

Bias-corrected and robust estimation of the bivariate stable tail dependence function

Mikael Escobar-Bach · Yuri Goegebeur ·
Armelle Guillou · Alexandre You

Received: date / Accepted: date

Abstract The stable tail dependence function gives a full characterisation of the extremal dependence between two or more random variables. In this paper, we propose an estimator for this function which is robust against outliers in the sample. The estimator is derived from a bivariate second order tail model together with a proper transformation of the bivariate observations, and its asymptotic properties are studied under some suitable regularity conditions. Our estimation procedure depends on two parameters: α , which controls the trade-off between efficiency and robustness of the estimator, and a second order parameter τ , which can be replaced by a fixed value or by an estimate. In case where τ has been replaced by the true value or by an external consistent estimator, our robust estimator is asymptotically unbiased, whereas in case where τ is mis-specified, one loses this property, but still our estimator performs quite well with respect to bias. The finite sample performance of our robust and bias-corrected estimator of the stable tail dependence function is examined on a simulation study involving uncontaminated and contaminated samples. In particular its behavior is illustrated for different values of the pair

This work was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416).

M. Escobar-Bach & Y. Goegebeur
Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
Tel.: +45-65504476
Fax: +45-65503345
E-mail: Yuri.Goegebeur@imada.sdu.dk

A. Guillou
Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS,
7 rue René Descartes, 67084 Strasbourg cedex, France
A. You
Société Générale Insurance - Sogessur, Direction Technique, 2, rue Jacques Daguerre, 92565
Rueil-Malmaison Cedex, France

(α, τ) and is compared with alternative estimators from the extreme value literature.

Keywords Multivariate extreme value statistics · stable tail dependence function · robustness · bias-correction

Mathematics Subject Classification (2000) MSC 62G05 · MSC 62G20 · MSC 62G32

1 Introduction

Modelling dependence among variables is a challenging topic in multivariate extreme value theory. For instance, in the case of environmental data, it can be of interest to link still water levels and wave heights in order to avoid flooding, whereas in finance measuring the dependence between risky asset returns can be crucial. Extremal dependence can be measured by coefficients that give a representative picture of the dependence, like e.g. the coefficient of tail dependence (Ledford and Tawn, 1996, 1997). Alternatively one could use functions that give a complete characterisation of the extreme dependence, e.g. the spectral distribution function (Einmahl et al., 1997) or the Pickands dependence function (Pickands, 1981). We also refer to Beirlant et al. (2004) and de Haan and Ferreira (2006) for recent accounts on the available approaches for describing tail dependence. In this paper, we focus on the bivariate stable tail dependence function, firstly introduced by Huang (1992) as follows. Let (X, Y) be a bivariate vector with continuous marginal distribution functions F_X and F_Y . The stable tail dependence function is defined for $(x, y) \in \mathbb{R}_+^2$ as

$$L(x, y) := \lim_{t \downarrow 0} \left\{ \frac{1}{t} \mathbb{P}(1 - F_X(X) \leq tx \text{ or } 1 - F_Y(Y) \leq ty) \right\}$$

which can be rewritten as

$$L(x, y) = \lim_{t \downarrow 0} \left\{ \frac{1}{t} [1 - F(F_X^{\leftarrow}(1 - tx), F_Y^{\leftarrow}(1 - ty))] \right\} \quad (1)$$

with F the bivariate distribution function of the vector (X, Y) and $F_\star^{\leftarrow}(t) := \inf\{z : F_\star(z) \geq t\}$ where \star denotes either X or Y . This convergence (1) is relevant in multivariate extreme value theory: indeed if F_X and F_Y are in the max-domains of attraction of some extreme value distributions G_X and G_Y , respectively, and if (1) is satisfied, then F is in the max-domain of attraction of an extreme value distribution G with marginals G_X and G_Y and with copula function determined by L . Estimating this function L is thus of interest and is the subject of this paper.

With this aim in mind, consider a sample of size n drawn from F and an intermediate sequence $m = m_n$, i.e. $m \rightarrow \infty$ as $n \rightarrow \infty$ with $m/n \rightarrow 0$. Let

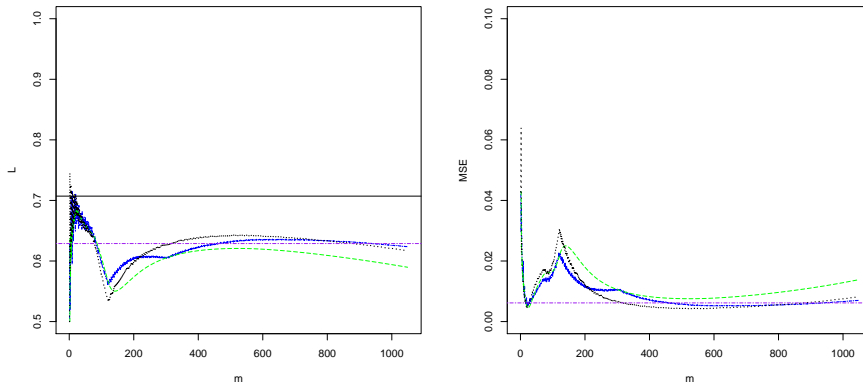


Fig. 1 Logistic distribution, $(x, y) = (0.5, 0.5)$. Comparison between the empirical estimator $\tilde{L}_m(x, y)$ (dotted line), a bias-corrected estimator $\tilde{L}_{m,0.4,990}(x, y)$ from Fougères et al. (2015) (dashed-dotted line) and its aggregated version \tilde{L}_{agg} (horizontal dashed-dotted line), and the bias-corrected estimator $\tilde{L}_{m,990}(x, y)$ proposed in Beirlant et al. (2016) (dashed line) in case of 5% of contamination (see Section 4 for their definitions). On the left: mean and on the right: MSE based on 500 samples of size 1000. The solid horizontal line in the left display is the true value of $L(x, y)$.

$X_{m,n}$, resp. $Y_{m,n}$, denote the m -th order statistic among n realisations of the margins X , resp. Y . The empirical estimator of L is then given by

$$\tilde{L}_m(x, y) = \frac{1}{m} \sum_{i=1}^n \mathbb{1}_{\{X_i \geq X_{n-[mx]+1, n} \text{ or } Y_i \geq Y_{n-[my]+1, n}\}}. \quad (2)$$

The weak convergence of this estimator, after proper normalisation, was established by Huang (1992). Other related references include Qi (1997), Drees and Huang (1998), Capéraà and Fougères (2000), Abdous and Ghoudi (2005), Schmidt and Stadtmüller (2006), Bücher et al. (2011). This estimator can suffer from bias as illustrated in Peng (2010), Fougères et al. (2015) or Beirlant et al. (2016) where bias-corrected estimators have been proposed. However, in practical data analysis, it can also happen that observations are contaminated in the sense that some outliers may have a disturbing effect on the estimators. Such a situation is illustrated in Figure 1 where we plot the mean and the mean squared error (MSE) of $\tilde{L}_m(x, y)$ and some recently introduced bias-corrected estimators for $L(x, y)$ as a function of m . The estimates are based on 500 samples of size 1000 from a Logistic distribution with 5% of contamination (as described in Section 4). As is clear from this figure, the classical estimators behave very poorly.

To solve this issue, it is important to propose robust estimators which also keep the nice property of being asymptotically unbiased as those proposed by Fougères et al. (2015) and Beirlant et al. (2016). This is the aim of the

present paper, where we propose to use the minimum density power divergence (MDPD) criterion introduced by Basu et al. (1998). This idea consists of defining the density power divergence between density functions f and h as follows

$$\Delta_\alpha(f, h) := \begin{cases} \int_{\mathbb{R}} \left[h^{1+\alpha}(y) - \left(1 + \frac{1}{\alpha}\right) h^\alpha(y) f(y) + \frac{1}{\alpha} f^{1+\alpha}(y) \right] dy, & \alpha > 0, \\ \int_{\mathbb{R}} \log \frac{f(y)}{h(y)} f(y) dy, & \alpha = 0. \end{cases}$$

Remark that for $\alpha = 0$ one recovers the Kullback-Leibler divergence, whereas setting $\alpha = 1$ leads to the L_2 divergence. Assume that the density function h depends on a parameter vector θ , and let Y_1, \dots, Y_n be a sample of independent and identically distributed random variables according to the density function f . The minimum density power divergence estimator (MDPDE) is the point $\hat{\theta}$ minimizing the empirical density power divergence

$$\hat{\Delta}_\alpha(\theta) := \begin{cases} \int_{\mathbb{R}} h^{1+\alpha}(y) dy - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n h^\alpha(Y_i), & \alpha > 0, \\ -\frac{1}{n} \sum_{i=1}^n \log h(Y_i), & \alpha = 0. \end{cases}$$

Note that in case $\alpha = 0$, the empirical density power divergence corresponds with minus the log-likelihood function. The parameter α controls the trade-off between efficiency and robustness of the MDPDE: it becomes more efficient but less robust against outliers as α gets closer to zero, whereas by increasing α the robustness increases and the efficiency decreases.

In the present paper we introduce a robust and bias-corrected estimator of the stable tail dependence function L . To the best of our knowledge, robust and bias-corrected estimation of L has not been considered in the extreme value literature. The remainder of the paper is organized as follows. In Section 2, we introduce our estimator of $L(x, y)$, while in Section 3 we establish its asymptotic properties. The finite sample performance of our procedure is examined in Section 4 on a simulation study. The proofs of our main results are postponed to the appendix, whereas the supplementary material contains the proof of a crucial but intermediate result (namely Theorem 1) together with some additional simulations.

2 Construction of the estimators

Let (X, Y) be a bivariate random vector with continuous marginal distribution functions F_X and F_Y satisfying

$$\mathbb{P}(1 - F_X(X) < x, 1 - F_Y(Y) < y) = x^{d_1} y^{d_2} g(x, y) \left(1 + \frac{1}{\eta} \delta(x, y)\right), \quad (3)$$

for $(x, y) \in [0, 1]^2 \setminus \{(0, 0)\}$, where d_1, d_2 are positive constants such that $d_1 + d_2 = \eta^{-1} \in (0, 1]$, g is a continuous homogeneous function of order 0 and δ a function of constant sign in the neighbourhood of zero, with $|\delta|$ being a bivariate regularly varying function, that is, there exists a function ξ such that

$$\lim_{t \downarrow 0} \frac{|\delta|(tx, ty)}{|\delta|(t, t)} = \xi(x, y), \quad (4)$$

for all $(x, y) \in [0, \infty)^2 \setminus \{(0, 0)\}$. Also, ξ is assumed to be continuous, homogeneous of order $\tau > 0$, and the convergence is uniform on $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$. Note that model (3) is essentially a condition on the copula function C of F . Indeed, one easily verifies that

$$\mathbb{P}(1 - F_X(X) < x, 1 - F_Y(Y) < y) = x + y - 1 + C(1 - x, 1 - y).$$

If $\eta \in (0, 1)$ then we are in the asymptotic independence case, and in that context $L(x, y) = x + y$ for $x, y > 0$. As such, estimating L is not relevant, and main interest is in estimating η , as has been done in Beirlant and Vandewalle (2002), Beirlant et al. (2011), Goegebeur and Guillou (2013), Dutang et al. (2014). On the contrary, if $\eta = 1$ then we are in the case of asymptotic dependence, and in that context estimating L is the subject of interest. Our starting point is thus model (3) with $d_1 + d_2 = 1$.

For convenience we suppose that the marginal distributions are unit Pareto. In that case, model (3) with $\eta = 1$ becomes

$$\mathbb{P}(X > x, Y > y) = x^{-d_1} y^{-d_2} g\left(\frac{1}{x}, \frac{1}{y}\right) \left(1 + \delta\left(\frac{1}{x}, \frac{1}{y}\right)\right), \quad x, y > 1. \quad (5)$$

Note that one can write

$$\mathbb{P}(X > x, Y > y) = \mathbb{P}\left(X > x, \frac{\omega}{1 - \omega} Y > x\right),$$

where $\omega := x/(x + y) \in (0, 1)$ can be interpreted as being a radial parameter. Thus the transformed variable $Z_\omega := \min(X, \frac{\omega}{1 - \omega} Y)$ admits the following survival function

$$\begin{aligned} \bar{F}_{Z_\omega}(x) &= \left(\frac{1 - \omega}{\omega}\right)^{-d_2} g\left(\frac{1}{x}, \frac{\omega}{1 - \omega} \frac{1}{x}\right) x^{-1} \left(1 + \delta\left(\frac{1}{x}, \frac{\omega}{1 - \omega} \frac{1}{x}\right)\right) \\ &= \left(\frac{1 - \omega}{\omega}\right)^{-d_2} g\left(1, \frac{\omega}{1 - \omega}\right) x^{-1} \left(1 + \delta\left(\frac{1}{x}, \frac{\omega}{1 - \omega} \frac{1}{x}\right)\right) \\ &=: K_\omega x^{-1} (1 + \delta_\omega(x)) \end{aligned} \quad (6)$$

using the homogeneity of order 0 of g .

Now, if we come back to our initial problem which is estimating the bivariate stable tail dependence function, recall that

$$\begin{aligned} L(x, y) &= \lim_{t \downarrow 0} \left\{ \frac{1}{t} \mathbb{P}(1 - F_X(X) \leq tx) + \frac{1}{t} \mathbb{P}(1 - F_Y(Y) \leq ty) \right. \\ &\quad \left. - \frac{1}{t} \mathbb{P}(1 - F_X(X) \leq tx, 1 - F_Y(Y) \leq ty) \right\} \\ &= \lim_{t \downarrow 0} \left\{ x + y - \frac{1}{t} \mathbb{P}\left(Z_{1-\omega} \geq \frac{1}{tx}\right) \right\} \\ &= \lim_{t \downarrow 0} \left\{ x + y - xK_{1-\omega}\left(1 + \delta_{1-\omega}\left(\frac{1}{tx}\right)\right) \right\} \end{aligned}$$

by (6). This leads to the simple estimator

$$\widehat{L}_m(x, y) = x + y - x\widehat{K}_{1-\omega, m}, \quad (7)$$

which requires an estimate for $K_{1-\omega}$. This estimator $\widehat{K}_{1-\omega, m}$ can now be obtained as follows. Consider the tail distribution of $Z_{1-\omega}$ given in (6). This distribution function belongs to the class of distribution functions introduced in Beirlant et al. (2009) and called Condition (\mathcal{R}) in Dierckx et al. (2013). As shown in Beirlant et al. (2009), the distribution function of the relative excesses $Z_{1-\omega}/u$ given that $Z_{1-\omega} > u$ can, for u large, be approximated by an extended Pareto distribution (EPD) function given by

$$H(y; \delta_{1-\omega}(u), \tau) := \begin{cases} 1 - y^{-1} [1 + \delta_{1-\omega}(u)(1 - y^{-\tau})]^{-1}, & y > 1, \\ 0, & y \leq 1, \end{cases}$$

where $\delta_{1-\omega}(u) > \max\{-1, -1/\tau\}$. Moreover, using Proposition 2.3 in Beirlant et al. (2009) the approximation error is uniformly $o(\delta_{1-\omega}(u))$ for $u \rightarrow \infty$. Using this property, one can estimate $\delta_{1-\omega}(u)$ by fitting the density function h associated with H to the relative excesses $Z_{1-\omega}/u$ given that $Z_{1-\omega} > u$ using e.g. the MDPD criterion.

In practice, the sample of pairs at our disposal, say $(X_1, Y_1), \dots, (X_n, Y_n)$, have unknown margins. Thus the first step is to transform the margins into unit Pareto distributions. To this aim, we can use the empirical distribution functions of the X and Y observations. This gives

$$\widetilde{Z}_{1-\omega, i} := \min\left(\frac{n+1}{n+1-R_i^X}, \frac{1-\omega}{\omega} \frac{n+1}{n+1-R_i^Y}\right)$$

with R_i^X and R_i^Y denoting the rank of X_i and Y_i , $i = 1, \dots, n$, in the respective samples. Then, in a second step, the parameter $\delta_{1-\omega}(u)$ of the EPD is estimated by fitting the density function h to the relative excesses $E_j := \widetilde{Z}_{1-\omega, n-j+1, n} / \widetilde{Z}_{1-\omega, n-m, n}$, $j = 1, \dots, m$, where $1 \leq m \leq n-1$ using the minimum density power divergence criterion. Note that, as usual in extreme value statistics, we use for the threshold u the $(m+1)$ largest observation $\widetilde{Z}_{1-\omega, n-m, n}$ within the sample. Concerning the parameter τ of the EPD, we fix it at some

value, either the true one or a mis-specified one, or we estimate it externally in a consistent way. External estimation of second order parameters is quite common in the univariate framework, see e.g. Gomes and Martins (2002), Gomes et al. (2008) and Dierckx et al. (2013). Joint estimation of $\delta_{1-\omega}(u)$ and τ has been considered, but leads to theoretical difficulties. In particular, we would need to introduce $\sqrt{mq_1}(k/n) \rightarrow \infty$, which is not in accordance with the assumptions of our Theorem 1 where the quantile process of $\tilde{Z}_{1-\omega}$ is studied. To get around this one would need to introduce a third order model with strict unit Pareto margins for deriving the asymptotic properties. The fact that we cannot use the empirical transformation of the margins (as we do in our methodology) due to the inversion procedure in the proof of Theorem 1, implies that such a method with strict Pareto margins would be very restrictive and without practical interest. Moreover, from a practical perspective, one can expect that joint estimation of $\delta_{1-\omega}$ and τ will lead to an increase in the asymptotic variance, as was already observed in the univariate context. Therefore, this approach is not pursued in this paper. Finally, in a last step, an estimator of $K_{1-\omega}$ follows

$$\hat{K}_{1-\omega,m} = \frac{m}{n} \frac{\tilde{Z}_{1-\omega,n-m,n}}{1 + \hat{\delta}_{1-\omega}(\tilde{Z}_{1-\omega,n-m,n})} \quad (8)$$

from which the estimator of $L(x, y)$ can be deduced by applying (7). The estimator (8) can be intuitively motivated as follows. Consider the tail distribution of $Z_{1-\omega}/u$ given that $Z_{1-\omega} > u$:

$$\frac{\bar{F}_{Z_{1-\omega}}(ux)}{\bar{F}_{Z_{1-\omega}}(u)} = \frac{K_{1-\omega}(ux)^{-1}(1 + \delta_{1-\omega}(ux))}{\bar{F}_{Z_{1-\omega}}(u)}.$$

Take $u = Z_{1-\omega,n-m,n}$, replace the denominator of the above display by its empirical value, being m/n , and set $x = 1$, to obtain

$$1 \approx \frac{n}{m} K_{1-\omega} Z_{1-\omega,n-m,n}^{-1} (1 + \delta_{1-\omega}(Z_{1-\omega,n-m,n})),$$

and thus

$$K_{1-\omega} \approx \frac{m}{n} \frac{Z_{1-\omega,n-m,n}}{1 + \delta_{1-\omega}(Z_{1-\omega,n-m,n})}.$$

The estimator (8) is then obtained by replacing $Z_{1-\omega,n-m,n}$ by its empirical analogue $\tilde{Z}_{1-\omega,n-m,n}$, and $\delta_{1-\omega}(\tilde{Z}_{1-\omega,n-m,n})$ by its MDPDE, denoted $\hat{\delta}_{1-\omega}(\tilde{Z}_{1-\omega,n-m,n})$.

3 Asymptotic properties

The crucial point in the methodology consists of computing the estimate for $\delta_{1-\omega}(\tilde{Z}_{1-\omega,n-m,n})$ by using the MDPD criterion. In the sequel, although this

parameter depends on the threshold $\tilde{Z}_{1-\omega, n-m, n}$, we do not make this dependence explicit but prefer to use the notation $\delta_{1-\omega}$ for simplicity. As we already mentioned, the parameter τ of the EPD is not estimated jointly with $\delta_{1-\omega}$ by the MDPD method but it is fixed at some value or estimated externally. The MDPDE of $\delta_{1-\omega}$ satisfies thus the estimating equation

$$\int_1^\infty h^\alpha(y; \delta_{1-\omega}, \tau) \frac{\partial h(y; \delta_{1-\omega}, \tau)}{\partial \delta_{1-\omega}} dy - \frac{1}{m} \sum_{j=1}^m h^{\alpha-1}(E_j; \delta_{1-\omega}, \tau) \frac{\partial h(E_j; \delta_{1-\omega}, \tau)}{\partial \delta_{1-\omega}} = 0, \quad (9)$$

where the density function h of the EPD is given by

$$h(y; \delta_{1-\omega}, \tau) = y^{-2} (1 + \delta_{1-\omega}(1 - y^{-\tau}))^{-2} (1 + \delta_{1-\omega}(1 - (1 - \tau)y^{-\tau})), \quad y > 1,$$

with $\tau > 0$ and $\delta_{1-\omega} > \max\{-1, -1/\tau\}$. Note that the estimating equation (9) depends only on the data through statistics of the form $(1/m) \sum_{j=1}^m E_j^s$ for $s < 0$. Thus, considering the tail quantile process $Q_n(t) := \tilde{Z}_{1-\omega, n-[mt], n}$, $0 < t < n/m$, these statistics can be expressed as the functional $\int_0^1 (Q_n(t)/Q_n(1))^s dt$.

For the sequel it is instructive to keep in mind that (3) satisfies the following second order multivariate regular variation condition introduced in Draisma et al. (2004):

Condition (SO): Let (X, Y) be a random vector with joint distribution function F and continuous marginal distribution functions F_X and F_Y such that

$$\lim_{t \downarrow 0} q_1(t)^{-1} \left(\frac{\mathbb{P}(1 - F_X(X) < tx, 1 - F_Y(Y) < ty)}{\mathbb{P}(1 - F_X(X) < t, 1 - F_Y(Y) < t)} - c(x, y) \right) =: c_1(x, y) \quad (10)$$

exists for all $x \geq 0, y \geq 0$ with $x + y > 0$, a function q_1 tending to zero as $t \downarrow 0$, and c_1 a function neither constant nor a multiple of c . Moreover, we assume that the convergence is uniform on $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$.

In case $\eta = 1$, the specific functions of Condition (SO) are for model (3) given by

$$\begin{cases} c(x, y) = x^{d_1} y^{d_2} \frac{g(x, y)}{g(1, 1)}, \\ q_1(t) = \tau \delta(t, t), \\ c_1(x, y) = c(x, y) \frac{\xi(x, y) - 1}{\tau}, \end{cases} \quad (11)$$

as shown in Lemma 1 of Dutang et al. (2014).

As a preliminary result, we establish in our first theorem the behavior of the tail quantile process $Q_n(t)$, when correctly normalized. To this aim, set $c_{1-\omega} := c(1, (1 - \omega)/\omega)$, $\xi_{1-\omega} := \xi(1, (1 - \omega)/\omega)$, $q(t) := \mathbb{P}(1 - F_X(X) < t, 1 - F_Y(Y) < t)$, let $k/n = q^{\leftarrow}(m/n)$, and denote $\ell = \lim_{n \rightarrow \infty} m/k$. Remark that this limit exists, see (20) infra.

Theorem 1. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of the random vector (X, Y) which has a joint distribution satisfying (3) with $\eta = 1$ such that

the function c given in (11) has continuous first order partial derivatives. For $m, k \rightarrow \infty$, as $n \rightarrow \infty$ such that $\sqrt{mq_1}(k/n) \rightarrow \lambda \in \mathbb{R}$ we have that there exist suitable versions of Q_n such that for all $t_0, \varepsilon > 0$

$$\sup_{0 < t \leq t_0} t^{\frac{3}{2} + \varepsilon} \left| \sqrt{m} \left(\frac{k}{n} Q_n(t) - \frac{c_{1-\omega}}{t} \right) - \frac{c_{1-\omega}}{t^2} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \lambda \frac{c_{1-\omega}}{t} \frac{\left(\frac{t}{c_{1-\omega}} \right)^\tau \xi_{1-\omega} - 1}{\tau} \right| = o_{\mathbb{P}}(1),$$

where

$$\overline{W}(x) := \widetilde{W} \left(x, x \frac{1-\omega}{\omega} \right)$$

with

$$\widetilde{W}(x, y) := \frac{1}{\sqrt{\ell}} [W(x, 0) + W(0, y) - W(x, y)] - \sqrt{\ell} [c_x(x, y)W(x, 0) + c_y(x, y)W(0, y)],$$

where c_x and c_y denote the derivatives of c with respect to x and y , respectively, and W a Gaussian process with mean zero and covariance structure given by

$$\begin{aligned} \mathbb{E}(W(x_1, y_1)W(x_2, y_2)) &= x_1 \wedge x_2 + y_1 \wedge y_2 \\ &\quad - \ell [c(x_1, y_1) + c(x_2, y_2) - c(x_1 \vee x_2, y_1 \vee y_2)]. \end{aligned}$$

Note that this theorem is in spirit close to the one derived in Dutang et al. (2014, Theorem 1) but here we consider the case $\eta = 1$ and we have a process \widetilde{W} much more complicated. We refer to the supplementary material for the proof of this theorem. The covariance structure of \overline{W} can be obtained after tedious computations involving mainly the use of the covariance structure of W and the homogeneity properties of c_x and c_y . It is given by

$$\begin{aligned} &\mathbb{E} \left(\overline{W} \left(\frac{t_1}{c_{1-\omega}} \right) \overline{W} \left(\frac{t_2}{c_{1-\omega}} \right) \right) \\ &= t_1 + t_2 - t_1 \vee t_2 \\ &\quad + \ell \left\{ \frac{t_1 \wedge t_2}{c_{1-\omega}} \left[c_x^2 \left(1, \frac{1-\omega}{\omega} \right) + c_y^2 \left(1, \frac{1-\omega}{\omega} \right) \frac{1-\omega}{\omega} \right] \right. \\ &\quad \quad - c_x \left(1, \frac{1-\omega}{\omega} \right) \left[t_1 \wedge t_2 + \frac{c(t_1 \wedge t_2, t_1 \vee t_2 \frac{1-\omega}{\omega})}{c_{1-\omega}} \right] \\ &\quad \quad \left. - c_y \left(1, \frac{1-\omega}{\omega} \right) \left[t_1 \wedge t_2 + \frac{c(t_1 \vee t_2, t_1 \wedge t_2 \frac{1-\omega}{\omega})}{c_{1-\omega}} \right] \right\} \\ &\quad + \frac{\ell^2}{c_{1-\omega}} c_x \left(1, \frac{1-\omega}{\omega} \right) c_y \left(1, \frac{1-\omega}{\omega} \right) \\ &\quad \quad \times \left[c \left(t_1 \wedge t_2, t_1 \vee t_2 \frac{1-\omega}{\omega} \right) + c \left(t_1 \vee t_2, t_1 \wedge t_2 \frac{1-\omega}{\omega} \right) \right]. \end{aligned}$$

We are now able to derive the asymptotic behavior of our main statistic appearing in the estimating equation (9).

Theorem 2. *Under the assumptions of Theorem 1, for any $s < 0$, we have that*

$$\begin{aligned} \frac{1}{m} \sum_{j=1}^m \left(\frac{\tilde{Z}_{1-\omega, n-j+1, n}}{\tilde{Z}_{1-\omega, n-m, n}} \right)^s &= \frac{1}{1-s} + \frac{s}{\sqrt{m}} \int_0^1 t^{-s} \left(\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right) dt \\ &\quad - \frac{\lambda s}{(1-s)(1-s+\tau)} \frac{\xi_{1-\omega}}{c_{1-\omega}^\tau} \frac{1}{\sqrt{m}} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right). \end{aligned}$$

From now on we will denote the true value of $\delta_{1-\omega}$ by δ_0 . Note that $\delta_0 = \delta_0(\tilde{Z}_{1-\omega, n-m, n})$. We want to prove the convergence in distribution of $\hat{L}_m(x, y)$. This requires to establish the weak convergence of $\hat{K}_{1-\omega, m}$ which is a function of the MDPDE $\hat{\delta}_{1-\omega}$ and $\tilde{Z}_{1-\omega, n-m, n} = Q_n(1)$. Consequently, in the next theorem, we derive the convergence in distribution of the vector $\begin{pmatrix} Q_n(1) \\ \hat{\delta}_{1-\omega} \end{pmatrix}$ correctly normalized, in case where τ is replaced by the true value τ_0 , by a mis-specified value, say $\tilde{\tau}$, possibly different from τ_0 , or by a suitable external estimator $\hat{\tau}$.

Theorem 3. *Under the assumptions of Theorem 1 we have that*

$$\sqrt{m} \begin{pmatrix} \frac{k}{n} Q_n(1) - c_{1-\omega} \\ \hat{\delta}_{1-\omega} - \delta_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} A_Q \\ A_\tau \end{pmatrix}$$

where

$$A_Q := c_{1-\omega} \overline{W} \left(\frac{1}{c_{1-\omega}} \right) + \lambda c_{1-\omega} \frac{c_{1-\omega}^{-\tau_0} \xi_{1-\omega} - 1}{\tau_0}$$

and

- in case $\tau = \tau_0$

$$\begin{aligned} A_{\tau_0} := & \frac{(1+2\alpha)(1+2\alpha+\tau_0)(1+2\alpha+2\tau_0)}{\tau_0^2(1+\tau_0+2\alpha\tau_0+4\alpha^2)} \left[2\alpha \int_0^1 t^{2\alpha} \left(\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right) dt \right. \\ & \left. - (2\alpha+\tau_0)(1+\tau_0) \int_0^1 t^{2\alpha+\tau_0} \left(\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right) dt \right] \end{aligned}$$

- in case $\tau = \tilde{\tau}$, a mis-specified value

$$\begin{aligned} A_{\tilde{\tau}} := & \frac{(1+2\alpha)(1+2\alpha+\tilde{\tau})(1+2\alpha+2\tilde{\tau})}{\tilde{\tau}^2(1+\tilde{\tau}+2\alpha\tilde{\tau}+4\alpha^2)} \left\{ 2\alpha \int_0^1 t^{2\alpha} \left(\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right) dt \right. \\ & \left. - (2\alpha+\tilde{\tau})(1+\tilde{\tau}) \int_0^1 t^{2\alpha+\tilde{\tau}} \left(\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right) dt \right. \\ & \left. + \lambda \frac{\xi_{1-\omega}}{c_{1-\omega}^{\tau_0}} \left[\frac{-2\alpha}{(1+2\alpha)(1+2\alpha+\tau_0)} + \frac{(2\alpha+\tilde{\tau})(1+\tilde{\tau})}{(1+2\alpha+\tilde{\tau})(1+2\alpha+\tau_0+\tilde{\tau})} \right. \right. \\ & \left. \left. - \frac{\tilde{\tau}^2(1+\tilde{\tau}+2\alpha\tilde{\tau}+4\alpha^2)}{\tau_0(1+2\alpha)(1+2\alpha+\tilde{\tau})(1+2\alpha+2\tilde{\tau})} \right] \right\} \end{aligned}$$

- in case $\tau = \hat{\tau}$, an external estimator satisfying $\sqrt{\bar{k}} q_1 \left(\frac{\bar{k}}{n} \right) (\hat{\tau} - \tau) = O_{\mathbb{P}}(1)$ with \bar{k} an intermediate sequence such that $\sqrt{\frac{m}{\bar{k}}} / q_1(\bar{k}/n) \rightarrow 0$

$$A_{\hat{\tau}} = A_{\tau_0}.$$

As is clear from Theorem 3, our estimator $\widehat{\delta}_{1-\omega}$ is asymptotically unbiased in the sense that the expectation of A_{τ_0} is zero, whatever the value of λ . This is also the case for $A_{\widehat{\tau}}$ but not for $A_{\widetilde{\tau}}$ if $\widetilde{\tau} \neq \tau_0$.

As a corollary of Theorem 3, we can deduce the convergence in distribution of $\widehat{L}_m(x, y)$.

Corollary 4. *Under the assumptions of Theorem 3 we have that*

$$\sqrt{m} \left(\widehat{L}_m(x, y) - L(x, y) \right) \xrightarrow{d} -x\ell A_Q + xK_{1-\omega}A_\tau + xK_{1-\omega}\lambda \frac{c_{1-\omega}^{-\tau_0} \xi_{1-\omega} - 1}{\tau_0}.$$

Again we can easily check that in case $\tau = \tau_0$ or if τ is replaced by an external estimator satisfying $\sqrt{\bar{k}} q_1 \left(\frac{\bar{k}}{n} \right) (\widehat{\tau} - \tau) = O_{\mathbb{P}}(1)$ with \bar{k} an intermediate sequence such that $\sqrt{\frac{m}{\bar{k}}} / q_1(\bar{k}/n) \rightarrow 0$, our estimator of $L(x, y)$ is asymptotically unbiased. When τ is mis-specified, one possibly loses the asymptotic unbiasedness. However, as illustrated in Section 4, in that case the estimators still perform quite well with respect to bias and they outperform estimators which are not corrected for bias. Note that our result in Corollary 4 is a pointwise weak convergence result. This could be extended to finite dimensional convergence, though this does not yield interesting additional insights. Obtaining weak convergence results for the stochastic process $\sqrt{m}(\widehat{L}_m(x, y) - L(x, y))$ is in our context very complicated as we do not have an explicit estimator for δ_0 . This will though be the subject of future research.

4 Simulation study

In this section, we illustrate the finite sample properties of our robust and bias-corrected estimator $\widehat{L}_m(x, y)$ through a simulation study. Using the homogeneity property, we consider only the estimation of $L(t, 1-t)$ for $t \in (0, 1)$, corresponding to the Pickands dependence function, and the following distributions:

- the symmetric logistic model, for which $L(x, y) = (x^{1/s} + y^{1/s})^s$. We set $s = 1/2$;
- the bivariate Pareto of type II model, called BP $II(p)$, for which $L(x, y) = x + y - (x^{-p} + y^{-p})^{-1/p}$. We set $p = 3$ and 4 ;
- the Cauchy distribution, for which $L(x, y) = (x^2 + y^2)^{1/2}$.

These distributions have already been considered in Fougères et al. (2015). They have shown in particular that they satisfy their second order condition. Since the latter implies our model (3), these distributions also satisfy our model. Note that other types of distributions have been studied but since the results are similar to those included we omit them, in order to keep the length of the paper under control.

For each pair (X_i, Y_i) , $i = 1, \dots, n$, independently from one of these distributions, we transform the margins into unit Pareto using the empirical

distribution functions. Then, we minimize the empirical density power divergence $\hat{\Delta}_\alpha(\delta_{1-\omega}, \tau)$ after replacing τ by 1, which can be the true value or a mis-specified one depending on the distribution, or by the external estimator $\hat{\tau}_{\bar{k}}$ introduced in Beirlant et al. (2016):

$$\hat{\tau}_{\bar{k}}(x, y) := \left(\frac{1}{\log r} \log \left| \frac{\Delta_{\bar{k}, a}(rx, ry)}{\Delta_{\bar{k}, a}(x, y)} \right| - 1 \right) \vee 0 \quad (12)$$

where

$$\Delta_{\bar{k}, a}(x, y) := a^{-1} \check{L}_{\bar{k}}(ax, ay) - \check{L}_{\bar{k}}(x, y)$$

with

$$\check{L}_{\bar{k}}(x, y) := \frac{1}{\bar{k}} \sum_{j=1}^{\bar{k}} a_j^{-1} K(a_j) \tilde{L}_{\bar{k}}(a_j x, a_j y)$$

and $a_j := \frac{j}{\bar{k}+1}$ and K a kernel such that $\int_0^1 K(u) u^{-\frac{1}{2}} du < \infty$. According to Proposition 1 in Beirlant et al. (2016), this estimator $\hat{\tau}_{\bar{k}}(x, y)$ satisfies the condition $\sqrt{\bar{k}q_1}(\bar{k}/n)(\hat{\tau} - \tau) = O_{\mathbb{P}}(1)$ required by our Theorem 3. Moreover as advocated in Beirlant et al. (2016), we use for a and r the value 0.4 and K is taken as the power kernel $K(t) := 6t^5 \mathbb{1}_{\{t \in (0,1)\}}$. Concerning \bar{k} , an extensive simulation study involving different percentages of contamination, different types of distributions and sample sizes, indicates that $\bar{k} = n/2$ performs quite well in all the contexts. Thus, these values for a, r, K and \bar{k} are those used throughout the paper. However, to avoid instabilities in the computation of our robust estimator of L due to the fact that $\hat{\tau}_{\bar{k}}(x, y)$ can be too close to 0, we set $\hat{\tau}_{\bar{k}}(x, y)$ at 0.3 if $\hat{\tau}_{\bar{k}}(x, y) \in [0, 0.3]$ and at its definition (12) otherwise. Similar proposals where second order rate parameters are bounded away from zero have already been used, for instance in Beirlant et al. (1999) in the univariate framework or Beirlant et al. (2016) in the multivariate setting. Finally, since any stable tail dependence function satisfies $\max(x, y) \leq L(x, y) \leq x + y$, this constraint has been imposed on all the estimators.

Concerning the contamination, we use the following algorithm, applied to each of the above mentioned distributions:

- We simulate n pairs (X_i, Y_i) , $i = 1, \dots, n$, independently with unit Fréchet margins;
- We draw two lines both with origin the point $(X_{990, n}, Y_{990, n})$ and to the points $(X_{n, n}, Y_j)$ and $(X_k, Y_{n, n})$, respectively, where j and k correspond to the indices associated to the maximum of the other component (note that the point $(X_{990, n}, Y_{990, n})$ is still in the main cloud, see Figure 2 for an illustration). We put $\lfloor n_0/2 \rfloor := \lfloor (n\varepsilon)/2 \rfloor$ variables on each of these lines, according to a Fréchet distribution. Our contaminated sample has a size $n^* = n + n_0$;
- We transform the margins of the contaminated sample empirically (with the ranks) into unit Pareto;
- We apply our MDPD criterion on the relative excesses.

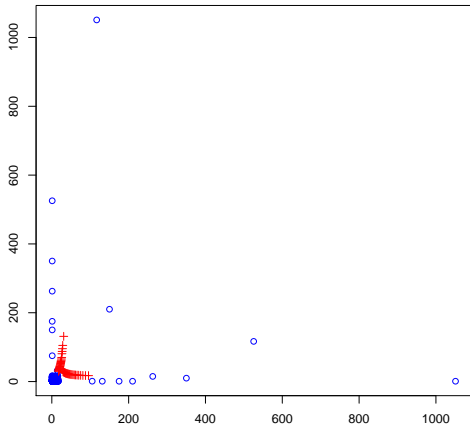


Fig. 2 Procedure of contamination

This procedure of contamination is illustrated in Figure 2, where the non-contaminated sample of 1000 pairs is represented as circles whereas the contaminated pairs are represented as crosses. Here, the percentage of contamination is 5%. This scatterplot is obtained after the empirical transformation of the margins into unit Pareto distributions. Thus this sample is the one on which our MDPD criterion will be applied.

The percentage of contamination is set to $\varepsilon = 0\%$ and 5% , while $n = 1000$ and the procedure is repeated 500 times. Two positions have been considered: $(x, y) = (0.5, 0.5)$ and $(x, y) = (0.4, 0.6)$. Note that other percentages of contamination have been used, and, as expected, increasing the percentage of contamination deteriorates all the estimators but it increases the superiority of our estimator compared to the other ones in particular in terms of MSE. Again, to keep the length of the paper reasonable, they are omitted. All the simulations have been done with Fortran and the NAG library. The code is available upon request from the authors.

First, we investigate the impact of the choice of the parameter α and of estimating externally the second order parameter τ by (12) rather than fixing it at the value one. We refer to the figures included in the supplementary material for the results of a simulation involving different values of α and where τ was either estimated or fixed at the value one. It is clear from the simulations that a value of α less than one should be used, since the difference between our estimators as a function of α is rather small for α less than 1 in both cases, τ estimated or fixed, whereas increasing its value leads to results which are again acceptable but worse than those obtained for $\alpha \leq 1$. This was also suggested in the original Basu et al. (1998) paper and more recently in papers devoted to this topic. A careful study of the figures provided in the supplementary material indicates however that, in case τ estimated or fixed, $\alpha = 0.25$ seems to be close to the best value, especially with respect to the MSE, although the differences between the values of α are rather small. Concerning the influence of estimating τ , except for the BPII(4) distribution, it is clear from all the other figures that using a fixed value for τ is the best approach, even if this fixed value is not the true one. Indeed, in general in that context, the bias is reduced and also the MSEs are smaller and more stable for a long range of values of m . The fact that this is not the case for the BPII(4) (which has $\tau_0 = 1/2$) is probably due to the proximity between the true value

τ_0 and the lower bound (0.3) imposed to avoid the instabilities. At first sight, the worse results obtained by estimating τ compared to mis-specifying it, can appear a bit surprising, but this has already been observed in the literature where also τ is often set at the value one (see e.g. Feuerverger and Hall, 1999) and is due, in our context, to the fact that our estimator $\hat{\tau}_k(x, y)$ defined in (12) is not robust (see also Figure 8 in Dierckx et al., 2013). Recall that the first paper dealing with the estimation of the second order parameter τ in the multivariate framework is due to Fougères et al. (2015), and even in the univariate context where this topic has been extensively studied, no robust estimator exists. In the sequel, we pursue the simulations with τ fixed at the value 1 and also we keep α at the value 0.25.

In order to evaluate the performance of our robust and bias-corrected estimator $\tilde{L}_m(x, y)$, we compare it with four other estimators already introduced in the literature:

- the empirical estimator $\tilde{L}_m(x, y)$;
- the bias-corrected estimator $\tilde{L}_{m,0.4,990}(x, y)$ proposed by Fougères et al. (2015):

$$\tilde{L}_{m,0.4,990}(x, y) := \frac{\tilde{L}_m(x, y)\tilde{\Delta}_{990,0.4}(0.4x, 0.4y) - \tilde{L}_m(0.4x, 0.4y)\tilde{\Delta}_{990,0.4}(x, y)}{\tilde{\Delta}_{990,0.4}(0.4x, 0.4y) - 0.4\tilde{\Delta}_{990,0.4}(x, y)}$$

where

$$\tilde{\Delta}_{990,0.4}(x, y) := 0.4^{-1}\tilde{L}_{990}(0.4x, 0.4y) - \tilde{L}_{990}(x, y);$$

- the aggregated estimator $\tilde{L}_{agg}(x, y)$ proposed by Fougères et al. (2015):

$$\tilde{L}_{agg}(x, y) := \text{Median}(\tilde{L}_{m,0.4,990}(x, y), m = 1, \dots, \kappa_n),$$

where κ_n is an appropriate fraction of n . Here we use $\kappa_n = \lfloor n(1 - \varepsilon) \rfloor$ which is the upper value for this parameter in case of contamination as outlined in Remark 11 in Fougères et al. (2015);

- the bias-corrected estimator proposed by Beirlant et al. (2016):

$$\bar{L}_{m,990}(x, y) := \frac{\tilde{L}_m(x, y) - \left(\frac{m}{990}\right)^{\hat{\tau}_{990}(x, y)} \check{\alpha}_{990}(x, y) \frac{1}{m} \sum_{j=1}^m K(a_j) a_j^{\hat{\tau}_{990}(x, y)}}{\frac{1}{m} \sum_{j=1}^m K(a_j)}$$

where

$$\check{\alpha}_{990}(x, y) := \frac{\sum_{j=1}^{990} \sum_{\ell=1}^{990} \left(a_j^{\hat{\tau}_{990}(x, y)} - a_\ell^{\hat{\tau}_{990}(x, y)} \right) a_j^{-1} \tilde{L}_{990}(a_j x, a_j y)}{\sum_{j=1}^{990} \sum_{\ell=1}^{990} a_j^{\hat{\tau}_{990}(x, y)} \left(a_j^{\hat{\tau}_{990}(x, y)} - a_\ell^{\hat{\tau}_{990}(x, y)} \right)}.$$

Note that in fact Fougères et al. (2015) have proposed two bias-corrected estimators for $L(x, y)$ but we show here only the best one and its aggregated version. The results are displayed in Figure 3 for 0% of contamination and in Figure 4 for 5% of contamination. Again $(x, y) = (0.5, 0.5)$ in these two figures. We can observe that in the presence of contamination, our estimator $\hat{L}_m(x, y)$

is always the best one in terms of bias and MSE, whatever the distribution. This is obviously expected since it is the only estimator which is robust, the other ones being very poor in that context. In case of no contamination, as illustrated in Figure 3, our robust and bias-corrected estimator $\widehat{L}_m(x, y)$ still performs quite well, always better than the empirical one $\widetilde{L}_m(x, y)$ and often similarly to $\overline{L}_{m,990}(x, y)$. Also note that in that context, $\widetilde{L}_{m,0.4,990}(x, y)$ has nice performance in terms of bias, but is worse than the others in terms of MSE. This is also the case of the aggregated version $\widetilde{L}_{agg}(x, y)$, the latter being almost perfect in case of the Cauchy distribution in terms of bias. Moreover, unlike the estimators proposed by Fougères et al. (2015), $\overline{L}_{m,990}(x, y)$ and our bias-corrected and robust estimator $\widehat{L}_m(x, y)$ have very smooth sample paths.

Finally, in Figure 5, we consider another position, $(0.4, 0.6)$, again for the Logistic and BPII(3) distributions with $\tau = 1$ and 0% and 5% of contamination. In that context, the estimation is slightly more difficult than in case $x = y$ with a reduced stable part of the mean and the MSE, but again the influence of α is rather small, the value of $\alpha = 0.25$ being again the best one.

5 Conclusions and future research

In this paper we introduced a robust and bias-corrected estimator for the stable tail dependence function in extreme value theory, and studied its weak convergence under some suitable regularity conditions. We evaluated our method by means of some simulations, involving both uncontaminated and contaminated samples, and learned that the estimator is very competitive with recent alternatives from the extreme value literature.

One possibility for future research could be the study of estimators for second order parameters in multivariate extreme value statistics. Unlike the univariate context, where several well working estimators are available, this area is highly unexplored in the multivariate context. Indeed, we are only aware of the recent proposals by Fougères et al. (2015) and Beirlant et al. (2016), though these estimators are not robust against outliers. Other possible topics are the extension of the methodology to the case where covariates are available and goodness-of-fit tests.

6 Appendix: Proofs

Proof of Theorem 2. Recall that $1/m \sum_{j=1}^m (\widetilde{Z}_{1-\omega, n-j+1, n} / \widetilde{Z}_{1-\omega, n-m, n})^s$ can be rewritten as $\int_0^1 (Q_n(t) / Q_n(1))^s dt$. Thus the main argument to prove Theorem 2 is to use the representation

$$Q_n(t) = \frac{n}{k} \frac{c_{1-\omega}}{t} \left[1 + \frac{1}{t\sqrt{m}} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) + \frac{\lambda}{\sqrt{m}} \frac{\left(\frac{t}{c_{1-\omega}} \right)^\tau \xi_{1-\omega} - 1}{\tau} + o_{\mathbb{P}} \left(\frac{t^{-(\frac{1}{2}+\varepsilon)}}{\sqrt{m}} \right) \right] \quad (13)$$

which follows from Theorem 1. To this aim, we use a Taylor expansion of the integrand and thus we need that all the terms appearing in the square parenthesis in (13) tend to 0 uniformly. To reach this goal we have to split the integrals into two parts, from 0 to $m^{-\kappa}$ and from $m^{-\kappa}$ to 1, for a suitable $\kappa \in (0, 1/(1+2\varepsilon))$.

Assuming $\kappa > 1/2$, we clearly have $\int_0^{m^{-\kappa}} \left(\frac{Q_n(t)}{Q_n(1)}\right)^s dt = o_{\mathbb{P}}(1/\sqrt{m})$. Also, for $\varepsilon < 0.5$

$$\begin{aligned} \int_{m^{-\kappa}}^1 \left(\frac{Q_n(t)}{Q_n(1)}\right)^s dt &= \int_{m^{-\kappa}}^1 t^{-s} dt + \frac{s}{\sqrt{m}} \int_{m^{-\kappa}}^1 t^{-s} \left[\frac{1}{t} \overline{W}\left(\frac{t}{c_{1-\omega}}\right) - \overline{W}\left(\frac{1}{c_{1-\omega}}\right) \right] dt \\ &\quad + \frac{\lambda s}{\sqrt{m}} \frac{\xi_{1-\omega}}{c_{1-\omega}^\tau} \int_{m^{-\kappa}}^1 t^{-s} \frac{t^\tau - 1}{\tau} dt + \frac{o_{\mathbb{P}}(1)}{\sqrt{m}} \int_{m^{-\kappa}}^1 t^{-s-\frac{1}{2}-\varepsilon} dt + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right) \\ &= \int_0^1 t^{-s} dt + \frac{s}{\sqrt{m}} \int_0^1 t^{-s} \left[\frac{1}{t} \overline{W}\left(\frac{t}{c_{1-\omega}}\right) - \overline{W}\left(\frac{1}{c_{1-\omega}}\right) \right] dt \\ &\quad + \frac{\lambda s}{\sqrt{m}} \frac{\xi_{1-\omega}}{c_{1-\omega}^\tau} \int_0^1 t^{-s} \frac{t^\tau - 1}{\tau} dt + \frac{o_{\mathbb{P}}(1)}{\sqrt{m}} \int_0^1 t^{-s-\frac{1}{2}-\varepsilon} dt + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right) \end{aligned}$$

from which Theorem 2 follows.

Proof of Theorem 3. The expression of A_Q follows from Theorem 1. Below we give the details about the proof which leads to A_{τ_0} . Concerning $A_{\widehat{\tau}}$ and $A_{\widehat{\tau}}$, the proofs are similar. Thus we will limit ourselves to the main differences.

Recall that $\left. \frac{\partial \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\widehat{\delta}_{1-\omega}} = 0$ by definition. Thus a Taylor series expansion around δ_0 combined with the boundedness of the third derivative of $\widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)$ with respect to $\delta_{1-\omega}$ leads to

$$0 = \left. \frac{\partial \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\delta_0} + \left. \frac{\partial^2 \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}^2} \right|_{\delta_{1-\omega}=\delta_0} (\widehat{\delta}_{1-\omega} - \delta_0)(1 + o_{\mathbb{P}}(1))$$

from which we deduce that

$$\sqrt{m}(\widehat{\delta}_{1-\omega} - \delta_0) = - \left(\left. \frac{\partial^2 \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}^2} \right|_{\delta_{1-\omega}=\delta_0} \right)^{-1} \sqrt{m} \left. \frac{\partial \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\delta_0} (1 + o_{\mathbb{P}}(1)). \quad (14)$$

Since we only need the dominant term of the second derivative, it is sufficient to compute it at $\delta_{1-\omega} = 0$. Doing so, we obtain

$$\begin{aligned}
\left. \frac{\partial^2 \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}^2} \right|_{\delta_{1-\omega}=0} &= 2(1+\alpha)(3+2\alpha) \int_1^\infty y^{-2(1+\alpha)}(1-y^{-\tau_0})^2 dy \\
&\quad - 4(1+\alpha)^2 \int_1^\infty y^{-2(1+\alpha)}(1-y^{-\tau_0})(1-(1-\tau_0)y^{-\tau_0}) dy \\
&\quad + \alpha(1+\alpha) \int_1^\infty y^{-2(1+\alpha)}(1-(1-\tau_0)y^{-\tau_0})^2 dy \\
&\quad - 2(\alpha+1)(2\alpha+1) \frac{1}{m} \sum_{j=1}^m E_j^{-2\alpha} (1-E_j^{-\tau_0})^2 \\
&\quad + 4\alpha(1+\alpha) \frac{1}{m} \sum_{i=1}^m E_j^{-2\alpha} (1-E_j^{-\tau_0})(1-(1-\tau_0)E_j^{-\tau_0}) \\
&\quad - (\alpha-1)(1+\alpha) \frac{1}{m} \sum_{j=1}^m E_j^{-2\alpha} (1-(1-\tau_0)E_j^{-\tau_0})^2 \\
&= \frac{\tau_0^2(1+\alpha)(1+\tau_0+2\alpha\tau_0+4\alpha^2)}{(1+2\alpha)(1+2\alpha+\tau_0)(1+2\alpha+2\tau_0)} (1+o_{\mathbb{P}}(1)) \quad (15)
\end{aligned}$$

by applying Theorem 2. Now, concerning the first derivative

$$\begin{aligned}
&\left. \frac{\partial \widehat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\delta_0} \\
&= -2(1+\alpha) \int_1^\infty y^{-2(1+\alpha)}(1-y^{-\tau_0})(1+\delta_0(1-y^{-\tau_0}))^{-3-2\alpha}(1+\delta_0(1-(1-\tau_0)y^{-\tau_0}))^{1+\alpha} dy \\
&\quad + (1+\alpha) \int_1^\infty y^{-2(1+\alpha)}(1-(1-\tau_0)y^{-\tau_0})(1+\delta_0(1-y^{-\tau_0}))^{-2(1+\alpha)}(1+\delta_0(1-(1-\tau_0)y^{-\tau_0}))^\alpha dy \\
&\quad + 2(1+\alpha) \frac{1}{m} \sum_{j=1}^m E_j^{-2\alpha} (1-E_j^{-\tau_0})(1+\delta_0(1-E_j^{-\tau_0}))^{-1-2\alpha}(1+\delta_0(1-(1-\tau_0)E_j^{-\tau_0}))^\alpha \\
&\quad - (1+\alpha) \frac{1}{m} \sum_{j=1}^m E_j^{-2\alpha} (1-(1-\tau_0)E_j^{-\tau_0})(1+\delta_0(1-E_j^{-\tau_0}))^{-2\alpha}(1+\delta_0(1-(1-\tau_0)E_j^{-\tau_0}))^{\alpha-1} \\
&= -\frac{2\alpha(1+\alpha)}{\sqrt{m}} \int_0^1 t^{2\alpha} \left[\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right] dt \\
&\quad + \frac{(1+\alpha)(2\alpha+\tau_0)(1+\tau_0)}{\sqrt{m}} \int_0^1 t^{2\alpha+\tau_0} \left[\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right] dt \\
&\quad - q_1 \left(\frac{k}{n} \right) \frac{\xi_{1-\omega} \tau_0 (1+\alpha)(1+\tau_0+2\alpha\tau_0+4\alpha^2)}{c_{1-\omega}^{\tau_0} (1+2\alpha)(1+2\alpha+\tau_0)(1+2\alpha+2\tau_0)} \\
&\quad + \delta_0 \frac{\tau_0^2(1+\alpha)(1+\tau_0+2\alpha\tau_0+4\alpha^2)}{(1+2\alpha)(1+2\alpha+\tau_0)(1+2\alpha+2\tau_0)} + O_{\mathbb{P}}(\delta_0^2) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right)
\end{aligned}$$

by again an application of Theorem 2 and using expansions of the integrands. Now recall that

$$\begin{aligned}
\delta_0 &= \delta_0(\tilde{Z}_{1-\omega, n-m, n}) = \delta\left(\frac{1}{\tilde{Z}_{1-\omega, n-m, n}}, \frac{1-\omega}{\omega} \frac{1}{\tilde{Z}_{1-\omega, n-m, n}}\right) \text{ by (6)} \\
&= \xi\left(1, \frac{1-\omega}{\omega}\right) \delta\left(\frac{1}{\tilde{Z}_{1-\omega, n-m, n}}, \frac{1}{\tilde{Z}_{1-\omega, n-m, n}}\right) (1 + o_{\mathbb{P}}(1)) \text{ by (4)} \\
&= \frac{\xi_{1-\omega}}{\tau_0} q_1\left(\frac{1}{\tilde{Z}_{1-\omega, n-m, n}}\right) (1 + o_{\mathbb{P}}(1)) \text{ by (11)} \\
&= \frac{\xi_{1-\omega}}{\tau_0} \frac{1}{c_{1-\omega}^{\tau_0}} q_1\left(\frac{k}{n}\right) (1 + o_{\mathbb{P}}(1)),
\end{aligned}$$

using the regularly varying property of q_1 and Theorem 1. Consequently

$$\begin{aligned}
\left. \frac{\partial \hat{\Delta}_\alpha(\delta_{1-\omega}, \tau_0)}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\delta_0} &= -\frac{2\alpha(1+\alpha)}{\sqrt{m}} \int_0^1 t^{2\alpha} \left[\frac{1}{t} \overline{W}\left(\frac{t}{c_{1-\omega}}\right) - \overline{W}\left(\frac{1}{c_{1-\omega}}\right) \right] dt \\
&\quad + \frac{(1+\alpha)(2\alpha+\tau_0)(1+\tau_0)}{\sqrt{m}} \int_0^1 t^{2\alpha+\tau_0} \left[\frac{1}{t} \overline{W}\left(\frac{t}{c_{1-\omega}}\right) - \overline{W}\left(\frac{1}{c_{1-\omega}}\right) \right] dt \\
&\quad + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right). \tag{16}
\end{aligned}$$

Combining (13) and (14) with (15) and (16) yields the expression of A_{τ_0} .

Now, concerning $A_{\tilde{\tau}}$, the limit in probability of the second derivative is the same as in (15) with τ_0 replaced by $\tilde{\tau}$. For the first derivative, similar arguments as for (16) lead to

$$\begin{aligned}
&\left. \frac{\partial \hat{\Delta}_\alpha(\delta_{1-\omega}, \tilde{\tau})}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\delta_0} \\
&= -\frac{2\alpha(1+\alpha)}{\sqrt{m}} \int_0^1 t^{2\alpha} \left[\frac{1}{t} \overline{W}\left(\frac{t}{c_{1-\omega}}\right) - \overline{W}\left(\frac{1}{c_{1-\omega}}\right) \right] dt \\
&\quad + \frac{(1+\alpha)(2\alpha+\tilde{\tau})(1+\tilde{\tau})}{\sqrt{m}} \int_0^1 t^{2\alpha+\tilde{\tau}} \left[\frac{1}{t} \overline{W}\left(\frac{t}{c_{1-\omega}}\right) - \overline{W}\left(\frac{1}{c_{1-\omega}}\right) \right] dt \\
&\quad + q_1\left(\frac{k}{n}\right) \frac{\xi_{1-\omega}}{c_{1-\omega}^{\tau_0}} \left[\frac{2\alpha(1+\alpha)}{(1+2\alpha)(1+2\alpha+\tau_0)} - \frac{(1+\alpha)(2\alpha+\tilde{\tau})(1+\tilde{\tau})}{(1+2\alpha+\tilde{\tau})(1+2\alpha+\tau_0+\tilde{\tau})} \right. \\
&\quad \quad \left. + \frac{\tilde{\tau}^2(1+\alpha)(1+\tilde{\tau}+2\alpha\tilde{\tau}+4\alpha^2)}{\tau_0(1+2\alpha)(1+2\alpha+\tilde{\tau})(1+2\alpha+2\tilde{\tau})} \right] \\
&\quad + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)
\end{aligned}$$

which leads to the expression of $A_{\tilde{\tau}}$.

Finally, concerning $A_{\hat{\tau}}$, the limit in probability of the second derivative is again the same as in (15) since $\hat{\tau}$ is a consistent estimator of τ . For the

first derivative, similar arguments as for (16), combined with the mean value theorem, lead to

$$\begin{aligned} \left. \frac{\partial \widehat{\Delta}_\alpha(\delta_{1-\omega}, \widehat{\tau})}{\partial \delta_{1-\omega}} \right|_{\delta_{1-\omega}=\delta_0} &= -\frac{2\alpha(1+\alpha)}{\sqrt{m}} \int_0^1 t^{2\alpha} \left[\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right] dt \\ &+ \frac{(1+\alpha)(2\alpha+\tau_0)(1+\tau_0)}{\sqrt{m}} \int_0^1 t^{2\alpha+\tau_0} \left[\frac{1}{t} \overline{W} \left(\frac{t}{c_{1-\omega}} \right) - \overline{W} \left(\frac{1}{c_{1-\omega}} \right) \right] dt \\ &+ o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right) + O_{\mathbb{P}} \left(\frac{1}{\sqrt{k} q_1 \left(\frac{\bar{k}}{n} \right)} \right). \end{aligned}$$

Using our assumption on the intermediate sequence \bar{k} , the expression of $A_{\widehat{\tau}}$ follows.

Proof of Corollary 4. Note that

$$\sqrt{m} \left(\widehat{L}_m(x, y) - L(x, y) \right) = -x\sqrt{m} \left(\widehat{K}_{1-\omega, m} - K_{1-\omega} \right). \quad (17)$$

Consequently, we have to look at

$$\begin{aligned} \sqrt{m} \left(\widehat{K}_{1-\omega, m} - K_{1-\omega} \right) &= \sqrt{m} \left(\frac{m}{k} \frac{\frac{k}{n} Q_n(1)}{1 + \widehat{\delta}_{1-\omega}} - K_{1-\omega} \right) \\ &= \sqrt{m} \frac{1 + \delta_0}{1 + \widehat{\delta}_{1-\omega}} \frac{m}{k} \left(\frac{\frac{k}{n} Q_n(1) - c_{1-\omega}}{1 + \delta_0} - \frac{k}{m} \left[1 + \frac{\widehat{\delta}_{1-\omega} - \delta_0}{1 + \delta_0} \right] K_{1-\omega} + \frac{c_{1-\omega}}{1 + \delta_0} \right) \\ &= \sqrt{m} \left(\frac{k}{n} Q_n(1) - c_{1-\omega} \right) \ell(1 + o_{\mathbb{P}}(1)) - \sqrt{m} (\widehat{\delta}_{1-\omega} - \delta_0) K_{1-\omega} (1 + o_{\mathbb{P}}(1)) \\ &\quad - \sqrt{m} \left(K_{1-\omega} - \frac{m}{k} \frac{c_{1-\omega}}{1 + \delta_0} \right) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (18)$$

We have the joint distribution of the two first terms, so we only have to look at the last one. Recall that, by definition (see (6))

$$K_{1-\omega} = \left(\frac{1-\omega}{\omega} \right)^{d_2} g \left(1, \frac{1-\omega}{\omega} \right).$$

Now using (11), we have

$$c \left(\frac{1}{x}, \frac{1-\omega}{\omega} \frac{1}{x} \right) = x^{-1} \left(\frac{1-\omega}{\omega} \right)^{d_2} \frac{g \left(\frac{1}{x}, \frac{1-\omega}{\omega} \frac{1}{x} \right)}{g(1, 1)} = x^{-1} \left(\frac{1-\omega}{\omega} \right)^{d_2} \frac{g \left(1, \frac{1-\omega}{\omega} \right)}{g(1, 1)}$$

by homogeneity of order 0 of g . This implies that

$$K_{1-\omega} = g(1, 1) x c \left(\frac{1}{x}, \frac{1-\omega}{\omega} \frac{1}{x} \right) = g(1, 1) c_{1-\omega}, \quad (19)$$

by homogeneity of order 1 of c . Now recall that $\frac{k}{n} = q^{\leftarrow} \left(\frac{m}{n} \right)$, i.e.

$$\begin{aligned} \frac{m}{k} &= \frac{n}{k} \mathbb{P} \left(1 - F_X(X) < \frac{k}{n}, 1 - F_Y(Y) < \frac{k}{n} \right) \\ &= g(1, 1) \left(1 + \delta \left(\frac{k}{n}, \frac{k}{n} \right) \right) \text{ by (3) with } \eta = 1 \text{ and homogeneity of } g. \end{aligned} \quad (20)$$

Combining (19) and (20), the quantity of interest can be rewritten as

$$\begin{aligned} \sqrt{m} \left(K_{1-\omega} - \frac{m}{k} \frac{c_{1-\omega}}{1 + \delta_0} \right) &= g(1, 1) c_{1-\omega} \sqrt{m} \left(1 - \frac{1 + \delta \left(\frac{k}{n}, \frac{k}{n} \right)}{1 + \delta_0} \right) \\ &= g(1, 1) c_{1-\omega} \sqrt{m} \left(\delta_0 - \delta \left(\frac{k}{n}, \frac{k}{n} \right) \right) + o_{\mathbb{P}}(1) \\ &= K_{1-\omega} \sqrt{m} q_1 \left(\frac{k}{n} \right) \frac{c_{1-\omega}^{-\tau_0} \xi_{1-\omega} - 1}{\tau_0} + o_{\mathbb{P}}(1) \end{aligned} \quad (21)$$

by (11) and using the facts that $\delta_0 = c_{1-\omega}^{-\tau_0} \xi_{1-\omega} \frac{1}{\tau_0} q_1 \left(\frac{k}{n} \right) (1 + o_{\mathbb{P}}(1))$ and $\sqrt{m} q_1(k/n) \rightarrow \lambda \in \mathbb{R}$.

Finally combining (17), (18) and (21) with our Theorem 3, Corollary 4 follows.

Acknowledgements The authors thank the associate editor and the reviewers for their helpful comments which led to improvements of their paper.

References

1. Abdous B, Ghoudi K (2005) Non-parametric estimators of multivariate extreme dependence functions. *J Nonparametr Stat* 17:915–935.
2. Basu A, Harris IR, Hjort NL, Jones MC (1998) Robust and efficient estimation by minimizing a density power divergence. *Biometrika* 85:549–559.
3. Beirlant J, Dierckx G, Goegebeur Y, Matthys G (1999) Tail index estimation and an exponential regression model. *Extremes* 2:177–200.
4. Beirlant J, Dierckx G, Guillou A (2011) Bias-reduced estimators for bivariate tail modelling. *Insur Math Econ* 49:18–26.
5. Beirlant J, Escobar-Bach M, Goegebeur Y, Guillou A (2016) Bias-corrected estimation of stable tail dependence function. *J Multivariate Anal* 143:453–466.
6. Beirlant J, Goegebeur Y, Segers J, Teugels J (2004) *Statistics of Extremes – Theory and Applications*. Wiley.
7. Beirlant J, Joossens E, Segers J (2009) Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *J Stat Plan Infer* 139:2800–2815.
8. Beirlant J, Vandewalle B (2002) Some comments on the estimation of a dependence index in bivariate extreme value statistics. *Stat Probabil Lett* 60:265–278.
9. Bücher A, Dette H, Volgushev S (2011) New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann Stat* 39:1963–2006.
10. Capérea P, Fougères A-L (2000) Estimation of a bivariate extreme value distribution. *Extremes* 3:311–329.
11. Dierckx G, Goegebeur Y, Guillou A. (2013) An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *J Multivariate Anal* 121:70–86.
12. Draisma G, Drees H, Ferreira A, de Haan L (2004) Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* 10:251–280.

13. Drees H, Huang X (1998) Best attainable rates of convergence for estimators of the stable tail dependence function. *J Multivariate Anal* 64:25–47.
14. Dutang C, Goegebeur Y, Guillou A (2014) Robust and bias-corrected estimation of the coefficient of tail dependence. *Insur Math Econ* 57:46–57.
15. Einmahl JHJ, de Haan L, Sinha AK (1997) Estimating the spectral measure of an extreme value distribution. *Stoch Proc Appl* 70:143–171.
16. Feuerverger A, Hall P (1999) Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann Stat* 27:760–781.
17. Fougères AL, de Haan L, Mercadier C (2015) Bias correction in multivariate extremes. *Ann Stat* 43:903–934.
18. Goegebeur Y, Guillou A (2013) Asymptotically unbiased estimation of the coefficient of tail dependence. *Scand J Stat* 40:174–189.
19. Gomes MI, de Haan L, Rodrigues LH (2008) Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J Roy Stat Soc B* 70:31–52.
20. Gomes MI, Martins MG (2002) "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* 5:5–31.
21. de Haan L, Ferreira A (2006) *Extreme Value Theory - An Introduction*. Springer, New-York.
22. Huang X (1992) *Statistics of bivariate extremes*. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.
23. Ledford AW, Tawn JA (1996) Statistics for near independence in multivariate extreme values. *Biometrika* 83:169–187.
24. Ledford AW, Tawn JA (1997) Modelling dependence within joint tail regions. *J Roy Stat Soc B* 59:475–499.
25. Peng L (2010) A practical way for estimating tail dependence functions. *Stat Sinica* 20:365–378.
26. Pickands J (1981) Multivariate extreme value distributions. *B Int Statist Inst* 49:859–878.
27. Qi Y (1997) Almost sure convergence of the stable tail empirical dependence function in multivariate extreme statistics. *Acta Math Appl Sin-E* 13:167–175.
28. Schmidt R, Stadtmüller U (2006) Non-parametric estimation of tail dependence. *Scand J Stat* 33:307–335.

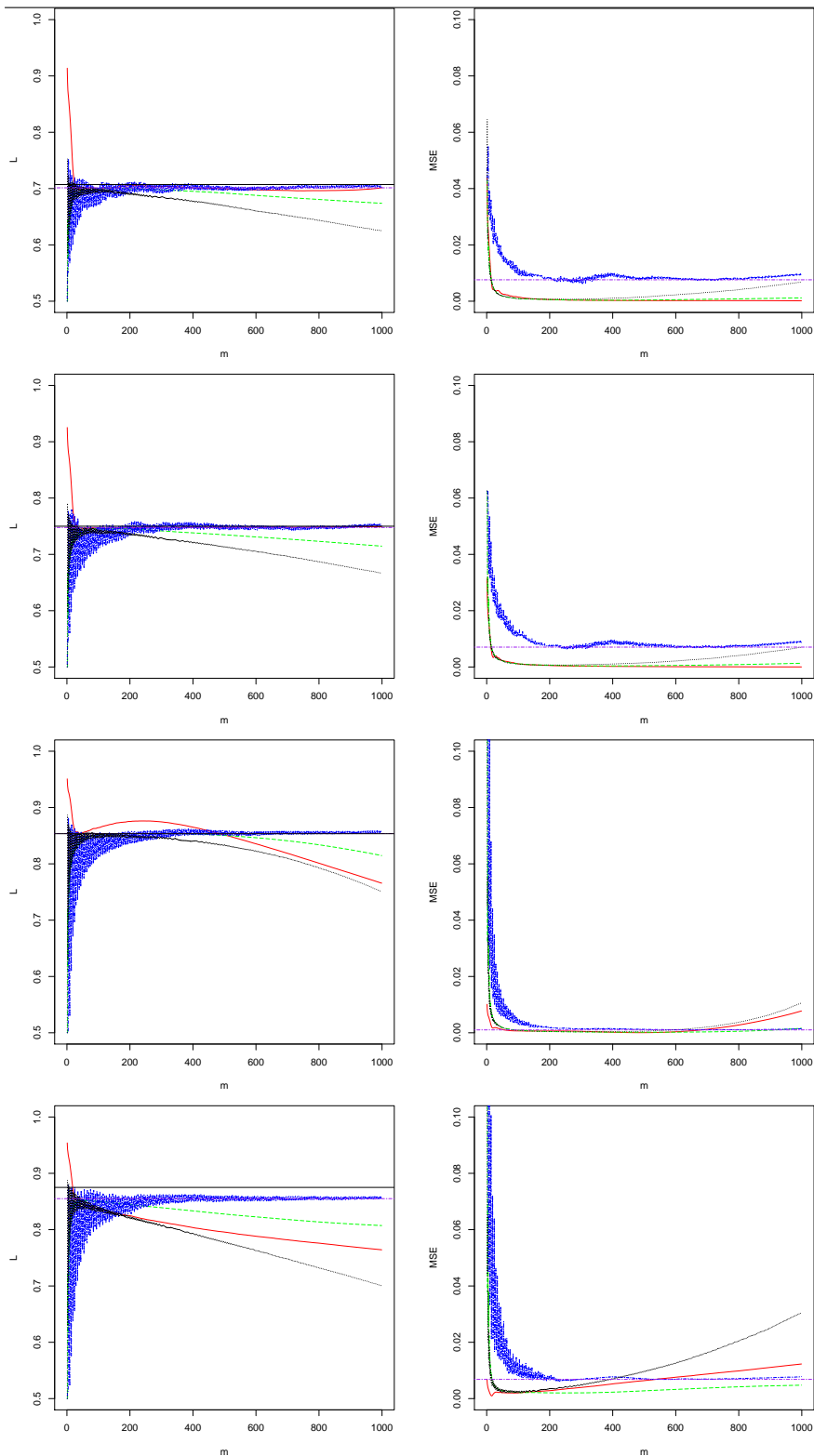


Fig. 3 Comparison between our robust and bias-corrected estimator $\hat{L}_m(0.5, 0.5)$ with $(\alpha, \tau) = (0.25, 1)$ (full line) and the empirical estimator $\tilde{L}_m(0.5, 0.5)$ (dotted line), the bias-corrected estimators $\tilde{L}_{m,0.4,990}(0.5, 0.5)$ (dashed-dotted line), $\tilde{L}_{agg}(x, y)$ (horizontal dashed-dotted line) and $\bar{L}_{m,990}(0.5, 0.5)$ (dashed line) in case of 0% of contamination. First row: Logistic; Second row: BPII(3); Third row: Cauchy; Fourth row: BPII(4) distributions. On the left: mean and on the right: MSE based on 500 replications. The full horizontal line on the left is the true value of the parameter.

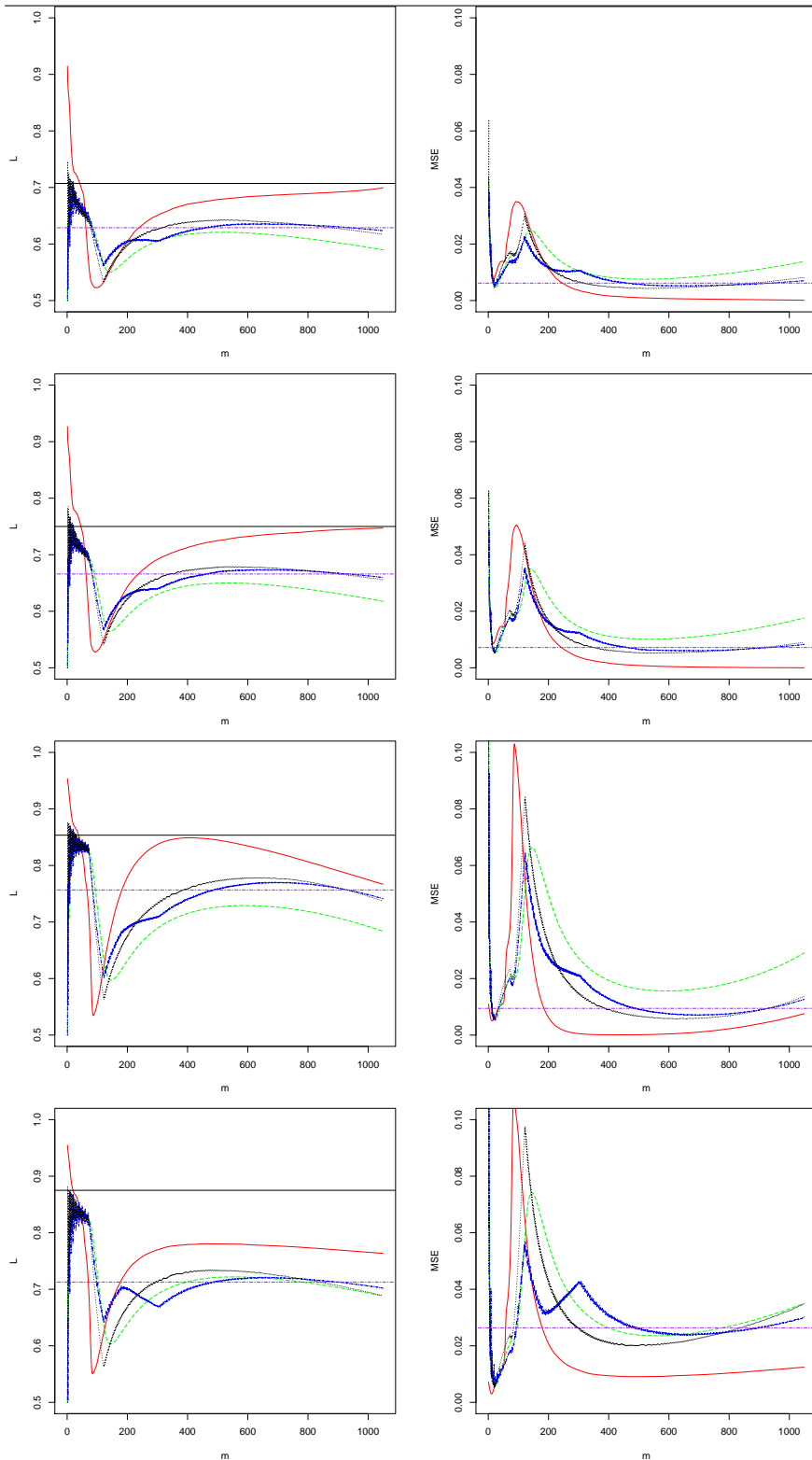


Fig. 4 Comparison between our robust and bias-corrected estimator $\hat{L}_m(0.5, 0.5)$ with $(\alpha, \tau) = (0.25, 1)$ (full line) and the empirical estimator $\tilde{L}_m(0.5, 0.5)$ (dotted line), the bias-corrected estimators $\tilde{L}_{m,0.4,990}(0.5, 0.5)$ (dashed-dotted line), $\tilde{L}_{agg}(x, y)$ (horizontal dashed-dotted line) and $\bar{L}_{m,990}(0.5, 0.5)$ (dashed line) in case of 5% of contamination. First row: Logistic; Second row: BPII(3); Third row: Cauchy; Fourth row: BPII(4) distributions. On the left: mean and on the right: MSE based on 500 replications. The full horizontal line on the left is the true value of the parameter.

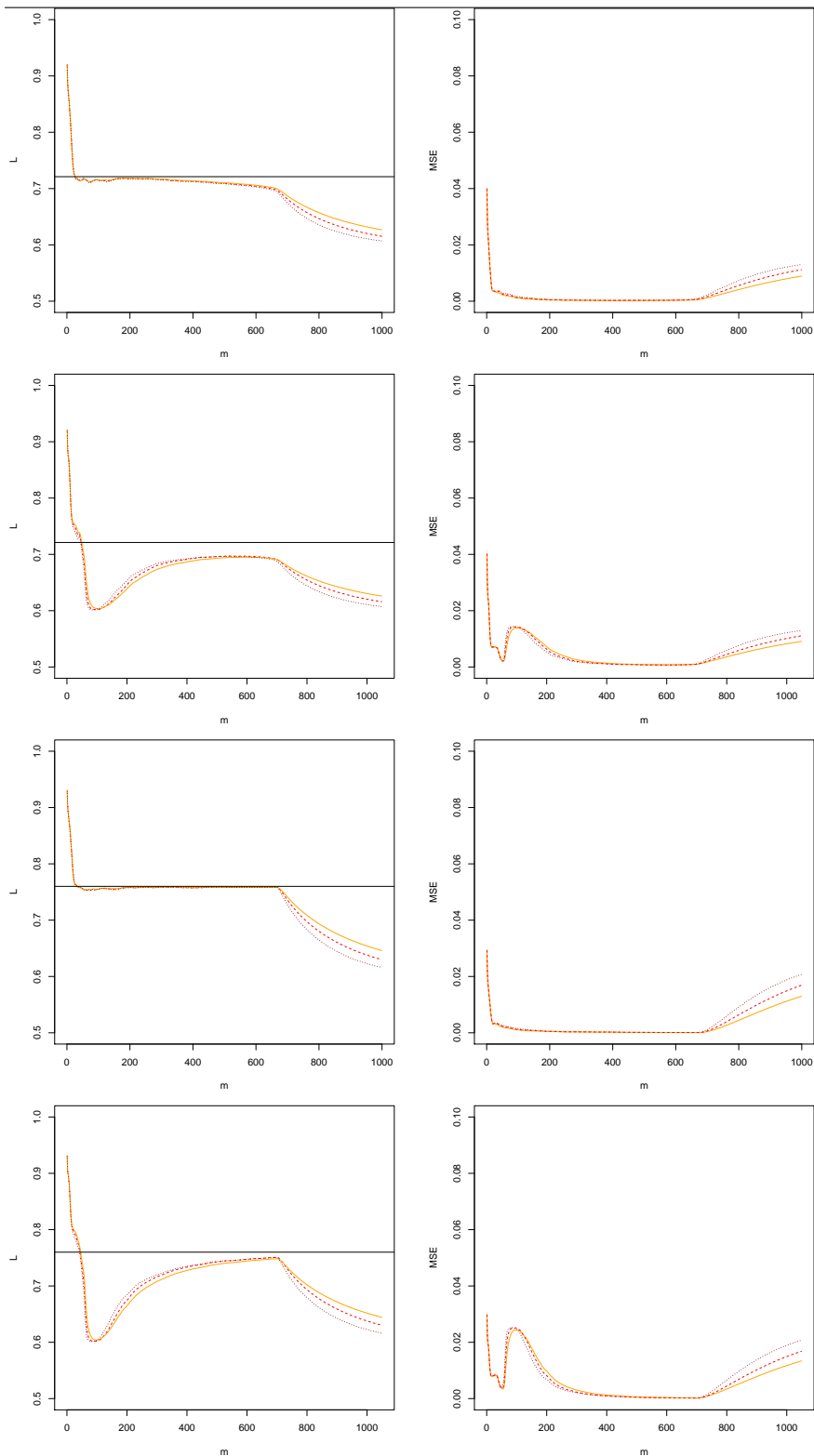


Fig. 5 Comparison between our robust and bias-corrected estimator $\hat{L}_m(0.4, 0.6)$ with $\alpha = 0.25$ (full line), 0.5 (dashed line) and 1 (dotted line). Here $\tau = 1$. Logistic distribution with 0% of contamination (first row), 5% (second row) and BPII(3) distribution with 0% of contamination (third row), 5% (fourth row). On the left: mean and on the right: MSE based on 500 replications. The full horizontal line on the left is the true value of the parameter.