakraborty, A., Cai, T.: Efficient and adaptive linear regression in semi-supervised settings. *Ann. Stat.* **46**, 1541–1572 (2018)

Csörgő, M., Csörgő, S., Horváth, L., Mason, D.M.: Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31–85 (1986)

de Haan, L., Ferreira, A.: Extreme value theory: an introduction. Springer, Berlin (2006)

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

- Einmahl, J.H.J.: The a.s. behavior of the weighted empirical process and the LIL for the weighted tail empirical process. *Ann. Probab.* **20**, 681–695 (1992)
- Gomes, M.I., Guillou, A.: Extreme value theory and statistics of univariate extremes: a review. *Int. Stat. Rev.* **83**, 263–292 (2015)
- Hill, B.: A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **3**, 1163–1174 (1975)
- Lay, T., Wallace, T.C.: *Modern global seismology*. Academic Press, Cambridge (1995)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.