

Robust and bias-corrected estimation of the coefficient of tail dependence

Christophe Dutang ^{*}
Yuri Goegebeur [†]
Armelle Guillou [‡]

January 29, 2014

Abstract

We introduce a robust and asymptotically unbiased estimator for the coefficient of tail dependence in multivariate extreme value statistics. The estimator is obtained by fitting a second order model to the data by means of the minimum density power divergence criterion. The asymptotic properties of the estimator are investigated. The efficiency of our methodology is illustrated on a small simulation study and by a real dataset from the actuarial context.

Keywords: bias-correction, tail dependence, robustness, tail quantile process

1 Introduction

Multivariate extreme value statistics deals with the estimation of the tail of a multivariate distribution function based on a random sample. Of particular interest is the estimation of the extremal dependence between two or more variables. Modelling tail dependence is a crucial issue in actuarial science (see e.g. Joe, 2011), firstly, because of the forthcoming Solvency II regulation framework which will oblige insurers and mutuals to compute 99.5% quantiles. Secondly, tail dependence can be used in the daily work of actuaries, for instance for pricing an excess-of-loss reinsurance treaty (see Cebrian *et al.*, 2003, and the references therein), and for approximating very large quantiles of the distribution of the sums of possibly dependent risks (Barbe *et al.*, 2006). In finance, obvious applications also arise, see e.g. Charpentier and Juri (2006), and Poon *et al.* (2004). Therefore, accurate modelling of extremal events is needed to better understand the relationship of possibly dependent risks at the tail.

^{*}Laboratoire Manceau de Mathématique, Université du Maine, Avenue Olivier Messiaen, 72000 - Le Mans, France (email: christophe.dutang@univ-lemans.fr).

[†]Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: yuri.goegebeur@imada.sdu.dk).

[‡]Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France (email: armelle.guillou@math.unistra.fr).

A full characterization of the extremal dependence between variables can be obtained from functions like the spectral distribution function or the Pickands dependence function. We refer to Beirlant *et al.* (2004), and de Haan and Ferreira (2006), and the references therein, for more details about this approach. Alternatively, similar to classical statistics one can try and summarize the extremal dependency in a number of well chosen coefficients that give a representative picture of the full dependency structure. In this paper we will consider the estimation of such a dependency coefficient, namely the coefficient of tail dependence.

The extremal dependence between the components of a continuous random vector (X, Y) with unit Fréchet margins (note that this can be assumed without loss of generality) can be analyzed with the model of Ledford and Tawn (1997):

$$\mathbb{P}(X > x, Y > y) = x^{-d_1} y^{-d_2} \ell(x, y), \quad x, y > 0,$$

where $d_1, d_2 > 0$ and ℓ is a bivariate slowly varying function, i.e.

$$\frac{\ell(tx, ty)}{\ell(t, t)} \rightarrow \zeta(x, y) \text{ as } t \rightarrow \infty, \text{ for all } x, y > 0,$$

and the function ζ is homogeneous of order zero. The parameter $\eta := (d_1 + d_2)^{-1}$ is called the coefficient of tail dependence. It satisfies $\eta \in (0, 1]$, and larger values of it indicate a stronger extremal dependence. As we can imagine, several attempts have been made to estimate η from data. Since

$$\mathbb{P}(\min(X, Y) > z) = \mathbb{P}(X > z, Y > z) = z^{-1/\eta} \ell(z, z),$$

i.e. the transformed variable $\min(X, Y)$ follows a Pareto-type model with index $1/\eta$, one can estimate η with classical estimators for the extreme value index like the Hill (Hill, 1975), Pickands (Pickands, 1975) or moment estimator (Dekkers *et al.*, 1989). However, this type of estimators typically suffers from bias and also they are not robust with respect to outliers. These issues will be addressed in the present paper.

In order to obtain a bias-corrected estimator we will, as usual in extreme value statistics, invoke a second order condition. In particular we will work under the following condition from Draisma *et al.* (2004), which can be seen as an extension of the above discussed Ledford and Tawn condition.

Condition \mathcal{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

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

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 $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$.

Essentially, this condition is a second order multivariate regular variation condition on the function $R(x, y) := \mathbb{P}(1 - F_X(X) < x, 1 - F_Y(Y) < y)$. It can be shown that $R(t, t)$ is regularly varying at zero with index $1/\eta$, $|q_1|$ is regularly varying at zero with index $\tau \geq 0$, and that the function c is homogeneous of order $1/\eta$, that is $c(tx, ty) = t^{1/\eta}c(x, y)$. Also, $c_1(x, x) = x^{1/\eta}(x^\tau - 1)/\tau$.

The robust and asymptotically unbiased estimator for η will be derived from a second order model obtained from condition (\mathcal{SO}) , which will be fitted to the data by the minimum density power divergence (MDPD) criterion. The specific second order model will be introduced in the next section. The density power divergence criterion was originally introduced by Basu *et al.* (1998) for the purpose of developing a robust estimation method. In particular, the density power divergence between density functions f and h is given by

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

Note 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 f be the true density function of the random variable under consideration. The idea is then to estimate θ by minimizing an empirical version of Δ_α based on a random sample Z_1, \dots, Z_n from f : if $\alpha > 0$ one considers

$$\widehat{\Delta}_\alpha(\theta) := \int_{\mathbb{R}} h^{1+\alpha}(z) dz - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n h^\alpha(Z_i),$$

whereas for $\alpha = 0$

$$\widehat{\Delta}_0(\theta) := -\frac{1}{n} \sum_{i=1}^n \log h(Z_i).$$

For $\alpha = 0$, one fits the model h to the data using the maximum likelihood method. The parameter α controls the trade-off between efficiency and robustness of the MDPD estimator: the estimator becomes more efficient but less robust against outliers as α gets closer to zero, whereas for increasing α the robustness increases and the efficiency decreases.

In Beirlant *et al.* (2011), an asymptotically unbiased estimator for η was proposed, based on fitting the extended Pareto distribution with the method of maximum likelihood to properly transformed random variables. Goegebeur and Guillou (2013) obtained asymptotic unbiasedness by taking a properly weighted sum of two biased estimators for η . However, these methods are not robust with respect to outliers.

The plan of the paper is as follows. In Section 2, we will introduce a second order Pareto-type model, which is derived from a submodel of condition (\mathcal{SO}) , and discuss the robust estimation method. In Section 3, the asymptotic properties of our estimator are established. In particular, we will establish a uniform consistency result for the tail quantile process and use this to obtain the limiting distribution of the robust estimator for η . The estimation method is illustrated with

a small simulation study in Section 4, and a real dataset concerning workers' compensation in Section 5. Section 6 contains some concluding remarks. The proofs of all results are deferred to the Appendix.

2 Model and assumptions

Let (X, Y) be a bivariate random vector with continuous marginal distributions satisfying

$$\mathbb{P}(1 - F_X(X) < x, 1 - F_Y(Y) < y) = x^{d_1} y^{d_2} g(x, y) \left(1 + \frac{1}{\eta} \delta(x, y) \right), \quad x \geq 0, y \geq 0, \quad (2)$$

where d_1, d_2 are positive constants, $\eta := (d_1 + d_2)^{-1} \in (0, 1)$ is the tail dependence coefficient, g is a continuous function that is homogeneous of order 0 and δ is 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

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

for all $x, y \geq 0$. We assume additionally that ξ is continuous, homogeneous of order $\tau > 0$, and that the convergence is uniform on $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$. Note that we exclude the case $\eta = 1$, as was also done in Beirlant and Vandewalle (2002), Beirlant *et al.* (2011), and Goegebeur and Guillou (2013). For the sequel, it is instructive to keep the following elementary property in mind.

Lemma 1 *Model (2) satisfies assumption (SO).*

Many commonly used joint distribution functions satisfy model (2). Note that this model is in fact a condition on the copula function C . Indeed, one easily verifies that

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

Example 1: the Farlie Gumbel Morgenstern (FGM) distribution

The Farlie Gumbel Morgenstern copula function is given by

$$C(x, y) = xy [1 + \beta(1 - x)(1 - y)], \quad (x, y) \in [0, 1]^2,$$

with $\beta \in [-1, 1]$. Straightforward calculations lead to

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

In the case where $\beta \in (-1, 1]$, we get that $d_1 = d_2 = 1$, $\eta = 1/2$, $g(x, y) = 1 + \beta$, $\delta(x, y) = -\eta\beta(x + y - xy)/(1 + \beta)$, $\xi(x, y) = (x + y)/2$ and $\tau = 1$. In terms of condition (SO) this gives then $c(x, y) = xy$, $c_1(x, y) = xy(x + y - 2)/2$ and $q_1(t) \sim -2\beta t/(1 + \beta)$.

In the case $\beta = -1$, we have $d_1 = d_2 = 3/2$, $\eta = 1/3$, $g(x, y) = (x + y)/\sqrt{xy}$, $\delta(x, y) = -xy/(3(x + y))$, $\xi(x, y) = 2xy/(x + y)$ and $\tau = 1$. Condition (SO) is also satisfied with

$c(x, y) = xy(x + y)/2$, $c_1(x, y) = xy(2xy - x - y)/2$ and $q_1(t) \sim -t/2$.

Example 2: the Frank distribution

The copula function for the Frank distribution is given by

$$C(x, y) = -\frac{1}{\beta} \log \left[1 - \frac{(1 - e^{-\beta x})(1 - e^{-\beta y})}{1 - e^{-\beta}} \right], \quad (x, y) \in [0, 1]^2,$$

where $\beta > 0$. Tedious computations based on expansions of the above copula function lead to the following approximation

$$\mathbb{P}(1 - F_X(X) < x, 1 - F_Y(Y) < y) = \frac{\beta}{1 - e^{-\beta}} xy \left[1 - \frac{\beta}{2} (x + y) + o(x) + o(y) \right],$$

from which we deduce that $d_1 = d_2 = 1$, $\eta = 1/2$, $g(x, y) = \beta/(1 - e^{-\beta})$, $\delta(x, y) \sim -\eta\beta(x + y)/2$, $\xi(x, y) = (x + y)/2$ and $\tau = 1$. In terms of (\mathcal{SO}) this gives $c(x, y) = xy$, $c_1(x, y) = xy(x + y - 2)/2$, $q_1(t) \sim -\beta t$.

Note that the FGM and the Frank copulas have been widely used in actuarial science and in finance, see e.g. Frees and Valdez (1998), Klugman and Parsa (1999), Embrechts *et al.* (2003), which provides further motivation for the importance of our model (2) in these two fields of applications.

For convenience we assume that the marginal distributions are unit Pareto. In this case model (2) becomes

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

where $g^*(x, y) := g(1/x, 1/y)$ and $\delta^*(x, y) := \delta(1/x, 1/y)$. This model is a slight generalization of the second order model that was considered in Beirlant *et al.* (2011), and will form the basis for the estimation procedure to be developed in this paper. Note that one can write

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

where $\omega := x/(x + y)$, $\omega \in (0, 1)$, can be interpreted as being a radial parameter. Thus, we consider the transformed variable $Z_\omega := \min(X, \frac{\omega}{1 - \omega} Y)$. For this variable one easily derives the survival function, given by

$$\mathbb{P}(Z_\omega > z) = C_\omega z^{-1/\eta} \left(1 + \frac{1}{\eta} \delta_\omega(z) \right), \quad z > 0, \quad (4)$$

where $C_\omega := (\omega/(1 - \omega))^{d_2} g^*(1, (1 - \omega)/\omega)$, and $|\delta_\omega|$ is a function of regular variation with index $-\tau$. This second order condition is identical to the one used in the univariate framework in Beirlant *et al.* (2009) and Dierckx *et al.* (2013), and therefore, as shown in these papers, one

can approximate the distribution of Z_ω/u , conditionally on $Z_\omega > u$, for large u , by the extended Pareto distribution. Formally, for u large

$$\frac{\mathbb{P}(Z_\omega > uz)}{\mathbb{P}(Z_\omega > u)} \approx \bar{H}(z; \eta, \delta_\omega(u), \tau) := \begin{cases} 1, & z \leq 1, \\ [z(1 + \delta_\omega(u) - \delta_\omega(u)z^{-\tau})]^{-1/\eta}, & z > 1. \end{cases}$$

Using this property, one can estimate η by fitting h , the density function associated with H , to relative excesses over some large threshold u . From the univariate context it is well-known that by taking the second order structure into account one obtains estimators with better bias properties than so-called first order estimators. We refer to Feuerverger and Hall (1999), Beirlant *et al.* (1999), and more recently Gomes *et al.* (2008).

Specifically, for a sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent random vectors from model (2), one transforms into unit Pareto margins by using the empirical distribution functions of the X and Y observations. This gives

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

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. The parameters η and δ_ω of the extended Pareto distribution are estimated by fitting the density function h to the relative excesses $Z_j := \tilde{Z}_{\omega, n-m+j, n} / \tilde{Z}_{\omega, n-m, n}$, $j = 1, \dots, m$, where $1 \leq m \leq n-1$, and $\tilde{Z}_{\omega, 1, n} \leq \dots \leq \tilde{Z}_{\omega, n, n}$ are the order statistics of $\tilde{Z}_{\omega, 1}, \dots, \tilde{Z}_{\omega, n}$, using the minimum density power divergence criterion. The density function of the extended Pareto distribution is given by

$$h(z; \eta, \delta_\omega, \tau) = \frac{1}{\eta} z^{-1/\eta-1} [1 + \delta_\omega(1 - z^{-\tau})]^{-1/\eta-1} [1 + \delta_\omega(1 - (1-\tau)z^{-\tau})], \quad z > 1,$$

where $\eta > 0$, $\tau > 0$, and $\delta_\omega > \max\{-1, -1/\tau\}$. Remember that the parameter δ_ω reflects in fact the function $\delta_\omega(u)$, where $\delta_\omega(u) \rightarrow 0$ as $u \rightarrow \infty$, but we do not make this dependence on the threshold explicit in the notation. In the sequel we use the common reparametrization $\rho = -\tau\eta$. The MDPD estimator for η and δ_ω satisfies the estimating equations

$$0 = \int_1^\infty h^\alpha(z; \eta, \delta_\omega, \rho) \frac{\partial h(z; \eta, \delta_\omega, \rho)}{\partial \eta} dz - \frac{1}{m} \sum_{j=1}^m h^{\alpha-1}(Z_j; \eta, \delta_\omega, \rho) \frac{\partial h(Z_j; \eta, \delta_\omega, \rho)}{\partial \eta}, \quad (5)$$

$$0 = \int_1^\infty h^\alpha(z; \eta, \delta_\omega, \rho) \frac{\partial h(z; \eta, \delta_\omega, \rho)}{\partial \delta_\omega} dz - \frac{1}{m} \sum_{j=1}^m h^{\alpha-1}(Z_j; \eta, \delta_\omega, \rho) \frac{\partial h(Z_j; \eta, \delta_\omega, \rho)}{\partial \delta_\omega}. \quad (6)$$

Note that only the parameters η and δ_ω are estimated with the MDPD method. The parameter ρ will in this paper be fixed at some value, either the correct value or a mis-specified one.

3 Asymptotic results

Consider the random variables $\tilde{Z}_{\omega, 1}, \dots, \tilde{Z}_{\omega, n}$, with order statistics $\tilde{Z}_{\omega, 1, n} \leq \dots \leq \tilde{Z}_{\omega, n, n}$. In order to obtain the limiting distribution of the MDPD estimators for η and δ_ω we need the

following result concerning the tail quantile process $Q_n(t) := \tilde{Z}_{\omega, n-[mt], n}$, $0 < t < n/m$, where m is an intermediate sequence, i.e. $m \rightarrow \infty$ as $n \rightarrow \infty$ with $m = o(n)$. Set $c_\omega := c(1, \omega/(1-\omega))$, $\xi_\omega := \xi(1, \omega/(1-\omega))$, $q(t) := \mathbb{P}(1 - F_X(X) < t, 1 - F_Y(Y) < t)$ and let $k/n = q^{\leftarrow}(m/n)$, where it is assumed that $k \rightarrow \infty$.

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 (2) such that the function c given in (8) has continuous first order partial derivatives. For $m, k \rightarrow \infty$ as $n \rightarrow \infty$ such that $\sqrt{m}q_1(k/n) \rightarrow \lambda \in \mathbb{R}$ we have that there exist suitable versions of Q_n , a suitable process $\bar{W}(\cdot/c_\omega)$, equal in distribution to a standard Brownian motion such that for all $t_0, \varepsilon > 0$*

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

This result can now be used in order to study the statistics

$$\begin{cases} A_{m,n}^{(1)}(s_1) & := \frac{1}{m} \sum_{j=1}^m \left(\frac{\tilde{Z}_{\omega, n-j+1, n}}{\tilde{Z}_{\omega, n-m, n}} \right)^{s_1} = \int_0^1 \left(\frac{Q_n(t)}{Q_n(1)} \right)^{s_1} dt, \\ A_{m,n}^{(2)}(s_2) & := \frac{1}{m} \sum_{j=1}^m \left(\frac{\tilde{Z}_{\omega, n-j+1, n}}{\tilde{Z}_{\omega, n-m, n}} \right)^{s_2} \log \frac{\tilde{Z}_{\omega, n-j+1, n}}{\tilde{Z}_{\omega, n-m, n}} = \int_0^1 \left(\frac{Q_n(t)}{Q_n(1)} \right)^{s_2} \log \frac{Q_n(t)}{Q_n(1)} dt, \\ A_{m,n}^{(3)}(s_3) & := \frac{1}{m} \sum_{j=1}^m \left(\frac{\tilde{Z}_{\omega, n-j+1, n}}{\tilde{Z}_{\omega, n-m, n}} \right)^{s_3} \left(\log \frac{\tilde{Z}_{\omega, n-j+1, n}}{\tilde{Z}_{\omega, n-m, n}} \right)^2 = \int_0^1 \left(\frac{Q_n(t)}{Q_n(1)} \right)^{s_3} \left(\log \frac{Q_n(t)}{Q_n(1)} \right)^2 dt, \end{cases}$$

with $s_1, s_2, s_3 \leq 0$. In fact, the estimating equations (5) and (6) depend only on the data through statistics of these proposed forms.

Theorem 2 *Under the assumptions of Theorem 1, we have that*

$$\sqrt{m} \begin{bmatrix} A_{m,n}^{(1)}(s_1) - \frac{1}{1-\eta s_1} \\ A_{m,n}^{(2)}(s_2) - \frac{\eta}{(1-\eta s_2)^2} \\ A_{m,n}^{(3)}(s_3) - \frac{2\eta^2}{(1-\eta s_3)^3} \end{bmatrix} \rightsquigarrow N_3(\lambda \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where

$$\boldsymbol{\mu} := -\xi_\omega c_\omega^{-\eta\tau} \begin{bmatrix} \frac{s_1 \eta^2}{(1-\eta s_1)(1-\eta(s_1-\tau))} \\ \frac{\eta^2(1+\eta\tau-\eta^2 s_2^2)}{(1-\eta s_2)^2(1-\eta(s_2-\tau))^2} \\ \frac{2\eta^2}{\tau} \left[\frac{1}{(1-\eta s_3)^3} - \frac{1+\eta\tau}{(1-\eta(s_3-\tau))^3} \right] \end{bmatrix},$$

and where Σ has elements given by

$$\begin{aligned}
\sigma_{11} &:= \frac{\eta^2 s_1^2}{(1 - s_1 \eta)^2 (1 - 2s_1 \eta)}, \\
\sigma_{22} &:= \eta^2 \left\{ \frac{2}{[1 - 2s_2 \eta]^3} - \frac{1}{(1 - s_2 \eta)^4} \right\}, \\
\sigma_{33} &:= \eta^4 \left\{ \frac{4!}{[1 - 2s_3 \eta]^5} - \frac{4}{(1 - s_3 \eta)^6} \right\}, \\
\sigma_{12} &:= \eta \left\{ \frac{1}{[1 - (s_1 + s_2) \eta]^2} - \frac{1}{(1 - s_1 \eta)(1 - s_2 \eta)^2} \right\}, \\
\sigma_{13} &:= 2\eta^2 \left\{ \frac{1}{[1 - (s_1 + s_3) \eta]^3} - \frac{1}{(1 - s_1 \eta)(1 - s_3 \eta)^3} \right\}, \\
\sigma_{23} &:= 2\eta^3 \left\{ \frac{3}{[1 - (s_2 + s_3) \eta]^4} - \frac{1}{(1 - s_2 \eta)^2 (1 - s_3 \eta)^3} \right\}.
\end{aligned}$$

From now on we will denote the true value of η and ρ by η_0 and ρ_0 , respectively. Let $\delta_{\omega, n} := \delta_{\omega}(\tilde{Z}_{\omega, n-m, n})$.

Theorem 3 *Under the conditions of Theorem 1 we have that*

$$\sqrt{m} \begin{bmatrix} \hat{\eta}_m - \eta_0 \\ \hat{\delta}_{\omega, n} - \delta_{\omega, n} \end{bmatrix} \rightsquigarrow N_2(\mathbf{0}, \mathbb{C}^{-1}(\rho_0) \mathbb{B}(\rho_0) \mathbb{D}(\rho_0) \mathbb{B}'(\rho_0) \mathbb{C}^{-1}(\rho_0)),$$

where

$$\mathbb{B}(\rho_0) := \eta_0^{-\alpha-2} \begin{bmatrix} \eta_0 & 0 & -1 \\ \eta_0 & -\eta_0(1 - \rho_0) & 0 \end{bmatrix},$$

the symmetric matrix $\mathbb{C}(\rho_0)$ has elements

$$\begin{aligned}
c_{11}(\rho_0) &:= \eta_0^{-\alpha-2} \frac{1 + \alpha^2(1 + \eta_0)^2}{[1 + \alpha(1 + \eta_0)]^3}, \\
c_{12}(\rho_0) &:= \eta_0^{-\alpha-2} \frac{\rho_0(1 - \rho_0)[1 + \alpha(1 + \eta_0) + \alpha^2(1 + \eta_0)^2] + \alpha^3 \rho_0(1 + \eta_0)^3}{[1 + \alpha(1 + \eta_0)]^2 [1 - \rho_0 + \alpha(1 + \eta_0)]^2}, \\
c_{22}(\rho_0) &:= \eta_0^{-\alpha-2} \frac{(1 - \rho_0)\rho_0^2 + \alpha\rho_0^2(1 + \eta_0)[\alpha(1 + \eta_0) - \rho_0]}{[1 + \alpha(1 + \eta_0)][1 - \rho_0 + \alpha(1 + \eta_0)][1 - 2\rho_0 + \alpha(1 + \eta_0)]},
\end{aligned}$$

and the symmetric matrix $\mathbb{D}(\rho_0)$ has elements

$$\begin{aligned}
d_{11}(\rho_0) &:= \frac{\alpha^2(1+\eta_0)^2}{[1+\alpha(1+\eta_0)]^2[1+2\alpha(1+\eta_0)]}, \\
d_{21}(\rho_0) &:= \frac{\alpha(1+\eta_0)[\alpha(1+\eta_0)-\rho_0]}{[1+\alpha(1+\eta_0)][1-\rho_0+\alpha(1+\eta_0)][1-\rho_0+2\alpha(1+\eta_0)]}, \\
d_{22}(\rho_0) &:= \frac{[\alpha(1+\eta_0)-\rho_0]^2}{[1-\rho_0+\alpha(1+\eta_0)]^2[1-2\rho_0+2\alpha(1+\eta_0)]}, \\
d_{31}(\rho_0) &:= \eta_0 \left(\frac{1}{[1+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^3} \right), \\
d_{32}(\rho_0) &:= \eta_0 \left(\frac{1}{[1-\rho_0+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^2[1-\rho_0+\alpha(1+\eta_0)]} \right), \\
d_{33}(\rho_0) &:= \eta_0^2 \left(\frac{2}{[1+2\alpha(1+\eta_0)]^3} - \frac{1}{[1+\alpha(1+\eta_0)]^4} \right).
\end{aligned}$$

The proof of this theorem is identical to that of Theorem 2 in Dierckx *et al.* (2013), and therefore it is omitted here. The estimator $\widehat{\eta}_n$ is asymptotically unbiased in the sense that the mean of the limiting distribution is zero, whatever the value of λ . Now, we consider the asymptotic behavior of the MDPD estimator when the parameter ρ is fixed at some value $\widetilde{\rho}$, possibly mis-specified.

Theorem 4 *Under the conditions of Theorem 1 we have that*

$$\sqrt{m} \begin{bmatrix} \widehat{\eta}_n - \eta_0 \\ \widehat{\delta}_{\omega,n} \end{bmatrix} \rightsquigarrow N_2 \left(-\lambda \mathbb{C}^{-1}(\widetilde{\rho}) \mathbb{B}(\widetilde{\rho}) \widetilde{\boldsymbol{\mu}}, \mathbb{C}^{-1}(\widetilde{\rho}) \mathbb{B}(\widetilde{\rho}) \mathbb{D}(\widetilde{\rho}) \mathbb{B}'(\widetilde{\rho}) \mathbb{C}^{-1}(\widetilde{\rho}) \right),$$

where the vector $\widetilde{\boldsymbol{\mu}}$ has elements

$$\begin{aligned}
\widetilde{\mu}_1 &:= \xi_\omega c_\omega^{\rho_0} \frac{\alpha \eta_0 (1 + \eta_0)}{[1 + \alpha(1 + \eta_0)][1 - \rho_0 + \alpha(1 + \eta_0)]}, \\
\widetilde{\mu}_2 &:= \xi_\omega c_\omega^{\rho_0} \frac{\eta_0 [\alpha(1 + \eta_0) - \widetilde{\rho}]}{[1 - \widetilde{\rho} + \alpha(1 + \eta_0)][1 - \rho_0 - \widetilde{\rho} + \alpha(1 + \eta_0)]}, \\
\widetilde{\mu}_3 &:= -\xi_\omega c_\omega^{\rho_0} \frac{\eta_0^2 [1 - \rho_0 - \alpha^2(1 + \eta_0)^2]}{[1 + \alpha(1 + \eta_0)]^2 [1 - \rho_0 + \alpha(1 + \eta_0)]^2}.
\end{aligned}$$

For the proof of this theorem we refer to the proof of Proposition 1 in Dierckx *et al.* (2013). As is clear from Theorem 4, if one mis-specifies the parameter ρ , then the mean of the limiting distribution is not necessarily zero, and hence one possibly loses the bias-correction. However, as shown in the simulation experiments of Goegebeur and Guillou (2013), the estimators are not very sensitive to such a mis-specification and setting $\widetilde{\rho} = -1$ often results in an η -estimator that outperforms estimators that are not corrected for bias. Also in Beirlant *et al.* (2011) the second order rate parameter was fixed at -1. A third possibility for dealing with the second order parameter ρ consists in estimating it externally. In the univariate framework several estimators for this parameter have been proposed that have a good practical performance, see e.g. Fraga Alves *et al.* (2003), Ciuperca and Mercadier (2010), Goegebeur *et al.* (2010) and Deme *et al.* (2013), but at this stage it is unclear if the ideas from the univariate context can be transported without any problem to the multivariate case, and therefore we do not explore this option theoretically in the present paper.

4 Simulation study

In this section, we illustrate the finite sample properties of our robust and bias-corrected estimator $\hat{\eta}_n$ through a small simulation study. In order to make the dependence on the tuning parameter α explicit, we use from now on the notation $\hat{\eta}_{n,\alpha}$.

At first, we simulate n pairs $(X_i, Y_i), i = 1, \dots, n$, independently from a FGM copula with Fréchet marginals. Two values of β are used, -1 and 1 , which allows to have two different values of η , $1/3$ and $1/2$ respectively. Similarly, we simulate n pairs from a Frank copula with $\beta = 2$ for which $\eta = 1/2$. Then we transform the margins into (approximately) unit Pareto using the empirical distribution functions. This gives us $Z_{\omega,i}, i = 1, \dots, n$. In all our simulations $\omega = 1/2$. Finally, we minimize the empirical density power divergence $\hat{\Delta}_\alpha$.

In each setting (FGM or Frank copulas), we also contaminated the sample as follows:

- we simulate independently \tilde{X}_j and $\tilde{Y}_j, j = 1, \dots, n_0 := \lfloor n\varepsilon \rfloor$, from a unit Fréchet distribution;
- all our above methodology is applied to our new $n + n_0$ pairs $(X_1, Y_1), \dots, (X_n, Y_n), (\tilde{X}_1 + X_{n,n}, \tilde{Y}_1 + Y_{n,n}), \dots, (\tilde{X}_{n_0} + X_{n,n}, \tilde{Y}_{n_0} + Y_{n,n})$, where $X_{n,n}$ and $Y_{n,n}$ are sample maxima of the X and Y observations, respectively.

The percentage of contamination is set to $\varepsilon = 1\%$ and 2% , while $n = 100$, and the procedure is repeated 1024 times. Thus n_0 is 1 or 2. Note that we run the computations $2^{10} = 1024$ times since we used parallelization on a 8-core computer.

Figure 1 shows the mean (left panels of the figure) and the mean squared error (MSE, right panels of the figure) of our estimator $\hat{\eta}_{n,\alpha}$ as a function of m for different values of α : 0 (full line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line) for the FGM copula with $\beta = -1$. Figures 2 and 3 are constructed analogously for the FGM copula with $\beta = 1$ and the Frank copula with $\beta = 2$, respectively. From our simulations, we can conclude the following:

- on uncontaminated datasets the choice $\alpha = 0$ seems to perform the best, although there are no big differences. This is not surprising since $\hat{\eta}_{n,0}$ corresponds to the maximum likelihood estimator which is well-known to be efficient, but not robust. Note also that the estimators have a very nice performance, especially if one takes the small sample size of $n = 100$ into account;
- the lack of robustness of the maximum likelihood estimator is illustrated in the contaminated cases where we clearly observe the superiority of $\hat{\eta}_{n,\alpha}$ with $\alpha > 0$. In view of the three figures, we conclude that for the contamination we consider, the value $\alpha = 0.5$ seems to be the most appropriate in terms of bias and MSE;
- as expected, increasing the fraction of contamination negatively affects the estimators;
- the superiority of our estimator $\hat{\eta}_{n,0.5}$ compared to the maximum likelihood estimator ($\alpha = 0$) is stronger in case $\eta = 1/3$ than in case $\eta = 1/2$. A possible explanation for this is that in the case $\eta = 1/3$ one disturbs the sample by outliers that have a different dependence structure, whereas for $\eta = 1/2$ the data are only disturbed by outliers, but their asymptotic dependence is the same as for the uncontaminated data;
- as is well-known in the extreme value framework, bias correction leads to estimators with rather stable sample paths as a function of m , which alleviates the choice of m to some extent.

We also redraw our simulation study for $n = 1000$ and also we take different values of ω . Since the conclusions remain the same, the figures are omitted for brevity from the paper.

5 Illustration: workers' compensation data

We consider the example of the workers' compensation data examined by Klugman (1992), Frees *et al.* (2001) and Frees (2010). The dataset is originally from the National Council on Compensation Insurance (NCCI). This database contains records of losses due to permanent or partial disability claims for workers' compensation insurance in the US. For each claim amount, the payroll is available as a measure of exposure units. A total of 847 data points is available coming from the observation of 121 risk classes over 7 years. Possible time dependency among the data points is ignored in this paper.

In Figures 4 (a) and 4 (b) we show the scatterplot of the original data and the data transformed to unit Pareto margins, respectively. We can see some points far away from the cloud and the question is whether these points are outliers, that could have an adverse effect on the estimation of tail dependence. To have an idea about this question, we plot our estimator for η as a function of m with three different values of α : 0 (solid line), 0.1 (dashed line) and 0.2 (dotted line). Indeed, if the results for the maximum likelihood method ($\alpha = 0$) are similar to those obtained with the robust method ($\alpha > 0$) this indicates that the dataset does not contain disturbing outliers. Figure 4 (c) represents the Pareto quantile plot of the $\tilde{Z}_{\omega,i}$, $i = 1, \dots, n$. This plot is more or less linear in the largest observations indicating a good fit of the model (4) to our data. In Figure 4 (d) we show the η estimates obtained on the original data with $\omega = 1/2$. For small values of m the three estimators are close to each other, which seems to indicate that on the contrary of what we can believe in view of Figures 4 (a) or (b), this dataset does not contain outliers. To reinforce this idea, we contaminate the dataset according to the algorithm described in Section 4, with $\varepsilon = 1\%$ and 2% . Again ω is set at the value $1/2$. The results are given in Figures 4 (e) and (f). As is clear the estimator with $\alpha = 0.2$ is more stable for small values of m whereas the one with $\alpha = 0$, corresponding to maximum likelihood, is pulled up for small m , but decreases rather quickly which is in accordance with what we already observed in the simulation section. The estimator with $\alpha = 0.1$ provides more robustness than the maximum likelihood estimator, though it is not enough for the considered contamination.

6 Conclusion

In this paper we introduced a robust and asymptotically unbiased estimator for the coefficient of tail dependence in extreme value statistics. The estimator is obtained from a second order bivariate tail model that is fitted to minima of properly transformed observations using the MDPD technique. A uniform convergence in probability result for the tail quantile process of the transformed observations is introduced, which is of general importance and in our context it is crucial for the theoretical study of the estimator. In future research we will extend the

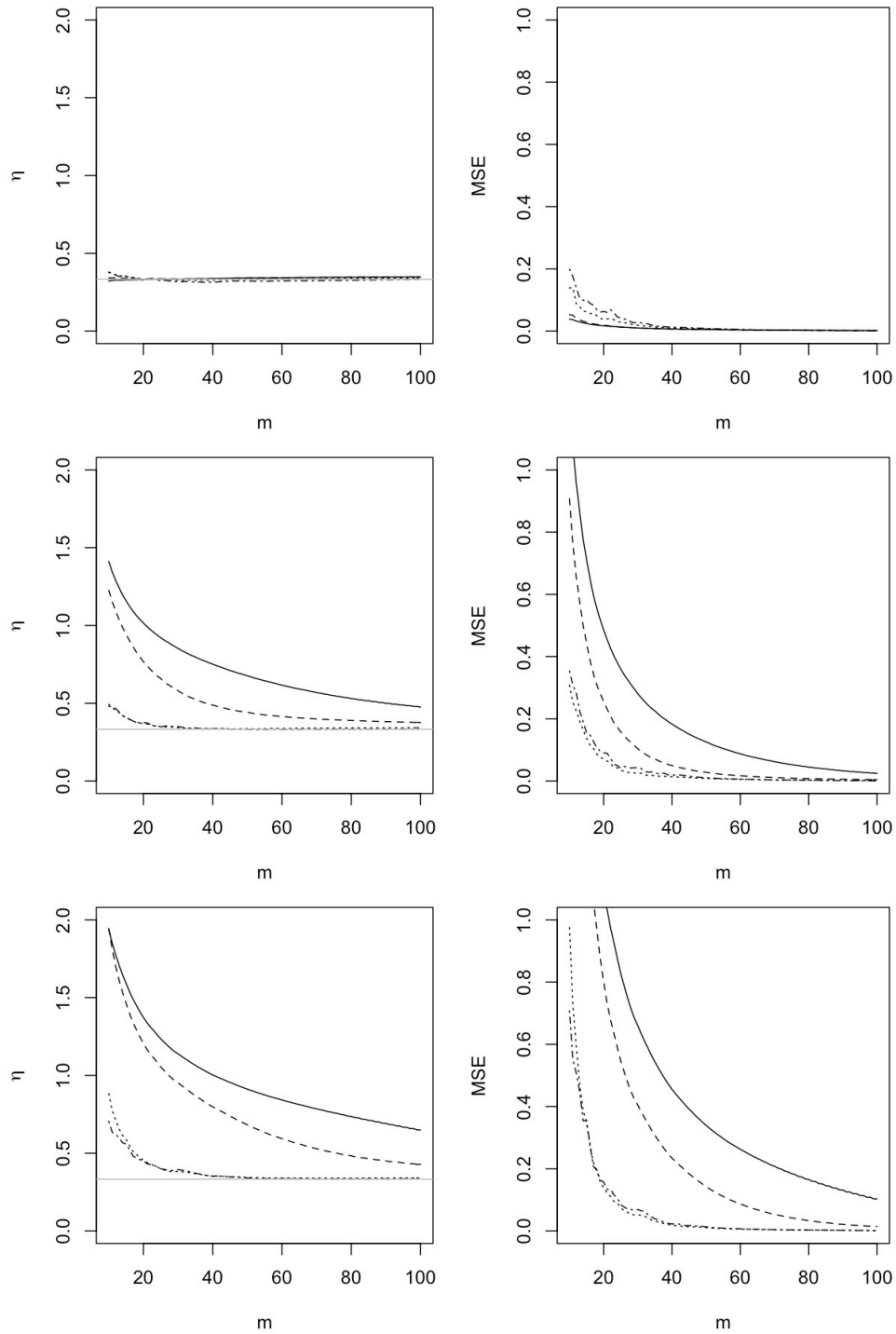


Figure 1: Farlie Gumbel Morgenstern copula with $\beta = -1$: mean (left), MSE (right) of $\hat{\eta}_{m,\alpha}$ for different values of α : 0 (full line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 1\%$, with 2% . The horizontal reference line in the left panels of the figure corresponds to the true value of η .

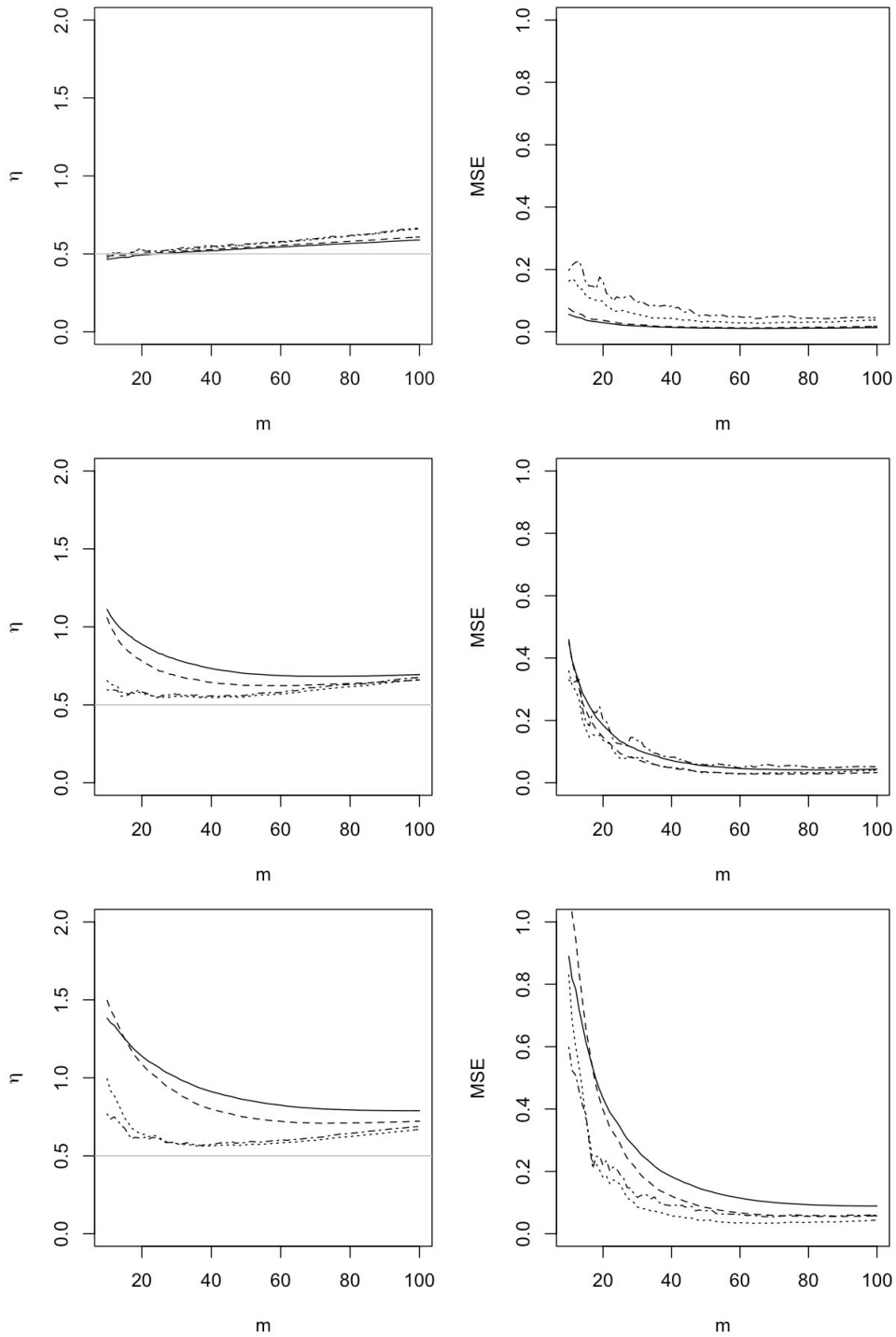


Figure 2: Farlie Gumbel Morgenstern copula with $\beta = 1$: mean (left), MSE (right) of $\hat{\eta}_{n,\alpha}$ for different values of α : 0 (full line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 1\%$, with 2% . The horizontal reference line in the left panels of the figure corresponds to the true value of η .

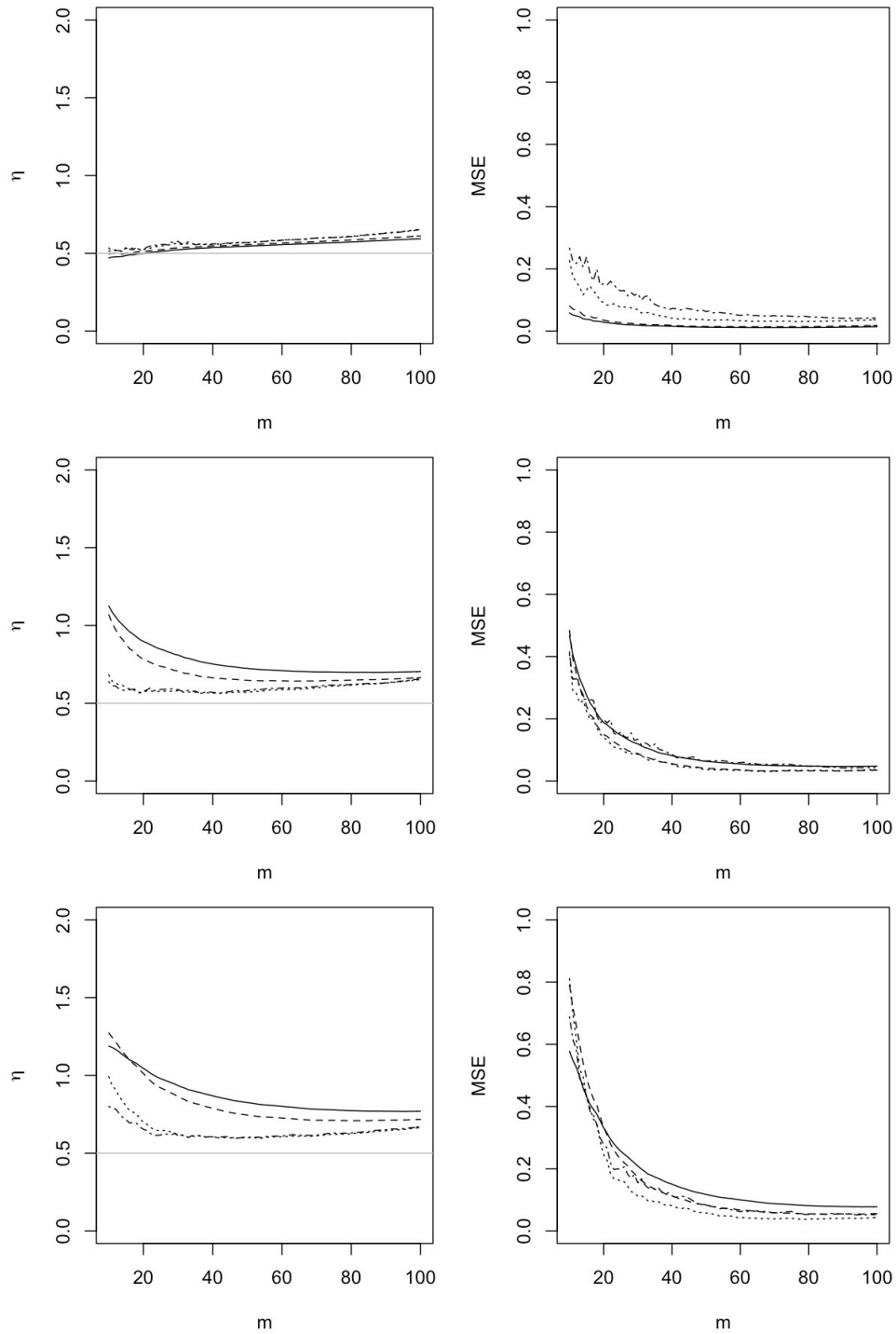


Figure 3: Frank copula with $\beta = 2$: mean (left), MSE (right) of $\hat{\eta}_{m,\alpha}$ for different values of α : 0 (full line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 1\%$, with 2% . The horizontal reference line in the left panels of the figure corresponds to the true value of η .

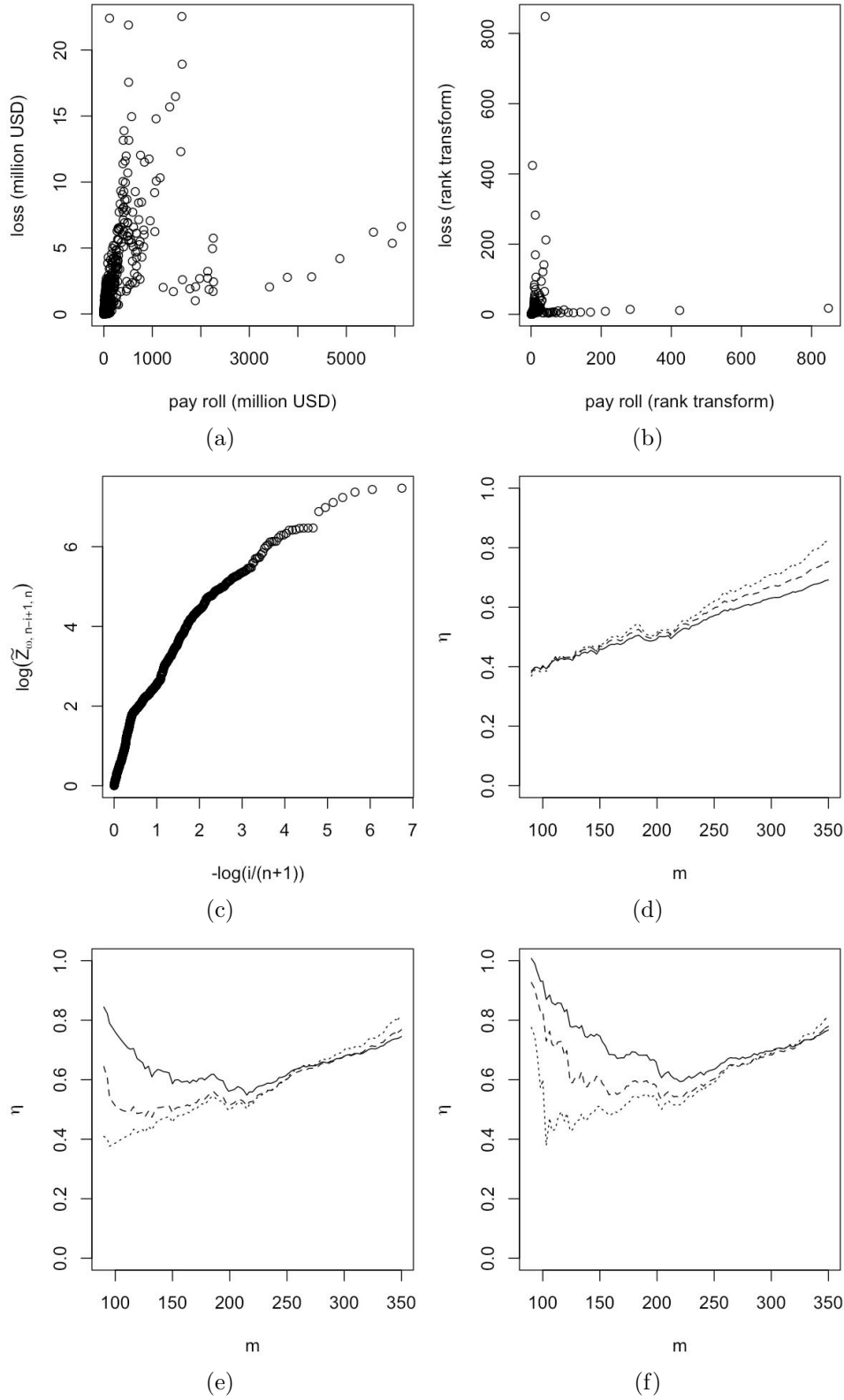


Figure 4: Workers' compensation dataset: (a) Scatterplot of the original data; (b) Scatterplot of the data transformed to unit Pareto margins; (c) Pareto quantile plot of the $\tilde{Z}_{\omega, i}$, $i = 1, \dots, n$; (d) Estimator of η with $\alpha = 0$ (solid line), $\alpha = 0.1$ (dashed line), $\alpha = 0.2$ (dotted line) in the case of no contamination; (e) Same in the case of contamination with $\varepsilon = 1\%$; (f) Same in the case of contamination with $\varepsilon = 2\%$.

developed methodology to the bias-corrected and robust estimation of probabilities of extreme failure sets.

Appendix

Proof of Lemma 1

Using model (2), one gets

$$\frac{\mathbb{P}(1 - F_X(X) < tx, 1 - F_Y(Y) < ty)}{\mathbb{P}(1 - F_X(X) < t, 1 - F_Y(Y) < t)} - c(x, y) = x^{d_1} y^{d_2} \frac{g(x, y)}{g(1, 1)} \frac{1 + \frac{1}{\eta} \delta(tx, ty)}{1 + \frac{1}{\eta} \delta(t, t)} - c(x, y). \quad (7)$$

Set

$$c(x, y) = x^{d_1} y^{d_2} \frac{g(x, y)}{g(1, 1)}, \quad (8)$$

this gives

$$c(tx, ty) = t^{d_1+d_2} x^{d_1} y^{d_2} \frac{g(x, y)}{g(1, 1)} = t^{d_1+d_2} c(x, y)$$

as required (see Draisma *et al.*, 2004). Plugging into (7) leads to

$$\begin{aligned} \frac{\mathbb{P}(1 - F_X(X) < tx, 1 - F_Y(Y) < ty)}{\mathbb{P}(1 - F_X(X) < t, 1 - F_Y(Y) < t)} - c(x, y) &= \frac{1}{\eta} \delta(t, t) c(x, y) \frac{\frac{\delta(tx, ty)}{\delta(t, t)} - 1}{1 + \frac{1}{\eta} \delta(t, t)} \\ &\sim \frac{\tau}{\eta} \delta(t, t) c(x, y) \frac{\xi(x, y) - 1}{\tau}, \quad t \downarrow 0 \end{aligned} \quad (9)$$

so we can take

$$\begin{aligned} q_1(t) &= \frac{\tau}{\eta} \delta(t, t) \\ c_1(x, y) &= c(x, y) \frac{\xi(x, y) - 1}{\tau}. \end{aligned}$$

Finally, by using the properties of bivariate regularly varying functions, one easily establishes that

$$c_1(x, x) = x^{1/\eta} \frac{x^\tau - 1}{\tau},$$

see e.g. p. 287 in Beirlant *et al.* (2004). The uniformity of the convergence in (2) applies also to (9). This concludes Lemma 1.

Proof of Theorem 1

Let $U_i := 1 - F_X(X_i)$ and $V_i := 1 - F_Y(Y_i)$, $i = 1, \dots, n$, with the convention that $U_{0,n} = V_{0,n} = 0$ and $U_{n+1,n} = V_{n+1,n} = 1$. First check that

$$\begin{aligned} \sum_{i=1}^n \mathbb{1}_{\{\tilde{Z}_{\omega,i} > r\}} &= \sum_{i=1}^n \mathbb{1}_{\left\{ \frac{n+1}{n+1-R_i^X} > r, \frac{n+1}{n+1-R_i^Y} \frac{\omega}{1-\omega} > r \right\}} \\ &= \sum_{i=1}^n \mathbb{1}_{\left\{ 1 - \frac{R_i^X}{n+1} < \frac{1}{r}, 1 - \frac{R_i^Y}{n+1} < \frac{\omega}{1-\omega} \frac{1}{r} \right\}} \\ &= \sum_{i=1}^n \mathbb{1}_{\left\{ U_i < U_{\lceil \frac{n+1}{r} \rceil, n}, V_i < V_{\lceil \frac{(n+1)\omega}{r(1-\omega)} \rceil, n} \right\}} \quad a.s. \end{aligned}$$

Thus

$$\begin{aligned} \bar{F}_n(r) &:= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\left\{ \frac{k}{n+1} \tilde{Z}_{\omega,i} > r \right\}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\left\{ U_i < U_{\lceil \frac{k}{r} \rceil, n}, V_i < V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right\}} \\ &=: \frac{1}{n} S \left(U_{\lceil \frac{k}{r} \rceil, n-}, V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n-} \right), \end{aligned}$$

where $S(x_-, y_-)$ denotes the left-hand side limit of S at (x, y) . Let $W(x, y)$ be a Gaussian process with mean zero and covariance structure defined as $\mathbb{E}(W(x_1, y_1)W(x_2, y_2)) = c(x_1 \wedge x_2, y_1 \wedge y_2)$. Under model (2), which implies condition (\mathcal{SO}) by Lemma 1, we have according to Peng (1999) and (6.2) in Draisma *et al.* (2004), and invoking a Skorohod construction (but keeping the same notation), that

$$\begin{aligned} \frac{S \left(\frac{k}{n}x, \frac{k}{n}y \frac{\omega}{1-\omega} \right)}{m} &= \frac{\mathbb{P} \left(U \leq \frac{k}{n}x, V \leq \frac{k}{n}y \frac{\omega}{1-\omega} \right)}{\mathbb{P} \left(U \leq \frac{k}{n}, V \leq \frac{k}{n} \right)} + \frac{W(x, y \frac{\omega}{1-\omega})}{\sqrt{m}} + o \left(\frac{1}{\sqrt{m}} \right) \\ &= \frac{\mathbb{P} \left(1 - F_X(X) < \frac{k}{n}x, 1 - F_Y(Y) < \frac{k}{n}y \frac{\omega}{1-\omega} \right)}{\mathbb{P} \left(1 - F_X(X) < \frac{k}{n}, 1 - F_Y(Y) < \frac{k}{n} \right)} + \frac{W(x, y \frac{\omega}{1-\omega})}{\sqrt{m}} + o \left(\frac{1}{\sqrt{m}} \right) \\ &= c \left(x, y \frac{\omega}{1-\omega} \right) + q_1 \left(\frac{k}{n} \right) c_1 \left(x, y \frac{\omega}{1-\omega} \right) + \frac{W(x, y \frac{\omega}{1-\omega})}{\sqrt{m}} + o \left(\frac{1}{\sqrt{m}} \right), \end{aligned}$$

a.s. uniformly on $[0, b]^2$. Using the condition $\sqrt{m}q_1(k/n) \rightarrow \lambda$ this can be rewritten as

$$\sup_{0 \leq x, y \leq b} \left| \sqrt{m} \left[\frac{S \left(\frac{k}{n}x, \frac{k}{n}y \frac{\omega}{1-\omega} \right)}{m} - c \left(x, y \frac{\omega}{1-\omega} \right) \right] - \lambda c_1 \left(x, y \frac{\omega}{1-\omega} \right) - W \left(x, y \frac{\omega}{1-\omega} \right) \right| \rightarrow 0 \quad a.s. \quad (10)$$

Now consider

$$\sup_{b^* \leq r < \infty} \left| \sqrt{m} \left(\frac{\bar{F}_n(r)}{q(k/n)} - r^{-\frac{1}{\eta}} c_\omega \right) - \lambda c_1 \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) - W \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) \right|.$$

By some elementary calculations this can be decomposed as

$$\begin{aligned}
& \left| \sqrt{m} \left(\frac{\bar{F}_n(r)}{q(k/n)} - r^{-\frac{1}{\eta}} c_\omega \right) - \lambda c_1 \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) - W \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) \right| \\
& \leq \left| \sqrt{m} \left(\frac{\bar{F}_n(r)}{q(k/n)} - c \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}, \frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right) \right) - \lambda c_1 \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}, \frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right) - W \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}, \frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right) \right| \\
& \quad + \sqrt{m} \left| c \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}, \frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right) - c \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) \right| + \lambda \left| c_1 \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}, \frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right) - c_1 \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) \right| \\
& \quad + \left| W \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}, \frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} \right) - W \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) \right| \\
& =: T_1 + T_2 + T_3 + T_4.
\end{aligned} \tag{11}$$

In order to analyze these terms we will use the following stochastic process property concerning uniform order statistics from Drees and Huang (1998): for any fixed value $T > 0$, with \widetilde{W} denoting an almost surely continuous Gaussian process, one has

$$\sup_{0 \leq x \leq T} \left| \sqrt{k} \left(\frac{n}{k} U_{\lceil kx \rceil, n} - x \right) + \widetilde{W}(x, 0) \right| \rightarrow 0 \text{ a.s. } , \tag{12}$$

$$\sup_{0 \leq y \leq T} \left| \sqrt{k} \left(\frac{n}{k} V_{\lceil ky \rceil, n} - y \right) + \widetilde{W}(0, y) \right| \rightarrow 0 \text{ a.s. } . \tag{13}$$

We now analyze each of the terms of (11) in turn. Concerning T_1 we use (10), (12) and (13) to obtain that $T_1 \rightarrow 0$ a.s. uniformly for $r \in [b^*, \infty)$. For T_2 , by the mean value theorem

$$T_2 = c_x(\tilde{U}_r, \tilde{V}_r) \sqrt{m} \left(\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n} - \frac{1}{r} \right) + c_y(\tilde{U}_r, \tilde{V}_r) \sqrt{m} \left(\frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n} - \frac{1}{r} \frac{\omega}{1-\omega} \right),$$

with c_x and c_y denoting the partial derivatives of c with respect to x and y , respectively, \tilde{U}_r being a random value between $\frac{n}{k} U_{\lceil \frac{k}{r} \rceil, n}$ and $1/r$, and \tilde{V}_r a random value between $\frac{n}{k} V_{\lceil \frac{\omega}{1-\omega} \frac{k}{r} \rceil, n}$ and $\frac{1}{r} \frac{\omega}{1-\omega}$. Using the continuity of c_x and c_y , (12) and (13), and the fact that $m = o(k)$ in case $\eta < 1$ (Draisma *et al.*, 2004), we have that $T_2 \rightarrow 0$ a.s. uniformly. Finally, for T_3 and T_4 we use the continuity of c_1 and W , (12) and (13), to get that $T_3 \rightarrow 0$ a.s. and $T_4 \rightarrow 0$ a.s. uniformly on $r \in [b^*, \infty)$.

Thus we have

$$\sqrt{m} \left(\frac{\bar{F}_n(r)}{q(k/n)} - r^{-\frac{1}{\eta}} c_\omega \right)_{b^* \leq r < \infty} \rightarrow \left(W \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) + \lambda c_1 \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) \right)_{b^* \leq r < \infty}$$

weakly in $D[b^*, \infty)$.

Using the function c_1 for model (2), see also the proof of Lemma 1, we obtain

$$\sqrt{m} \left(\frac{\bar{F}_n(r)}{q(k/n)} - r^{-\frac{1}{\eta}} c_\omega \right)_{b^* \leq r < \infty} \rightarrow \left(W \left(\frac{1}{r}, \frac{1}{r} \frac{\omega}{1-\omega} \right) + \lambda r^{-\frac{1}{\eta}} c_\omega \frac{r^{-\tau} \xi_\omega - 1}{\tau} \right)_{b^* \leq r < \infty}$$

weakly in $D[b^*, \infty)$. Thus, this implies in particular that

$$\sqrt{m} \left(\frac{\overline{F}_n(r^{-\eta})}{q(k/n)} - r c_\omega \right)_{0 < r \leq b'} \longrightarrow \left(W \left(r^\eta, r^\eta \frac{\omega}{1-\omega} \right) + \lambda r c_\omega \frac{r^{\eta\tau} \xi_\omega - 1}{\tau} \right)_{0 < r \leq b'}$$

weakly in $D(0, b']$. By inversion, using Vervaat's lemma (Vervaat, 1972) we have

$$\begin{aligned} \sqrt{m} \left(\left[\overline{F}_n^\leftarrow(tq(k/n)) \right]^{-1/\eta} - \frac{t}{c_\omega} \right)_{0 < t \leq c_\omega b'} &\longrightarrow -\frac{1}{c_\omega} \left(W \left(\frac{t^\eta}{c_\omega^\eta}, \frac{t^\eta}{c_\omega^\eta} \frac{\omega}{1-\omega} \right) + \lambda t \frac{\left(\frac{t}{c_\omega} \right)^{\eta\tau} \xi_\omega - 1}{\tau} \right)_{0 < t \leq c_\omega b'} \\ &=: -\frac{1}{c_\omega} \left(\overline{W} \left(\frac{t}{c_\omega} \right) + \lambda t \frac{\left(\frac{t}{c_\omega} \right)^{\eta\tau} \xi_\omega - 1}{\tau} \right)_{0 < t \leq c_\omega b'} \end{aligned} \quad (14)$$

weakly in $D(0, c_\omega b']$.

Using the fact that $W(x^\eta, x^\eta)$ is a Brownian motion, we can infer that it is also the case for $\overline{W}(x/c_\omega)$. Indeed, it has zero mean and it is such that

$$\begin{aligned} \mathbb{E} \left(\overline{W} \left(\frac{t_1}{c_\omega} \right) \overline{W} \left(\frac{t_2}{c_\omega} \right) \right) &= \mathbb{E} \left(W \left(\frac{t_1^\eta}{c_\omega^\eta}, \frac{t_1^\eta}{c_\omega^\eta} \frac{\omega}{1-\omega} \right) W \left(\frac{t_2^\eta}{c_\omega^\eta}, \frac{t_2^\eta}{c_\omega^\eta} \frac{\omega}{1-\omega} \right) \right) \\ &= c \left(\frac{t_1^\eta}{c_\omega^\eta} \wedge \frac{t_2^\eta}{c_\omega^\eta}, \frac{t_1^\eta}{c_\omega^\eta} \frac{\omega}{1-\omega} \wedge \frac{t_2^\eta}{c_\omega^\eta} \frac{\omega}{1-\omega} \right) \\ &= c \left(\frac{1}{c_\omega^\eta} (t_1 \wedge t_2)^\eta, \frac{\omega}{1-\omega} \frac{1}{c_\omega^\eta} (t_1 \wedge t_2)^\eta \right) \\ &= \frac{t_1 \wedge t_2}{c_\omega} c \left(1, \frac{\omega}{1-\omega} \right) = t_1 \wedge t_2. \end{aligned}$$

Recall now that $\overline{F}_n^\leftarrow(tq(k/n)) = \frac{k}{n} Q_n(t)$. Thus (14) combining with the delta method leads to

$$\left| \sqrt{m} \left(\frac{k}{n} Q_n(t) - \left(\frac{t}{c_\omega} \right)^{-\eta} \right) - \eta t^{-(\eta+1)} c_\omega^\eta \overline{W} \left(\frac{t}{c_\omega} \right) - \lambda \eta \left(\frac{t}{c_\omega} \right)^{-\eta} \frac{\left(\frac{t}{c_\omega} \right)^{\eta\tau} \xi_\omega - 1}{\tau} \right| = o_{\mathbb{P}}(1).$$

Now, following the lines of proof of Lemma 6.2 in Draisma *et al.* (2004), with some adjustments due to our different model, we can deduce that for all $t_0, \varepsilon > 0$:

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

Proof of Theorem 2

From Theorem 1, we can deduce that

$$Q_n(t) = \frac{n}{k} \left(\frac{t}{c_\omega} \right)^{-\eta} \left[1 + \frac{\eta}{t\sqrt{m}} \overline{W} \left(\frac{t}{c_\omega} \right) + \lambda \frac{\eta}{\sqrt{m}} \frac{\left(\frac{t}{c_\omega} \right)^{\eta\tau} \xi_\omega - 1}{\tau} + o_{\mathbb{P}} \left(\frac{t^{-(\frac{1}{2} + \varepsilon)}}{\sqrt{m}} \right) \right]. \quad (15)$$

In order to study the statistics $A_{m,n}^{(i)}(s_i)$, $i = 1, 2, 3$, we use a Taylor expansion of each integrand and thus we need that all the terms appearing in the square parenthesis in (15) tend to 0 uniformly. To this aim, we have to split the integrals into two parts, from 0 to $m^{-\kappa}$ and from $m^{-\kappa}$ to 1, for a suitable $\kappa \in (0, 1/(1+2\varepsilon))$. A similar argument was used in Section 3.6.2 of de Haan and Ferreira (2006).

Assuming $\kappa > 1/2$, we clearly have $\int_0^{m^{-\kappa}} \left(\frac{Q_n(t)}{Q_n(1)}\right)^{s_1} dt = o_{\mathbb{P}}(1/\sqrt{m})$. Also, for $\varepsilon < 0.5$

$$\begin{aligned} \int_{m^{-\kappa}}^1 \left(\frac{Q_n(t)}{Q_n(1)}\right)^{s_1} dt &= \int_{m^{-\kappa}}^1 t^{-\eta s_1} \left[1 + \left(-\frac{s_1 \eta}{\sqrt{m}} \overline{W} \left(\frac{1}{c_\omega} \right) - \lambda \frac{s_1 \eta}{\sqrt{m}} \frac{c_\omega^{-\eta \tau} \xi_\omega - 1}{\tau} \right. \right. \\ &\quad \left. \left. + \frac{s_1 \eta}{t \sqrt{m}} \overline{W} \left(\frac{t}{c_\omega} \right) + \lambda \frac{s_1 \eta}{\sqrt{m}} \frac{\left(\frac{t}{c_\omega}\right)^{\eta \tau} \xi_\omega - 1}{\tau} + o_{\mathbb{P}} \left(\frac{t^{-(\frac{1}{2} + \varepsilon)}}{\sqrt{m}} \right) \right) (1 + o_{\mathbb{P}}(1)) \right] dt \end{aligned}$$

where the $o_{\mathbb{P}}(1)$ -term is uniform. Thus

$$\begin{aligned} \int_{m^{-\kappa}}^1 \left(\frac{Q_n(t)}{Q_n(1)}\right)^{s_1} dt &= \frac{1}{1 - \eta s_1} - \lambda \frac{s_1 \eta^2}{(1 - \eta s_1)(1 - \eta(s_1 - \tau))} \frac{1}{\sqrt{m}} c_\omega^{-\eta \tau} \xi_\omega \\ &\quad + \frac{s_1 \eta}{\sqrt{m}} \int_0^1 t^{-(1+\eta s_1)} \overline{W} \left(\frac{t}{c_\omega} \right) dt - \frac{\eta s_1}{1 - \eta s_1} \frac{1}{\sqrt{m}} \overline{W} \left(\frac{1}{c_\omega} \right) + o_{\mathbb{P}}(m^{-1/2}). \end{aligned}$$

This yields

$$\begin{aligned} \sqrt{m} \left(A_{m,n}^{(1)}(s_1) - \frac{1}{1 - \eta s_1} \right) &= s_1 \eta \int_0^1 t^{-(1+\eta s_1)} \overline{W} \left(\frac{t}{c_\omega} \right) dt - \frac{\eta s_1}{1 - \eta s_1} \overline{W} \left(\frac{1}{c_\omega} \right) \\ &\quad - \lambda \frac{s_1 \eta^2}{(1 - \eta s_1)(1 - \eta(s_1 - \tau))} \xi_\omega c_\omega^{-\eta \tau} + o_{\mathbb{P}}(1). \end{aligned}$$

Similarly, we can derive that

$$\begin{aligned} \sqrt{m} \left(A_{m,n}^{(2)}(s_2) - \frac{\eta}{(1 - \eta s_2)^2} \right) &= -s_2 \eta^2 \int_0^1 t^{-(1+\eta s_2)} \log t \overline{W} \left(\frac{t}{c_\omega} \right) dt \\ &\quad + \eta \int_0^1 t^{-(1+\eta s_2)} \overline{W} \left(\frac{t}{c_\omega} \right) dt - \frac{\eta}{(1 - \eta s_2)^2} \overline{W} \left(\frac{1}{c_\omega} \right) \\ &\quad - \lambda \frac{\eta^2 (1 + \eta \tau - \eta^2 s_2^2)}{(1 - \eta s_2)^2 (1 - \eta(s_2 - \tau))^2} \xi_\omega c_\omega^{-\eta \tau} + o_{\mathbb{P}}(1). \end{aligned}$$

Here we used, setting $m^{-\kappa} = j/m$,

$$\sqrt{m} \int_0^{m^{-\kappa}} \left(\frac{Q_n(t)}{Q_n(1)}\right)^{s_2} \log \frac{Q_n(t)}{Q_n(1)} dt \leq m^{1/2-\kappa} (\log \tilde{Z}_{\omega,n,n} - \log \tilde{Z}_{\omega,n-m,n}),$$

with, from (15),

$$\log \tilde{Z}_{\omega,n-m,n} = O_{\mathbb{P}}(\log n/k) \text{ and } \log \tilde{Z}_{\omega,n,n} = O_{\mathbb{P}}(\log n/k + \log m).$$

Keeping in mind that $k/n = q^{\leftarrow}(m/n)$ with q being regularly varying at zero, the result follows.

By the same arguments,

$$\begin{aligned} \sqrt{m} \left(A_{m,n}^{(3)}(s_3) - \frac{2\eta^2}{(1-\eta s_3)^3} \right) &= s_3 \eta^3 \int_0^1 t^{-(1+\eta s_3)} (\log t)^2 \overline{W} \left(\frac{t}{c_\omega} \right) dt \\ &\quad - 2\eta^2 \int_0^1 t^{-(1+\eta s_3)} \log t \overline{W} \left(\frac{t}{c_\omega} \right) dt - 2 \frac{\eta^2}{(1-\eta s_3)^3} \overline{W} \left(\frac{1}{c_\omega} \right) \\ &\quad - 2\lambda \frac{\eta^2}{\tau} \xi_\omega c_\omega^{-\eta\tau} \left[\frac{1}{(1-\eta s_3)^3} - \frac{1+\eta\tau}{(1-\eta(s_3-\tau))^3} \right] + o_{\mathbb{P}}(1). \end{aligned}$$

The asymptotic variances and covariances can be obtained from straightforward but tedious calculations, and are therefore omitted from the paper.

Acknowledgement

The authors are very grateful to the referee for the constructive comments on the paper. The suggestions have definitely improved the presentation of the material.

References

- [1] Barbe, P., Fougères, A.-L., Genest, C., 2006. On the tail behavior of sums of dependent risks. *ASTIN Bulletin*, 36, 361–373.
- [2] Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C., 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. *Insurance: Mathematics and Economics*, 49, 18–26.
- [5] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes – Theory and Applications*. Wiley.
- [6] Beirlant, J., Joossens, E., Segers, J., 2009. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 139, 2800–2815.
- [7] Beirlant, J., Vandewalle, B., 2002. Some comments on the estimation of a dependence index in bivariate extreme value in statistics. *Statistics and Probability Letters*, 60, 265–278.
- [8] Cebrián, A.C., Denuit, M., Lambert, P., 2003. Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database. *Belgian Actuarial Bulletin*, 3, 33–41.

- [9] Charpentier, A., Juri, A., 2006. Limiting dependence structures for tail events, with applications to credit derivatives. *Journal of Applied Probability*, 43, 563–586.
- [10] Ciuperca, G., Mercadier, C., 2010. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13, 55–87.
- [11] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer.
- [12] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17, 1833–1855.
- [13] Deme, E., Gardes, L., Girard, S., 2013. On the estimation of the second order parameter for heavy-tailed distributions. *REVSTAT - Statistical Journal*, 11, 277–299.
- [14] Dierckx, G., Goegebeur, Y., Guillou, A., 2013. An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *Journal of Multivariate Analysis*, 121, 70–86.
- [15] Draisma, G., Drees, H., Ferreira, A., de Haan, L., 2004. Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10, 251–280.
- [16] Drees, H., Huang, X., 1998. Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64, 25–47.
- [17] Embrechts, P., Lindskog, F., McNeil, A., 2003. Modelling dependence with copulas and applications to risk management. *Handbook of heavy tailed distributions in finance*, 8, 329–384.
- [18] Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, 27, 760–781.
- [19] Fraga Alves, M.I., Gomes, M.I., de Haan, L., 2003. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60, 193–213.
- [20] Frees, E.W., 2010. *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press.
- [21] Frees, E.W., Valdez, E.A., 1998. Understanding relationships using copulas. *North American Actuarial Journal*, 2, 1–25.
- [22] Frees, E.W., Young, V.R., Luo, Y., 2001. Case studies using panel data models. *North American Actuarial Journal*, 5, 24–42.
- [23] Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, 140, 2632–2652.
- [24] Goegebeur, Y., Guillou, A., 2013. Asymptotically unbiased estimation of the coefficient of tail dependence. *Scandinavian Journal of Statistics*, 40, 174–189.

- [25] Gomes, M.I., de Haan, L., Rodrigues, L.H., 2008. Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society Series B*, 70, 31–52.
- [26] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.
- [27] Joe, H., 2011. Tail dependence in vine copulae. In *Dependence Modeling*, World Sci. Publ., Hackensack, NJ, Eds Kurowicka, D., Joe, H., 165–187.
- [28] Klugman, S.A., 1992. *Bayesian Statistics in Actuarial Science*. With emphasis on credibility. Huebner International Series on Risk, Insurance, and Economic Security. Kluwer Academic Publishers, Boston, MA.
- [29] Klugman, S.A., Parsa, R., 1999. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, 24, 139–148.
- [30] Ledford, A.W., Tawn, J.A., 1997. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society Series B*, 59, 475–499.
- [31] Peng, L., 1999. Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters*, 43, 399–409.
- [32] Pickands, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.
- [33] Poon, S.H., Rockinger, M., Tawn, J., 2004. Extreme-value dependence in financial markets: diagnostics, models, and financial implications. *Review of Financial Studies*, 17, 581–610.
- [34] Vervaat, W., 1972. Functional limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 23, 245–253.