CrossMark

Robust and Bias-Corrected Estimation of the Probability of Extreme Failure Sets

Christophe Dutang
Université du Maine, Le Mans, France

Yuri Goegebeur
University of Southern Denmark, Odense M, Denmark

Armelle Guillou
Université de Strasbourg et CNRS, Strasbourg Cedex, France

Abstract

In multivariate extreme value statistics, the estimation of probabilities of extreme failure sets is an important problem, with practical relevance for applications in several scientific disciplines. Some estimators have been introduced in the literature, though so far the typical bias issues that arise in application of extreme value methods and the non-robustness of such methods with respect to outliers were not addressed. We introduce a bias-corrected and robust estimator for small tail probabilities. The estimator is obtained from a second order model that is fitted to properly transformed bivariate observations by means of the minimum density power divergence technique. The asymptotic properties are derived under some mild regularity conditions and the finite sample performance is evaluated through an extensive simulation study. We illustrate the practical applicability of the method on a dataset from the actuarial context.

AMS (2000) *Subject Classification.* 62G32, 62H12, 62G20.
Keywords and phrases. Failure set, Bias-correction, Tail dependence, robustness, Tail quantile process.

1 Introduction

Multivariate extreme value statistics deals with the estimation of the joint tail of a distribution function based on a random sample. Within this area, the problem of estimating the probability of extreme failure sets, i.e. regions of the sample space where one or several random variables are large, is of practical relevance for applications in many scientific disciplines. For instance, financial institutions are often interested in assessing the risk of simultaneous large negative returns on several stocks or other assets

(Charpentier and Juri, 2006). In the design of coastal defence structures, engineers might want to estimate the probability that a seawall fails due to a combination of high waves and high still water levels (de Haan and de Ronde, 1998). The estimation of the probability of extreme failure sets is the topic of the present paper.

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 analysed 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.

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

and the function ς 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. Moreover, since estimators for a bivariate tail probability are typically based on such estimators for η, one can expect that they will also be affected by bias and outliers. These specific 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 model.

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*

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

$$(1)$$

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

Recently, several papers have addressed bias-corrected estimation of the coefficient of tail dependence η, but not too much attention has been paid to the estimation of the probability of extreme failure sets. Concerning η, we refer to Beirlant et al. (2011) and Goegebeur and Guillou (2013). Also, Dutang et al. (2014) introduced a robust and bias-corrected estimator for the coefficient of tail dependence, based on a submodel of ($\mathcal{SO}$) that is fitted to the data by means of the minimum density power divergence technique. The estimators of probabilities of extreme failure sets already proposed in the literature are quite simple in nature, but their asymptotics is very difficult to handle and requires very technical conditions. This is for instance the case of the estimator proposed by Draisma et al. (2004) where only the convergence in probability is established. Peng (1999) also proposed an estimator but without showing convergence nor asymptotic normality. Moreover, bias-correction and robustness were not considered so far for estimators of probabilities of extreme regions. This constitutes the topic of the present paper.

The remainder of this paper is organised as follows. In the next section, we introduce the model and the estimator for probabilities of extreme failure sets. The asymptotic results of the proposed estimator are investigated in Section 3, and the finite sample performance is illustrated through an extensive simulation study in Section 4. Section 5 contains an application of the proposed methodology on a dataset from the actuarial context. The proofs of the main results can be found in the Appendix.

2 Model Assumptions and Estimators

The robust and bias-corrected estimator for probabilities of extreme failure sets will be obtained from a submodel of the second order condition ($\mathcal{SO}$), that will be fitted to properly transformed bivariate observations by means of the minimum density power divergence (MDPD) technique.

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 arbitrary density functions f and h is given by

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

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 minimising an empirical version of Δ_α based on a random sample $Z_1, \ldots, Z_n$ 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. The criterion has been applied in the univariate extreme value context of heavy tailed distributions with the objective to obtain robust estimators of tail parameters: Kim and Lee (2008) constructed a robust version of the Hill estimator based on MDPD estimation, whereas Dierckx et al. (2013) fitted a second order model by this method in order to achieve a tail index estimator that is both robust and bias-corrected. Recently, the MDPD criterion was used by Dierckx et al. (2014) in the framework of conditional heavy tailed models and by Dutang et al. (2014) in the multivariate framework.

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

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

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

$$\lim_{t\downarrow 0} \frac{|\delta|(tx, ty)}{|\delta|(t, t)} = \xi(x, y), \tag{3}$$

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).

Lemma 1. *Model (2) satisfies assumption ($\mathcal{SO}$) with*

$$c(x, y) = x^{d_1} y^{d_2} \frac{g(x, y)}{g(1, 1)}, \tag{4}$$

$$q_1(t) = \frac{\tau}{\eta} \delta(t, t), \tag{5}$$

$$c_1(x, y) = c(x, y) \frac{\xi(x, y) - 1}{\tau}. \tag{6}$$

Many commonly used joint distribution functions satisfy model (2), as for instance the Farlie Gumbel Morgenstern or the Frank distributions, see Section 4.

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

$$\mathbb{P}(X > x, Y > y) = x^{-d_1} y^{-d_2} g^*(x, y) \left(1 + \frac{1}{\eta} \delta^*(x, y)\right) \tag{7}$$

where $g^*(x, y) := g(1/x, 1/y)$ and $\delta^*(x, y) := \delta(1/x, 1/y)$. Note that the choice of unit Pareto margins is not only for convenience in order to use some results from Dutang et al. (2014) already established under this assumption, but also, because as observed by Drees (1998a, b) the classical estimators of extreme value index have larger bias for Fréchet distributions than for Pareto distributions. Model (7) is a slight generalisation 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 and it has been introduced in order to estimate probabilities in

joint tail regions. This ray parametrization was also used in Ramos and Ledford (2009). This parameter is of little practical relevance for estimation of the coefficient of tail dependence, but it will play an important role for failure set probability estimation. 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

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

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, using the common reparametrization $\rho = -\tau\eta$, for u large

$$\frac{\mathbb{P}(Z_\omega > uz)}{\mathbb{P}(Z_\omega > u)} \approx \overline{H}(z; \eta, \delta_\omega(u), \rho) := \begin{cases} 1, & z \le 1, \\ [z(1 + \delta_\omega(u) - \delta_\omega(u)z^{\rho/\eta})]^{-1/\eta}, & z > 1, \end{cases} \quad (8)$$

where H has the density function

$$h(z; \eta, \delta_\omega(u), \rho) = \begin{cases} 0 & z \le 1, \\ \frac{1}{\eta}\, z^{-1/\eta-1}[1 + \delta_\omega(u)(1 - z^{\rho/\eta})]^{-1/\eta-1}\left[1 + \delta_\omega(u)\left(1 - \left(1 + \frac{\rho}{\eta}\right)z^{\rho/\eta}\right)\right], & z > 1, \end{cases}$$

($\eta \in (0,1)$, $\rho < 0$, and $\delta_\omega(u) > \max\{-1, \eta/\rho\}$). Using this property, one can estimate η by fitting h to the relative excesses over some large threshold u. Approximation (8) means that the difference between the two sides is uniformly of order $o(|\delta_\omega(u)|)$ (see Proposition 2.3 in Beirlant et al., 2009) and thus this new extended Pareto model can be fitted to a larger portion of the data. This is in line with other approaches to bias-corrected estimation in the extreme value context. Indeed, in general, by taking the second order structure of extreme value models explicitly into account in the estimation stage, one obtains estimators with better bias properties than estimators based on first order models.

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

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

with R_i^X and R_i^Y denoting the rank of X_i and Y_i, $i = 1, \ldots, n$, 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 $Z_j := \widetilde{Z}_{\omega,n-m+j,n}/\widetilde{Z}_{\omega,n-m,n}$, $j = 1, \ldots, m$, where $1 \le m \le n - 1$, and $\widetilde{Z}_{\omega,1,n} \le \ldots \le \widetilde{Z}_{\omega,n,n}$ are the order statistics of $\widetilde{Z}_{\omega,1}, \ldots, \widetilde{Z}_{\omega,n}$, using the minimum density power divergence criterion. Remember that the parameter δ_ω reflects in fact the function $\delta_\omega(u)$, where $\delta_\omega(u) \to 0$ as $u \to \infty$, but we do not make this dependence on the threshold explicit in the notation. The MDPD estimator for η and δ_ω satisfies the estimating equations

$$0 = \int_1^\infty h^\alpha(z; \eta, \delta_\omega, \rho) \frac{\partial h(z; \eta, \delta_\omega, \rho)}{\partial \eta} dz - \frac{1}{m} \sum_{j=1}^m h^{\alpha-1}(Z_j; \eta, \delta_\omega, \rho) \frac{\partial h(Z_j; \eta, \delta_\omega, \rho)}{\partial \eta}, \quad (9)$$

$$0 = \int_1^\infty h^\alpha(z; \eta, \delta_\omega, \rho) \frac{\partial h(z; \eta, \delta_\omega, \rho)}{\partial \delta_\omega} dz - \frac{1}{m} \sum_{j=1}^m h^{\alpha-1}(Z_j; \eta, \delta_\omega, \rho) \frac{\partial h(Z_j; \eta, \delta_\omega, \rho)}{\partial \delta_\omega}. \quad (10)$$

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 the correct value, or fixed at some arbitrary value, usually the canonical choice $\rho = -1$, and thus possibly mis-specified. Several estimators for ρ were proposed in the univariate context, and we refer here to Fraga Alves et al. (2003), Ciuperca and Mercadier (2010) and Goegebeur et al. (2010) for examples of estimators with a good practical performance. In the bivariate context the estimation of ρ (or τ) is a challenging problem. Goegebeur and Guillou (2013) introduced a consistent estimator for τ, based on the stochastic process results of Draisma et al. (2004) for the empirical distribution function. However, this estimator does not seem to perform well in practice. Therefore, these authors did not use it in their simulation experiment, and also resorted to fixing the second order parameter at some value. We will address the estimation of second order parameters in the bivariate extreme value context in future research.

Our goal in this paper is to estimate the tail probability $p_n := \mathbb{P}(X > z_n, Y > y_n)$, with X and Y being unit Pareto random variables, and where $z_n \to \infty$ as $n \to \infty$, and $y_n = \widetilde{\omega} z_n$ for some $\widetilde{\omega} > 0$, i.e. we estimate a tail probability along a ray. We have thus that $p_n := \overline{F}_{Z_\omega}(z_n)$, where $\omega = (1 + \widetilde{\omega})^{-1}$. To this aim, let m be an intermediate sequence, i.e. $m \to \infty$ as $n \to \infty$ with $m = o(n)$. Assume that p_n satisfies

$$\frac{np_n}{m} \longrightarrow \beta \in [0, 1), \qquad n \to \infty. \quad (11)$$

Using the estimators $\widehat{\eta}_n$ and $\widehat{\delta}_{\omega,n}$ satisfying the estimating equations (9)–(10), we can construct the following estimator for p_n based on (8), setting $u = \widetilde{Z}_{\omega,n-m,n}$ and replacing $\overline{F}_{Z_\omega}(u)$ by the empirical proportion m/n:

$$\widehat{p}_n := \frac{m}{n} \, \overline{H}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}; \widehat{\eta}_n, \widehat{\delta}_{\omega,n}, \widetilde{\rho}\right) \tag{12}$$

where $\widetilde{\rho}$ is either the true value of ρ or a mis-specified one.

Note that, at first sight, once the minima Z_ω are being considered, our problem of estimation might appear a standard problem of tail estimation in the classical univariate framework. However, this is not the case: indeed, the $\widetilde{Z}_{\omega,i}, i = 1, \ldots, n$, are not independent because of the rank transformation, and hence the theoretical study of their order statistics is very complicated, involving convergence results for stochastic processes. All subsequent results are proved in Section 3 using the stochastic process representation, which requires arguments that are different from the univariate independent and identically distributed (i.i.d.) case.

3   Asymptotic Results

Consider the random variables $\widetilde{Z}_{\omega,1}, \ldots, \widetilde{Z}_{\omega,n}$, with order statistics $\widetilde{Z}_{\omega,1,n} \leq \ldots \leq \widetilde{Z}_{\omega,n,n}$. In order to obtain the limiting distribution of $\widehat{p}_n$, we need some preliminary results given in Theorems 1 and 2. These theorems are based on Theorem 1 in Dutang et al. (2014) which is for completeness recalled in the Appendix. In the following, we denote the tail quantile process as $Q_n(t) := \widetilde{Z}_{\omega,n-[mt],n}$, $0 < t < n/m$. 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 $k \to \infty$.

The estimating equations (9) and (10) depend only on the data through the following statistics

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

with $s_1, s_2, s_3 \leq 0$ whose joint asymptotic normality follows from Theorem 5 (cf. Appendix). However, in order to derive the limiting distribution of $\widehat{p}_n$ defined in (12), we need to add an additional component, $Q_n(1) = \widetilde{Z}_{\omega,n-m,n}$, correctly rescaled. The aim of Theorem 1 is thus to establish the joint

limiting distribution of the main statistics. Here and in the sequel the arrow $\rightsquigarrow$ denotes convergence in distribution.

THEOREM 1. *Let* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *be independent copies of the random vector* (X, Y) *which has a joint distribution satisfying (2) such that the function* c *given in (4) has continuous first order partial derivatives. For* $m, k \to \infty$ *as* $n \to \infty$ *such that* $\sqrt{m} q_1(k/n) \to \lambda \in \mathbb{R}$ *we have*

$$
\sqrt{m}
\begin{bmatrix}
\frac{k}{n} Q_n(1) - c_\omega^\eta \\
A_{m,n}^{(1)}(s_1) - \frac{1}{1 - \eta s_1} \\
A_{m,n}^{(1)}(s_1') - \frac{1}{1 - \eta s_1'} \\
A_{m,n}^{(2)}(s_2) - \frac{\eta}{(1 - \eta s_2)^2} \\
A_{m,n}^{(3)}(s_3) - \frac{2\eta^2}{(1 - \eta s_3)^3}
\end{bmatrix}
\rightsquigarrow N_5(\lambda \boldsymbol{\mu}, \boldsymbol{\Sigma})
$$

where

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

and where $\boldsymbol{\Sigma}$ *has elements given by*

$$
\sigma_{11} := \eta^2 c_\omega^{2\eta}
$$

$$
\sigma_{22} := \frac{\eta^2 s_1^2}{(1 - s_1 \eta)^2 (1 - 2 s_1 \eta)}
$$

$$
\sigma_{33} := \frac{\eta^2 s_1'^2}{(1 - s_1' \eta)^2 (1 - 2 s_1' \eta)}
$$

$$
\sigma_{44} := \eta^2 \left\{ \frac{2}{(1 - 2 s_2 \eta)^3} - \frac{1}{(1 - s_2 \eta)^4} \right\}
$$

$$
\sigma_{55} := \eta^4 \left\{ \frac{4!}{(1 - 2 s_3 \eta)^5} - \frac{4}{(1 - s_3 \eta)^6} \right\}
$$

$$
\sigma_{12} = \sigma_{13} = \sigma_{14} = \sigma_{15} := 0
$$

$$
\sigma_{23} := \frac{\eta^2 s_1 s_1'}{(1 - s_1 \eta)(1 - s_1' \eta)[1 - (s_1 + s_1')\eta]}
$$

$$
\sigma_{24} := \eta \left\{ \frac{1}{[1 - (s_1 + s_2)\eta]^2} - \frac{1}{(1 - s_1 \eta)(1 - s_2 \eta)^2} \right\}
$$

$$\sigma_{25} := 2\eta^2 \left\{ \frac{1}{[1-(s_1+s_3)\eta]^3} - \frac{1}{(1-s_1\eta)(1-s_3\eta)^3} \right\}$$

$$\sigma_{34} := \eta \left\{ \frac{1}{[1-(s_1'+s_2)\eta]^2} - \frac{1}{(1-s_1'\eta)(1-s_2\eta)^2} \right\}$$

$$\sigma_{35} := 2\eta^2 \left\{ \frac{1}{[1-(s_1'+s_3)\eta]^3} - \frac{1}{(1-s_1'\eta)(1-s_3\eta)^3} \right\}$$

$$\sigma_{45} := 2\eta^3 \left\{ \frac{3}{[1-(s_2+s_3)\eta]^4} - \frac{1}{(1-s_2\eta)^2(1-s_3\eta)^3} \right\}.$$

The proof of this Theorem 1 is an adaptation of the proof of Theorem 2 in Dutang et al. (2014) combined with Theorem 5. Note that the asymptotic independence between the random threshold (first component of the $(5 \times 1)-$vector in Theorem 1) and the other components has already been observed in the univariate framework, see for instance Lemma 5.3 in Dekkers et al. (1989). Indeed in the univariate case this property can be explained by the fact that for $U_1, \ldots, U_n$ i.i.d. $U(0,1)$ random variables, with order statistics $U_{1,n} \leq \ldots \leq U_{n,n}$, one has that $V_{j,k} := U_{j,n}/U_{k+1,n}$, $j = 1, \ldots, k$, behave jointly as the order statistics of a random sample of size k from the $U(0,1)$ distribution, independently from $U_{k+1,n}$. Note however that the formal establishment of the asymptotic independence property in our situation is different from the classical univariate i.i.d. framework since we work with the order statistics of the (non-independent) variables $\widetilde{Z}_{\omega,i}$, $i = 1, \ldots, n$. In our situation, the independence can be intuitively explained by the fact that the asymptotic distribution of $(A_{m,n}^{(1)}(s_1), A_{m,n}^{(1)}(s_1'), A_{m,n}^{(2)}(s_2), A_{m,n}^{(3)}(s_3))$ is totally determined by that of $\left(\frac{\widetilde{Z}_{\omega,n,n}}{\widetilde{Z}_{\omega,n-m,n}}, \ldots, \frac{\widetilde{Z}_{\omega,n-m+1,n}}{\widetilde{Z}_{\omega,n-m,n}} \right)$ whereas the asymptotic distribution of the first component in our Theorem 1, $Q_n(1)$, is totally determined by that of $\widetilde{Z}_{\omega,n-m,n}$.

From now on, we will denote the true value of η and ρ by η_0 and ρ_0, respectively. Let $\delta_{\omega,n} := \delta_\omega(\widetilde{Z}_{\omega,n-m,n})$. In the next theorem, we establish the joint limiting distribution of $Q_n(1)$, $\widehat{\eta}_n$ and $\widehat{\delta}_{\omega,n}$, when properly normalised. The parameter ρ in the estimating equations (9) and (10) is denoted in the sequel as $\widetilde{\rho}$ and is either the true value ρ_0 or a mis-specified one.

THEOREM 2. *Under the conditions of Theorem 1 we have that*

$$\sqrt{m} \begin{bmatrix} \frac{k}{n}Q_n(1) - c_\omega^{\eta_0} \\ \widehat{\eta}_n - \eta_0 \\ \widehat{\delta}_{\omega,n} - \delta_{\omega,n} \end{bmatrix} \rightsquigarrow \begin{bmatrix} Z \\ \Gamma \\ \Delta \end{bmatrix} \tag{13}$$

where

$$\begin{bmatrix} Z \\ \Gamma \\ \Delta \end{bmatrix} \sim N_3 \left(\lambda [\mathbb{C}^{-1}(\widetilde{\rho})\mathbb{B}(\widetilde{\rho})\widetilde{\boldsymbol{\mu}} - \boldsymbol{\delta}_*], \ \mathbb{C}^{-1}(\widetilde{\rho})\mathbb{B}(\widetilde{\rho})\mathbb{D}(\widetilde{\rho})\mathbb{B}'(\widetilde{\rho})\mathbb{C}^{-1}(\widetilde{\rho}) \right), \quad (14)$$

with

$$\widetilde{\boldsymbol{\mu}} \ := \ \begin{bmatrix} -\eta_0^2 c_\omega^{\eta_0} \frac{c_\omega^{\rho_0} \xi_\omega - 1}{\rho_0} \\ \xi_\omega c_\omega^{\rho_0} \frac{\alpha\eta_0(1+\eta_0)}{(1+\alpha(1+\eta_0))(1-\rho_0+\alpha(1+\eta_0))} \\ \xi_\omega c_\omega^{\rho_0} \frac{\eta_0[\alpha(1+\eta_0)-\widetilde{\rho}]}{(1-\widetilde{\rho}+\alpha(1+\eta_0))(1-\rho_0-\widetilde{\rho}+\alpha(1+\eta_0))} \\ -\xi_\omega c_\omega^{\rho_0} \frac{\eta_0^2[1-\rho_0-\alpha^2(1+\eta_0)^2]}{(1+\alpha(1+\eta_0))^2(1-\rho_0+\alpha(1+\eta_0))^2} \end{bmatrix},$$

$$\boldsymbol{\delta}_*' \ := \ [0, 0, -\eta_0^2 \xi_\omega c_\omega^{\rho_0}/\rho_0],$$

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

the symmetric matrix $\mathbb{C}(\widetilde{\rho})$ *has elements*

$$c_{11}(\widetilde{\rho}) \ := \ 1$$
$$c_{12}(\widetilde{\rho}) \ := \ 0$$
$$c_{13}(\widetilde{\rho}) \ := \ 0$$
$$c_{22}(\widetilde{\rho}) \ := \ \eta_0^{-\alpha-2} \frac{1+\alpha^2(1+\eta_0)^2}{[1+\alpha(1+\eta_0)]^3}$$
$$c_{23}(\widetilde{\rho}) \ := \ \eta_0^{-\alpha-2} \frac{\widetilde{\rho}(1-\widetilde{\rho})[1+\alpha(1+\eta_0)+\alpha^2(1+\eta_0)^2]+\alpha^3\widetilde{\rho}(1+\eta_0)^3}{[1+\alpha(1+\eta_0)]^2[1-\widetilde{\rho}+\alpha(1+\eta_0)]^2}$$
$$c_{33}(\widetilde{\rho}) \ := \ \eta_0^{-\alpha-2} \frac{(1-\widetilde{\rho})\widetilde{\rho}^2+\alpha\widetilde{\rho}^2(1+\eta_0)[\alpha(1+\eta_0)-\widetilde{\rho}]}{[1+\alpha(1+\eta_0)][1-\widetilde{\rho}+\alpha(1+\eta_0)][1-2\widetilde{\rho}+\alpha(1+\eta_0)]}$$

and the symmetric matrix $\mathbb{D}(\widetilde{\rho})$ *has elements*

$$d_{11}(\widetilde{\rho}) \ := \ \eta_0^2 c_\omega^{2\eta_0}$$
$$d_{12}(\widetilde{\rho}) \ = \ d_{13}(\widetilde{\rho}) = d_{14}(\widetilde{\rho}) := 0$$
$$d_{22}(\widetilde{\rho}) \ := \ \frac{\alpha^2(1+\eta_0)^2}{[1+\alpha(1+\eta_0)]^2[1+2\alpha(1+\eta_0)]}$$
$$d_{23}(\widetilde{\rho}) \ := \ \frac{\alpha(1+\eta_0)[\alpha(1+\eta_0)-\widetilde{\rho}]}{[1+\alpha(1+\eta_0)][1-\widetilde{\rho}+\alpha(1+\eta_0)][1-\widetilde{\rho}+2\alpha(1+\eta_0)]}$$
$$d_{33}(\widetilde{\rho}) \ := \ \frac{[\alpha(1+\eta_0)-\widetilde{\rho}]^2}{[1-\widetilde{\rho}+\alpha(1+\eta_0)]^2[1-2\widetilde{\rho}+2\alpha(1+\eta_0)]}$$
$$d_{24}(\widetilde{\rho}) \ := \ \eta_0 \left(\frac{1}{[1+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^3} \right)$$

$$d_{34}(\widetilde{\rho}) \quad := \quad \eta_0 \left(\frac{1}{[1 - \widetilde{\rho} + 2\alpha(1 + \eta_0)]^2} - \frac{1}{[1 + \alpha(1 + \eta_0)]^2 [1 - \widetilde{\rho} + \alpha(1 + \eta_0)]} \right)$$

$$d_{44}(\widetilde{\rho}) \quad := \quad \eta_0^2 \left(\frac{2}{[1 + 2\alpha(1 + \eta_0)]^3} - \frac{1}{[1 + \alpha(1 + \eta_0)]^4} \right).$$

Note that if $\widetilde{\rho} = \rho_0$ in (14), the mean of the limiting normal distribution is $(-\lambda \eta_0^2 c_\omega^{\eta_0} \frac{c_\omega^{\rho_0} \xi_\omega - 1}{\rho_0}, 0, 0)'$. Thus, 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. Obviously, if $\widetilde{\rho}$ is not the true value, this property of unbiasedness is lost. Nevertheless, despite the loss of asymptotic unbiasedness when fixing $\widetilde{\rho}$ at a mis-specified value, like e.g. -1, one has in the univariate context often observed that the resulting second order estimators still have some bias-correcting effect and typically outperform estimators that are not corrected for bias, both in terms of bias and mean squared error; see e.g. Gomes and Martins (2004), Feuerverger and Hall (1999). A similar observation was made by Goegebeur and Guillou (2013) in the context of bias-corrected estimation of η. Keep in mind that in Theorem 2, $Z = \eta_0 c_\omega^{\eta_0} \overline{W}(1/c_\omega) - \lambda \eta_0^2 c_\omega^{\eta_0} (c_\omega^{\rho_0} \xi_\omega - 1)/\rho_0$, where $\overline{W}(1/c_\omega)$ comes from Theorem 5, and satisfies $\overline{W}(1/c_\omega) \sim N(0, 1)$.

Now we derive the asymptotic normality of the estimator $\widehat{p}_n$. In first instance we consider the situation where $np_n/m \to \beta \in (0, 1)$.

THEOREM 3. *Under the conditions of Theorem 1 and (11) with $\beta \in (0, 1)$, we have*

$$\sqrt{m} \left(\frac{\widehat{p}_n}{p_n} - 1 \right) \rightsquigarrow \overline{W} \left(\frac{1}{c_\omega} \right) - \frac{\ln \beta}{\eta_0} \Gamma - \frac{1 - \beta^{-\widetilde{\rho}}}{\eta_0} \Delta + \lambda \frac{\eta_0 \xi_\omega c_\omega^{\rho_0}}{\rho_0} (\beta^{-\rho_0} - \beta^{-\widetilde{\rho}}).$$

The result of Theorem 3 indicates that the tail probability estimator is asymptotically unbiased if one uses the correct value for ρ. We now consider the case where $np_n/m \to 0$, corresponding to a more extreme failure set. Let $d_n := m/(np_n)$.

THEOREM 4. *Under the conditions of Theorem 1, and assuming additionally $np_n = o(m)$ and $\ln np_n = o(\sqrt{m})$ we have that*

$$\frac{\eta_0 \sqrt{m}}{\ln d_n} \left(\frac{\widehat{p}_n}{p_n} - 1 \right) \rightsquigarrow \Gamma.$$

4   Simulation Study

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

Note that model (2), under which our estimator has been constructed, is in fact a condition on the copula function. Indeed

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

As already mentioned, this condition is satisfied by many commonly used bivariate distribution functions. In particular, we illustrate the behaviour of our estimator $\widehat{p}_{n,\alpha}$ on the following two well-known copulas:

- the Farlie Gumbel Morgenstern (FGM) copula function given by

$$C(x, y) = xy\left[1 + \zeta(1 - x)(1 - y)\right], \qquad (x, y) \in [0, 1]^2, \quad \zeta \in [-1, 1].$$

  We take two different values of ζ: $\zeta = 1$ which implies $\eta = 1/2$, and $\zeta = -1$ which leads to $\eta = 1/3$. For the FGM copula $\tau = 1$.

- the Frank copula function given by

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

In that case $\eta = 1/2$ and $\tau = 1$.

In our simulation, the value of z is chosen such that the (theoretical) survival probability $p_{z,\widetilde{\omega}} = \mathbb{P}(X > z, Y > \widetilde{\omega}z)$ is equal to 2 %, 0.1 % and 0.01 %, see Table 1.

The data are generated as follows. First, we simulate n pairs $(X_i, Y_i), i = 1, ..., n$, independently from a FGM copula with unit Fréchet marginals. Similarly, we simulate n pairs from a Frank copula. Each time, we transform the margins into (approximately) unit Pareto using the empirical distribution functions. This gives us $\widetilde{Z}_{\omega,i}, i = 1, ..., n$. Finally, we minimise the empirical density power divergence $\widehat{\Delta}_\alpha$.

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

Table 1: Value of z such that $p_{z,\widetilde{\omega}} = \mathbb{P}(X > z, Y > \widetilde{\omega}z)$ with $\omega = (1 + \widetilde{\omega})^{-1}$.

	FGM		Frank	FGM	
	$\zeta = 1$	$\zeta = -1$		$\zeta = 1$	
$\omega = 0.45$	10.486				
$\omega = 0.5$	9.488	4.461	9.828	44.219	140.92
$\omega = 0.55$	8.579				
$p_{z,\widetilde{\omega}}$	2 %			0.1 %	0.01 %

- we simulate independently $\breve{X}_j$ and $\breve{Y}_j, j = 1, ..., n_0 := \lfloor n\varepsilon \rfloor$, from a unit Fréchet distribution;

- all our above methodology is applied to our new $n + n_0$ pairs (X_1, Y_1), ..., (X_n, Y_n), $(\breve{X}_1 + X_{n,n}, \breve{Y}_1 + Y_{n,n})$, ..., $(\breve{X}_{n_0} + X_{n,n}, \breve{Y}_{n_0} + Y_{n,n})$, 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 $\varepsilon = 0, 2, 5$ %, while $n = 100$, thus $n_0 = 0, 2, 5$. The procedure is repeated $2^{10} = 1024$ times since we used parallelization on a $8-$core computer.

The left panel of Figure 1 represents the mean of our estimator $\widehat{p}_{n,\alpha}$ of $p_{z,1} = 2$ % and the right panel the mean squared error of the ratio $\widehat{p}_{n,\alpha}/p_{z,1}$ as a function of m for the FGM copula with $\zeta = 1$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line), 1 (dash-dotted line). Three levels of contamination have been considered: $\varepsilon = 0$ (first row), 2 % (second row), 5 % (third row). The parameter ρ is mis-specified to the value -1. Analogously, Figures 2 and 3 are built for the FGM copula with $\zeta = -1$ and the Frank copula, respectively, both in case $p_{z,1} = 2$ %. Note that in these figures, $\widetilde{\omega}$ is set to the value 1, i.e. ω to $1/2$. Based on our simulations, we can draw the following conclusions:

- on uncontaminated simulations, the choice $\alpha = 0$ is the best one, whatever the distribution is. This is expected, since in this no contamination framework, a procedure of robustification is not necessary. However, if we keep α small, we preserve the good behaviour of the estimator, in terms of bias and MSE;

- when contamination occurs, we clearly observe the superiority of our estimator $\widehat{p}_{n,\alpha}$ for $\alpha > 0$. In terms of bias as well as MSE, $\alpha = 0.5$ seems to be nearly always the best choice;

- as expected, increasing the percentage of contamination deteriorates the estimation procedure;

- the superiority of our estimator $\widehat{p}_{n,\alpha}$ on $\widehat{p}_{n,0}$ is stronger in case FGM with $\zeta = -1$ compared to FGM with $\zeta = 1$ in terms of bias and MSE (see the different scale on the $y-$axis in Figures 1 and 2). This can be explained by the fact that one disturbs the sample by outliers that have a different dependence structure in Figure 2 where $\eta = 1/3$, whereas for Figure 1 where $\eta = 1/2$, the data are only disturbed by outliers with a similar dependence structure as for the uncontaminated data;
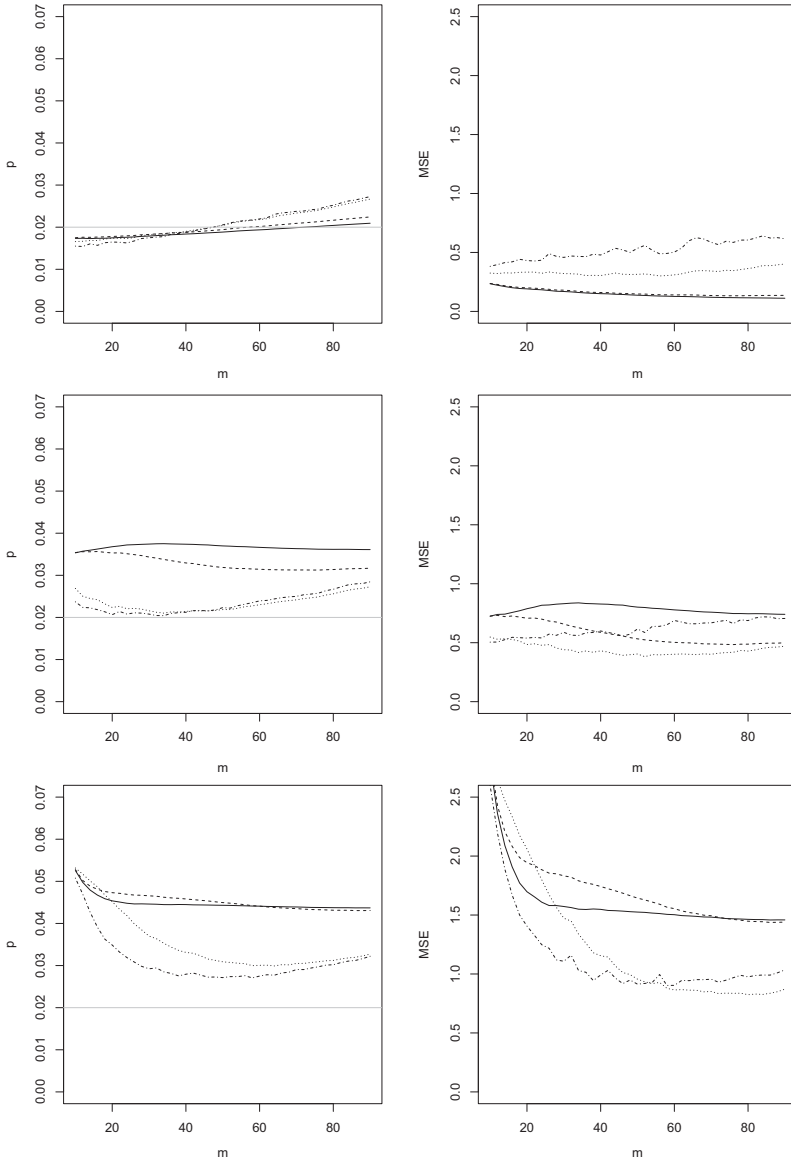
Figure 1: Farlie Gumbel Morgenstern copula with $\zeta = 1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 100, p_{z,1} = \frac{2}{n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 2\%$, and $\varepsilon = 5\%$. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,1}$.
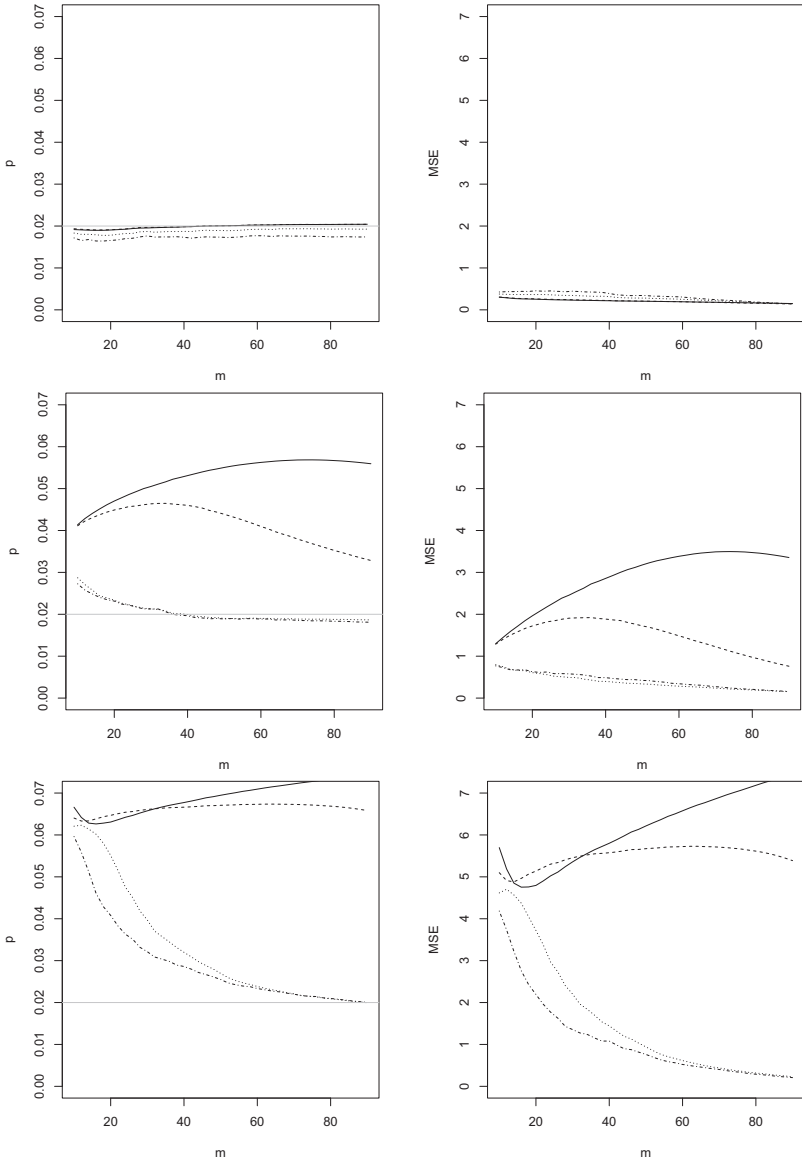
Figure 2: Farlie Gumbel Morgenstern copula with $\zeta = -1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 100, p_{z,1} = \frac{2}{n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 2\,\%$, and $\varepsilon = 5\,\%$. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,1}$.
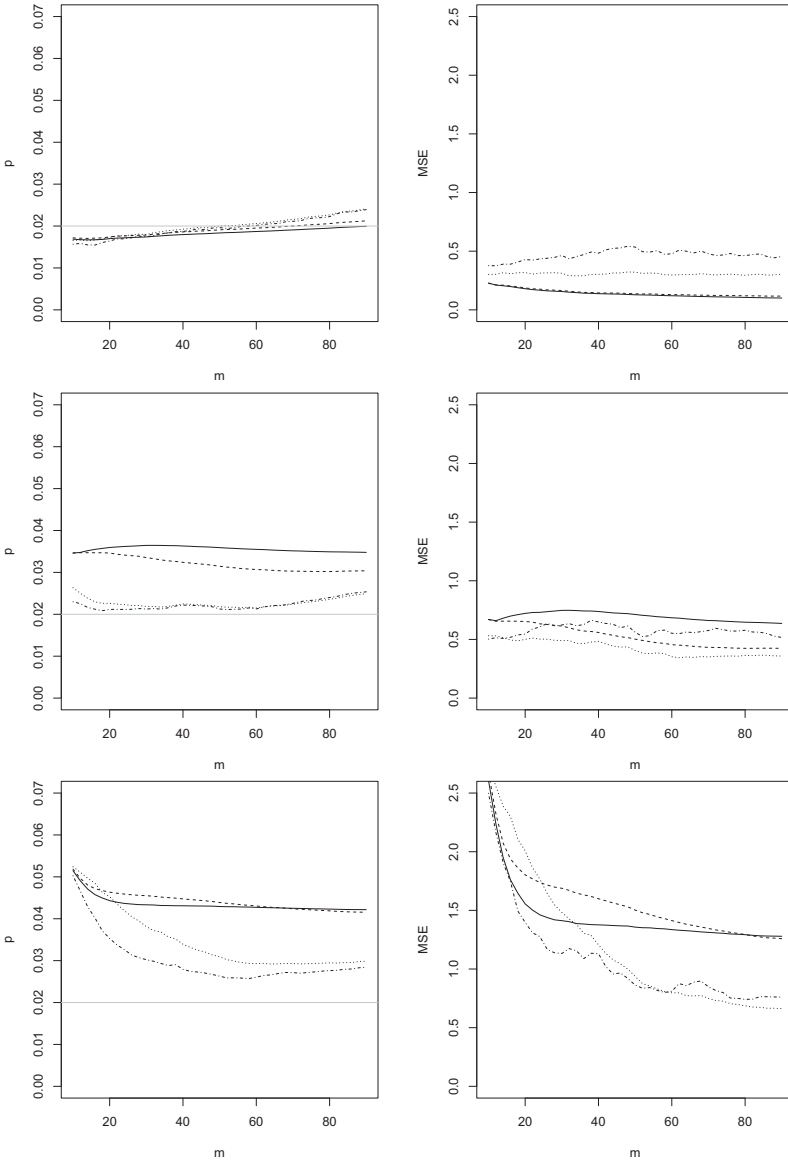
Figure 3: Frank copula: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 100, p_{z,1} = \frac{2}{n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 2\ \%$, and $\varepsilon = 5\ \%$. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,1}$.

Table 2: Values of the breakdown points, respectively for FGM $\zeta = 1, -1$, and Frank ($n = 100$).

	m	10	20	30	40	50	60	70	80	90
	$\alpha = 0$	3	3	2	2	2	2	3	3	3
FGM	$\alpha = 0.1$	3	3	3	3	3	3	3	4	4
$\zeta = 1$	$\alpha = 0.5$	3	4	5	5	6	6	7	7	7
	$\alpha = 1$	4	4	5	6	6	6	7	7	6
	$\alpha = 0$	2	1	1	1	1	1	1	1	1
FGM	$\alpha = 0.1$	2	2	2	2	2	2	2	3	3
$\zeta = -1$	$\alpha = 0.5$	3	3	4	5	6	7	9	10	> 10
	$\alpha = 1$	3	4	4	5	7	8	10	> 10	> 10
	$\alpha = 0$	3	3	3	3	3	3	3	3	3
Frank	$\alpha = 0.1$	3	3	3	3	3	4	4	4	4
	$\alpha = 0.5$	3	4	5	5	6	7	7	8	8
	$\alpha = 1$	4	5	5	5	6	7	8	8	8

- when a proper choice of α is used (all considered values of α in case of no contamination and $\alpha = 0.5$ or 1 when contamination is present), the estimators are clearly very stable as a function of m, and stay close to the true value of $p_{z,1}$ for a wide range of values for m, which illustrates the bias-correction of our procedure.

We also run our simulation study for smaller values of $p_{z,1}$ (out of the sample, see the right panel in Table 1) and larger sample size. The results are displayed in Figures 4, 5 and 6. To keep the length of the paper reasonable, we only show the results for the FGM distribution with $\zeta = 1$, which is, as explained above, the more challenging case. From the simulation we deduce:

- when contamination occurs, the superiority of our estimator $\widehat{p}_{n,\alpha}$ for $\alpha > 0$ is again observed. In terms of bias as well as MSE, $\alpha = 0.5$ seems to be nearly always the best choice;

- the estimators are again very stable as a function of m illustrating the bias-correction of our procedure;

- the MSEs are larger than in Figures 1–3 due to the fact that such estimation is much more challenging, since they are out of the sample.

As usual in the robust framework, we compute also the breakdown point of our estimator for contamination as follows:
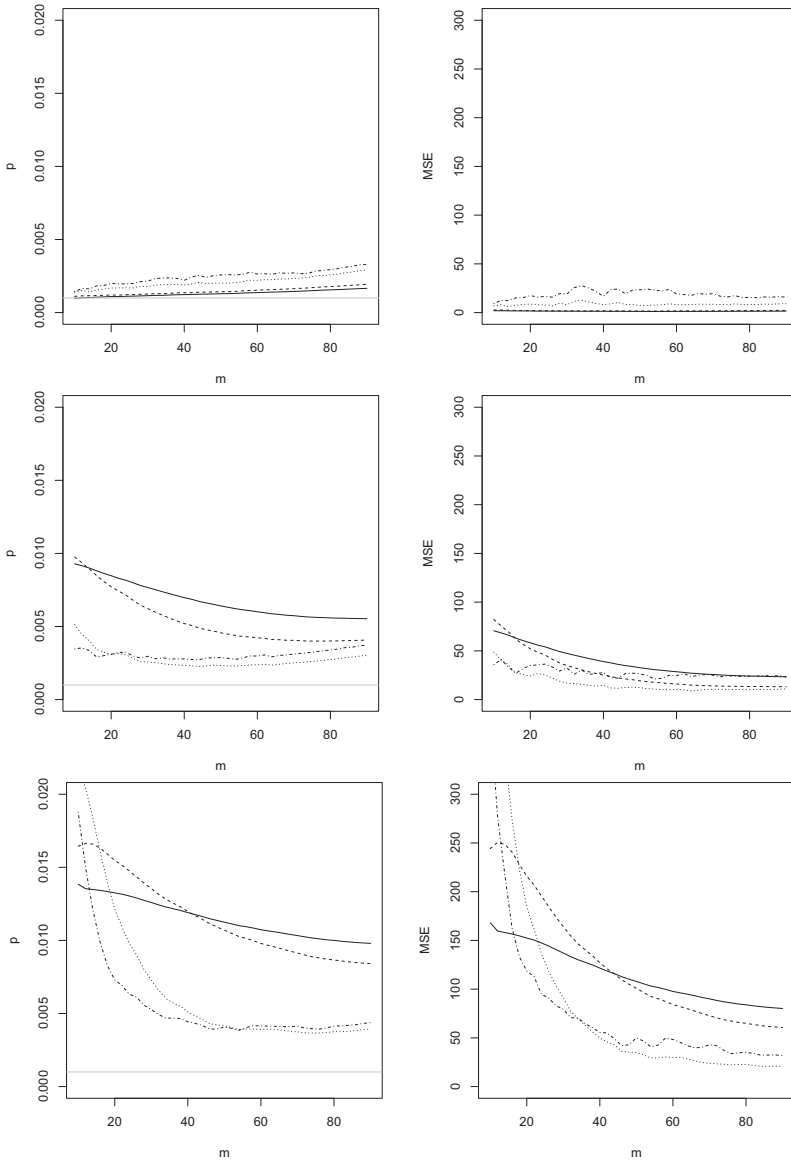
Figure 4: Farlie Gumbel Morgenstern copula with $\zeta = 1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 100, p_{z,1} = \frac{1}{10\,n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 2\,\%$, and $\varepsilon = 5\,\%$. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,1}$.
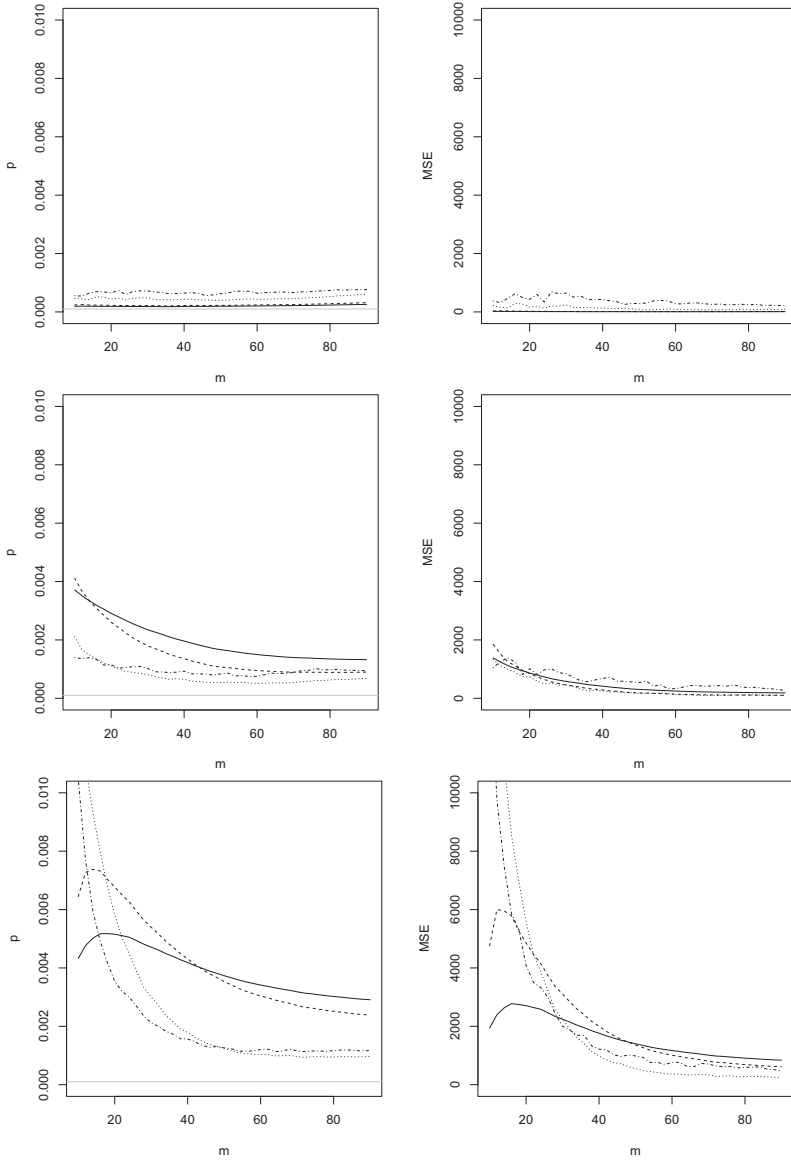
Figure 5: Farlie Gumbel Morgenstern copula with $\zeta = 1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 100, p_{z,1} = \frac{1}{100\,n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 2$ %, and $\varepsilon = 5$ %. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,1}$.
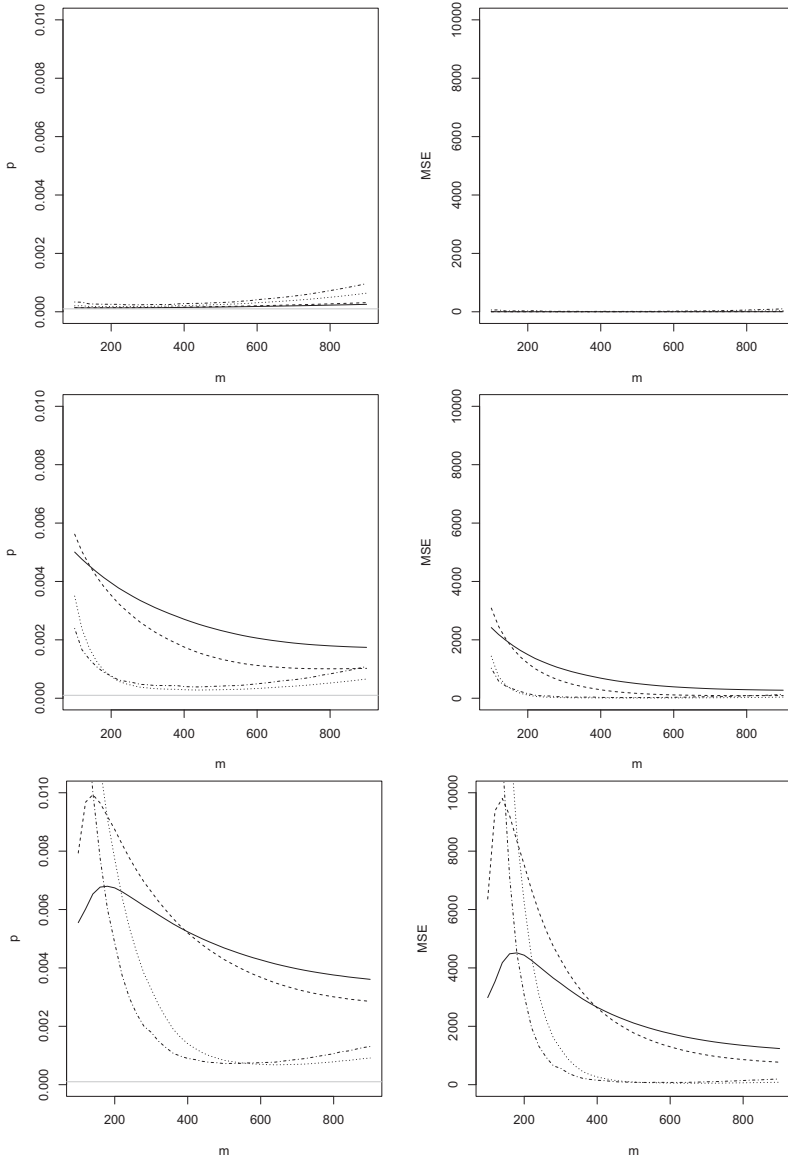
Figure 6: Farlie Gumbel Morgenstern copula with $\zeta = 1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 1000$, $p_{z,1} = \frac{1}{10\,n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with $\varepsilon = 2\ \%$, and $\varepsilon = 5\ \%$. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,1}$.

- we simulate 1024 samples of size n (100 and 1000) and take a percentage of contamination $\varepsilon = 0, 1\ \%, ..., 10\ \%$;

- for each value of ε and each sample size n, we compute our estimator $\widehat{p}_{n,\alpha}$ of $p_{z,1} = 2\ \%$ and the mean squared error of the ratio $\widehat{p}_{n,\alpha}/p_{z,1}$ as a function of m;

- for a fixed value of m, the breakdown point is defined as the smallest value of ε such that the MSE exceeds 1.

This breakdown value is reported in Table 2 for $n = 100$ and in Table 3 for $n = 1000$. Each table contains three subtables corresponding to the three distributions under consideration (FGM with $\zeta = 1, -1$, and Frank) and for each distribution we consider four values of α (0, 0.1, 0.5 and 1). For both sample sizes and whatever the distribution, the higher the value of α, the larger is the breakdown point. In particular, for $\alpha = 0.5$ or 1, the breakdown point is particularly high illustrating the robustness of our estimator $\widehat{p}_{n,\alpha}$, whereas for $\alpha = 0$, the breakdown point is low.

In order to examine the accuracy of the asymptotic results in finite samples we calculated the coverage probabilities of 95 % confidence intervals for $p_n = 0.01\ \%$ based on the result of Theorem 4, for the Farlie Gumbel Morgenstern copula with $\zeta = 1$ in case of a sample of size $n = 1000$. The results are displayed in Table 4. Note that the coverage probabilities are

Table 3: Values of the breakdown points, respectively for FGM $\zeta = 1, -1$, and Frank ($n = 1000$).

	m	100	200	300	400	500	600	700	800	900
	$\alpha = 0$	3	3	2	2	2	2	2	2	2
FGM	$\alpha = 0.1$	3	3	3	3	3	3	3	3	3
$\zeta = 1$	$\alpha = 0.5$	3	4	5	6	7	7	8	8	8
	$\alpha = 1$	3	5	6	7	8	9	9	8	6
	$\alpha = 0$	2	1	1	1	1	1	1	1	1
FGM	$\alpha = 0.1$	2	2	2	2	2	2	2	2	3
$\zeta = -1$	$\alpha = 0.5$	3	4	5	6	7	8	10	> 10	> 10
	$\alpha = 1$	3	4	5	6	7	9	> 10	> 10	> 10
	$\alpha = 0$	3	3	2	2	2	2	2	2	2
Frank	$\alpha = 0.1$	3	3	3	3	3	3	3	3	3
	$\alpha = 0.5$	3	4	5	6	6	7	8	8	8
	$\alpha = 1$	3	4	6	7	8	9	9	9	9

Table 4: FGM $\zeta = 1$: coverage probabilities of 95 % confidence intervals for $p_n = 0.01$ % based on Theorem 4 with $n = 1000$.

m	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$
200	0.9365	0.9287	0.9160	0.9111
400	0.9053	0.8975	0.8789	0.8750
600	0.7363	0.7109	0.6777	0.6562
800	0.9814	0.9854	0.9814	0.9756

generally somehow below 95 % but can be considered fairly good, especially if one takes into account that the use of the result of Theorem 4 is not so obvious in practice since in fact $p_n \to 0$ when $n \to \infty$, together with some additional assumptions on the involved sequences.

Furthermore, we consider the estimation of our probability $p_{z,\widetilde{\omega}}$, but this time outside the diagonal, that is for $\widetilde{\omega} \neq 1$. Figure 7 illustrates the behaviour of our estimator $\widehat{p}_{n,\alpha}$ in the uncontaminated case for the FGM copula with $\zeta = 1$, the same values of α, and two values of ω: 0.45 (first row) and 0.55 (second row). In both cases, we can observe the almost best behaviour of $\widehat{p}_{n,0}$ compared to $\widehat{p}_{n,\alpha}$, $\alpha > 0$, as expected in this framework. However, the most notable point compared to Figure 1 is that the estimation is much more difficult, with a quite stable bias appearing. This can be partly explained by our Theorems 3 and 4. Indeed, if we look at the bias term of our estimator $\widehat{p}_{n,\alpha}$, we have to compute the expectations of Γ and Δ which can be deduced from our Theorem 2. Straightforward computations lead to an expectation for both variables of the form $\lambda \xi_\omega c_\omega^{\rho_0} f(\eta_0, \rho_0, \alpha)$ where the function $f(\eta_0, \rho_0, \alpha)$ does not depend on ω. As the sense of variation of the function $\omega \mapsto \xi_\omega c_\omega^{\rho_0}$ depends if ω is smaller or larger than $1/2$ for our three bivariate distribution functions, the estimation on the diagonal is simpler than outside.

Finally, to illustrate that mis-specifying the parameter ρ is not crucial, we show in Figure 8 the behaviour of our estimator $\widehat{p}_{n,\alpha}$ for the FGM copula with $\zeta = 1$ in the uncontaminated framework in case ρ is replaced by the true $\rho_0 = -1/2$. The same values of α have been used and ω is set to $1/2$. This figure is indeed very similar to the first row of Figure 1 where ρ is mis-specified to -1.

5   Illustration on an Actuarial Dataset

We consider an actuarial dataset published by the Australian regulator of insurance services which consists of company-wide performance indicators for insurers operating in Australia between 2005 and 2010, see
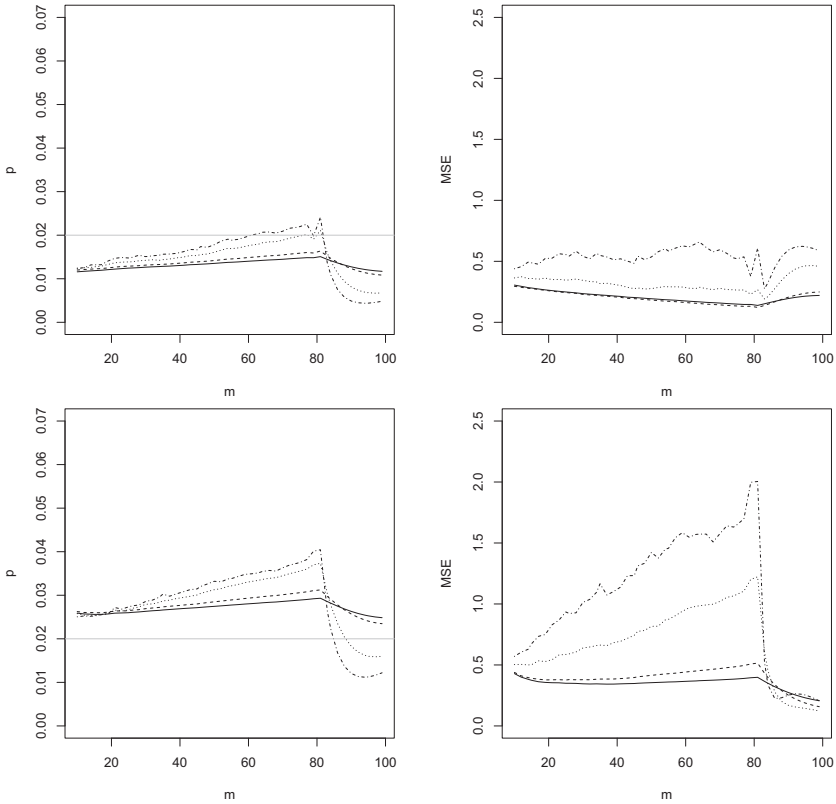
Figure 7: Farlie Gumbel Morgenstern copula with $\zeta = 1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,\widetilde{\omega}}$ for $n = 100, p_{z,1} = \frac{2}{n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations in the uncontaminated case. The first row corresponds to $\omega = 0.45$ and the second row to $\omega = 0.55$, where $\omega = (1 + \widetilde{\omega})^{-1}$. The parameter ρ is mis-specified to the value -1. The horizontal reference line in the left panels of the figure corresponds to the true value of $p_{z,\widetilde{\omega}}$.

http://www.apra.gov.au/. This dataset contains $n = 422$ observations of (gross) loss ratios X and expense ratios Y of 173 private insurance companies. Possible time dependency among the data points is ignored in this paper.

Figures 9(a) and (b) show the scatterplot of the original data and the data transformed into unit Pareto margins, respectively. The Pareto quantile plot of the $\widetilde{Z}_{\frac{1}{2},i}$, $i = 1, ..., n$, is depicted in Figure 9(c). The three points labelled with crosses correspond to possible outliers in the dataset. Their values
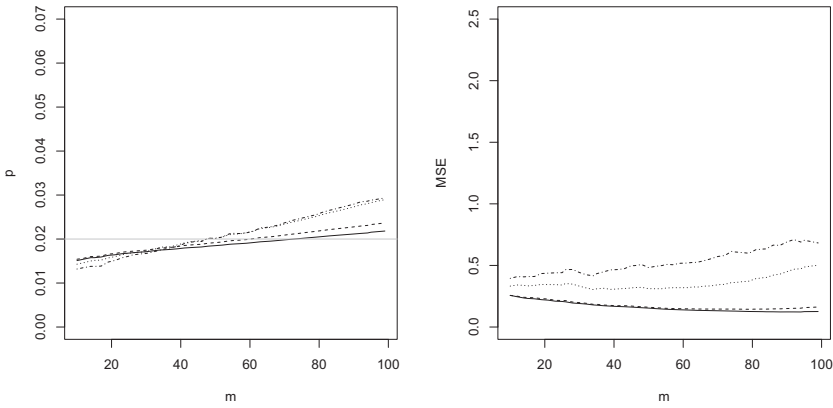
Figure 8: Farlie Gumbel Morgenstern copula with $\zeta = 1$: mean (left) of $\widehat{p}_{n,\alpha}$, MSE (right) of $\widehat{p}_{n,\alpha}/p_{z,1}$ for $n = 100, p_{z,1} = \frac{2}{n}$ and different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations in the uncontaminated case. The parameter ρ is set to its true value $\rho_0 = -1/2$. The horizontal reference line in the left panel of the figure corresponds to the true value of $p_{z,1}$.

are given in Table 5. Since X and Y are ratios (in terms of gross written premium), we may reasonably qualify these points $(X_i, Y_i \gg 1)$ as outliers. In the sequel, we will denote by X and Y the original random variables, and by $\widetilde{X}$ and $\widetilde{Y}$ the random variables on the unit Pareto scale, as obtained from the inverse probability integral transform.
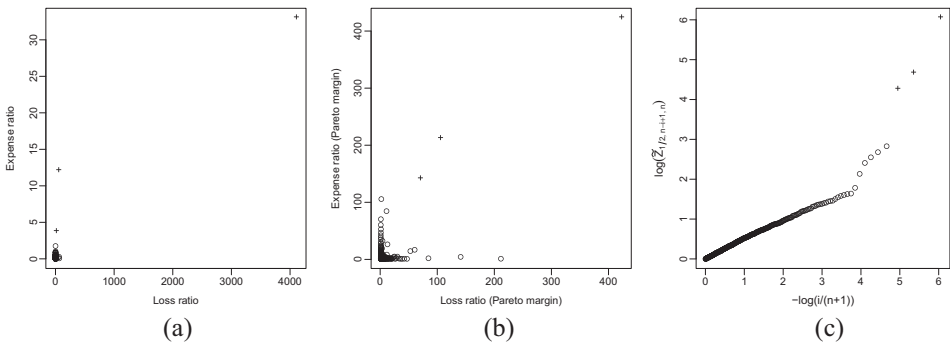

Figure 9: Loss/expense ratios dataset: (**a**) Scatterplot of the original data; (**b**) Scatterplot of the data transformed into unit Pareto margins; (**c**) Pareto quantile plot of the $\widetilde{Z}_{\frac{1}{2},i}$, $i = 1, ..., n$.

Table 5: Three points labelled with crosses in Figure 9.

i	X_i	Y_i	$\widetilde{Z}_{\frac{1}{2},i}$
184	4108	33	423
101	55.133	12.067	105.75
21	14.52	3.72	70.5

Now, in order to detect if these three points listed in Table 5 are outliers or not, we consider the estimation of the probability $\mathbb{P}(\widetilde{X} > 10, \widetilde{Y} > 10)$, with respect to the presence or the removal of these three points. The choices $z = 10$ and $\widetilde{\omega} = 1$ can be justified by the fact that $\mathbb{P}(\widetilde{X} > 10, \widetilde{Y} > 10)$ is a possible indicator of bankruptcy since it is unlikely that an insurer can deal with both high losses and high expenses. Figure 10 shows the estimators $\widehat{p}_{n,\alpha}$ as a function of m for different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line). In Figure 10(a), we use the whole dataset $i \in N := \{1, \ldots, n\}$ to estimate $\widehat{p}_{n,\alpha}$, whereas in Figure 10(b) we use $N \setminus \{184, 101, 21\}$. Deleting these three points clearly lead to estimators close to each other, for m not too small, whatever the value of α is. This corroborates our feeling that the three points are outliers. Indeed, as observed in Figure 10(a) the estimator $\widehat{p}_{n,\alpha}$ differs considerably
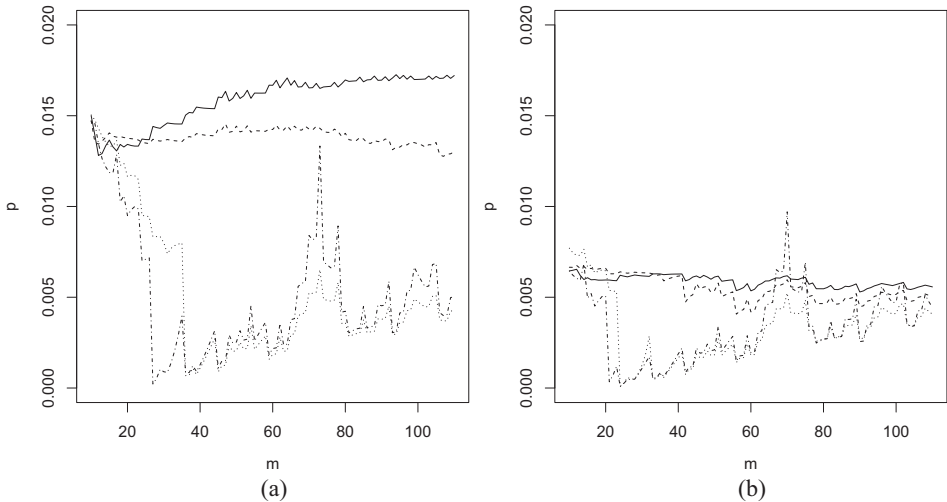

Figure 10: Loss/expense ratios dataset: Estimator $\widehat{p}_{n,\alpha}$ of $\mathbb{P}(\widetilde{X} > 10, \widetilde{Y} > 10)$ for different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line). (a) Whole dataset; (b) Dataset without the three points.
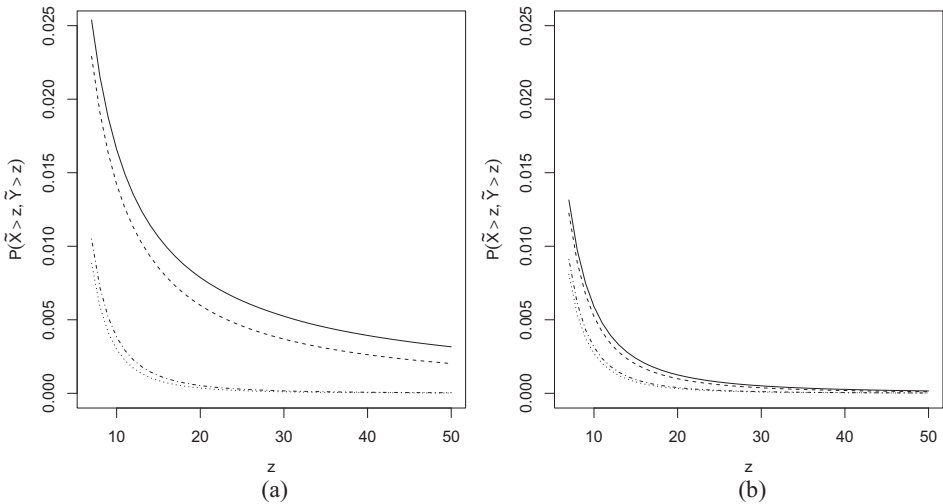
Figure 11: Median of the estimator $\widehat{p}_{n,\alpha}$ of $\mathbb{P}(\widetilde{X} > z, \widetilde{Y} > z)$ over $m = 50, \ldots, 80$ for different values of α: 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line). (a) Whole dataset; (b) Dataset without the three points.

depending on the value of α which is expected in the presence of outliers. As in the simulation study, the estimator $\widehat{p}_{n,0.5}$ remains stable both in the situation with and without outliers.

Finally, Figure 11 shows the median over $m = 50, \ldots, 80$ of our probability estimates $\widehat{p}_{n,\alpha}$ of $\mathbb{P}(\widetilde{X} > z, \widetilde{Y} > z)$ for increasing values of z and different values of α. Again, two cases have been considered: the whole dataset in panel (a) and the dataset without the three points in panel (b). The same conclusions follow, i.e. the estimators are close to each others if the three points are removed from the initial dataset and $\widehat{p}_{n,0.5}$ exhibits a more stable pattern than $\widehat{p}_{n,\alpha}$ for other values of α in case of the whole dataset. Both numerical illustrations of this section advocate $\alpha = 0.5$ as a good value, in line with the results of Section 4.

Appendix

The key element of our proofs is the following theorem established in Dutang et al. (2014).

THEOREM 5. (Dutang et al., 2014).   *Under the assumptions of Theorem 1, we have that there exist suitable versions of Q_n, a suitable process $\overline{W}(./c_\omega)$, equal in distribution to a standard Brownian motion, such that for all $t_0, \varepsilon > 0$*

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

$$\left. + \lambda \eta^2 \left(\frac{t}{c_\omega} \right)^{-\eta} \frac{\left(\frac{t}{c_\omega} \right)^{-\rho} \xi_\omega - 1}{\rho} \right| = o_\mathbb{P}(1).$$

PROOF OF THEOREM 2.   We follow the lines of proof of Theorem 2 in Dierckx et al. (2013). We give here only the adjustements needed. Define

$$\overline{A}_{m,n}(\widetilde{\rho}) := \sqrt{m} \begin{bmatrix} \frac{k}{n} Q_n(1) - c_\omega^{\eta_0} \\ A_{m,n}^{(1)} \left(-\frac{\alpha(1+\eta_0)}{\eta_0} \right) - \frac{1}{1 + \alpha(1+\eta_0)} \\ A_{m,n}^{(1)} \left(-\frac{\alpha(1+\eta_0) - \widetilde{\rho}}{\eta_0} \right) - \frac{1}{1 - \widetilde{\rho} + \alpha(1+\eta_0)} \\ A_{m,n}^{(2)} \left(-\frac{\alpha(1+\eta_0)}{\eta_0} \right) - \frac{\eta_0}{(1 + \alpha(1+\eta_0))^2} \end{bmatrix}.$$

We can easily infer from Theorem 1 that $\overline{A}_{m,n}(\widetilde{\rho}) \rightsquigarrow A(\widetilde{\rho}) \sim N_4(\lambda \widetilde{\boldsymbol{\mu}}, \mathbb{D}(\widetilde{\rho}))$. From the proof of Theorem 2 in Dierckx et al. (2013), we deduce that

$$\sqrt{m} \begin{bmatrix} \frac{k}{n} Q_n(1) - c_\omega^{\eta_0} \\ \widehat{\eta}_n - \eta_0 \\ \widehat{\delta}_{\omega,n} \end{bmatrix} \rightsquigarrow \mathbb{C}^{-1}(\widetilde{\rho}) \mathbb{B}(\widetilde{\rho}) A(\widetilde{\rho}) \sim N_3(\lambda \mathbb{C}^{-1}(\widetilde{\rho}) \mathbb{B}(\widetilde{\rho}) \widetilde{\boldsymbol{\mu}}, \mathbb{C}^{-1}(\widetilde{\rho}) \mathbb{B}(\widetilde{\rho}) \mathbb{D}(\widetilde{\rho}) \mathbb{B}'(\widetilde{\rho}) \mathbb{C}^{-1}(\widetilde{\rho})).$$

Taking into account that $\sqrt{m} \delta_{\omega,n} \xrightarrow{\mathbb{P}} -\lambda \eta_0^2 \xi_\omega c_\omega^{\rho_0} / \rho_0$, Theorem 2 follows.

PROOF OF THEOREM 3.   We use the following decomposition

$$\ln \frac{\widehat{p}_n}{p_n} = \ln \frac{\frac{m}{n} \overline{H} \left(\frac{z_n}{\widetilde{Z}_{\omega, n-m, n}}; \widehat{\eta}_n, \widehat{\delta}_{\omega,n}, \widetilde{\rho} \right)}{\overline{F}_{Z_\omega}(z_n)}$$

$$= -\ln \left(\frac{n}{m} \overline{F}_{Z_\omega}(\widetilde{Z}_{\omega, n-m, n}) \right)$$

$$+ \left[\ln \overline{H} \left(\frac{z_n}{\widetilde{Z}_{\omega, n-m, n}}; \eta_0, \delta_{\omega,n}, \rho_0 \right) - \ln \frac{\overline{F}_{Z_\omega} \left(\frac{z_n}{\widetilde{Z}_{\omega, n-m, n}} \widetilde{Z}_{\omega, n-m, n} \right)}{\overline{F}_{Z_\omega}(\widetilde{Z}_{\omega, n-m, n})} \right]$$

$$+ \left[\ln \overline{H} \left(\frac{z_n}{\widetilde{Z}_{\omega, n-m, n}}; \widehat{\eta}_n, \widehat{\delta}_{\omega,n}, \widetilde{\rho} \right) - \ln \overline{H} \left(\frac{z_n}{\widetilde{Z}_{\omega, n-m, n}}; \eta_0, \delta_{\omega,n}, \rho_0 \right) \right]$$

$$=: T_1 + T_2 + T_3.$$

We will study the three terms separately. Let $Z_{\omega,1,n} \leq \ldots \leq Z_{\omega,n,n}$ denote the order statistics of a random sample of size n from F_{Z_ω}.

Concerning T_1. This term can be rewritten as

$$
\begin{aligned}
T_1 &= -\ln \frac{\overline{F}_{Z_\omega}(\widetilde{Z}_{\omega,n-m,n})}{\overline{F}_{Z_\omega}(Z_{\omega,n-m,n})} - \ln \left(\frac{n}{m} \overline{F}_{Z_\omega}(Z_{\omega,n-m,n}) \right) \\
&= \frac{1}{\eta_0} \ln \frac{\widetilde{Z}_{\omega,n-m,n}}{Z_{\omega,n-m,n}} - \ln \frac{1 + \frac{1}{\eta_0}\delta_{\omega,n}}{1 + \frac{1}{\eta_0}\delta_\omega(Z_{\omega,n-m,n})} - \ln \left(\frac{n}{m} \overline{F}_{Z_\omega}(Z_{\omega,n-m,n}) \right) \\
&=: \ T_{11} + T_{12} + T_{13}.
\end{aligned}
$$

From Theorem 5, we easily infer that

$$
\frac{k}{n} \widetilde{Z}_{\omega,n-m,n} = c_\omega^{\eta_0} + \frac{\eta_0}{\sqrt{m}} c_\omega^{\eta_0} \overline{W}\left(\frac{1}{c_\omega}\right) - \frac{\lambda}{\sqrt{m}} \eta_0^2 c_\omega^{\eta_0} \frac{c_\omega^{\rho_0}\xi_\omega - 1}{\rho_0} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right).
$$

Thus, if we denote by Q_{Z_ω} the quantile function of the variable Z_ω and if $U_{k,n}$ denotes the k-th order statistic of a random sample of n uniform $(0,1)$ random variables, we deduce that

$$
\begin{aligned}
\ln \frac{\widetilde{Z}_{\omega,n-m,n}}{Z_{\omega,n-m,n}} \ &\overset{d}{=} \ \ln \frac{\frac{n}{k} c_\omega^{\eta_0} \left(1 + \frac{\eta_0}{\sqrt{m}}\overline{W}\left(\frac{1}{c_\omega}\right) - \frac{\lambda}{\sqrt{m}}\eta_0^2 \frac{c_\omega^{\rho_0}\xi_\omega - 1}{\rho_0} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)\right)}{Q_{Z_\omega}(U_{n-m,n})} \\
&\overset{d}{=} \ \ln \frac{\frac{n}{k} c_\omega^{\eta_0} \left(1 + \frac{\eta_0}{\sqrt{m}}\overline{W}\left(\frac{1}{c_\omega}\right) - \frac{\lambda}{\sqrt{m}}\eta_0^2 \frac{c_\omega^{\rho_0}\xi_\omega - 1}{\rho_0} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)\right)}{C_\omega^{\eta_0}\left(\frac{m}{n}\right)^{-\eta_0} \left(\frac{n}{m}U_{m+1,n}\right)^{-\eta_0}(1 + \delta_\omega(C_\omega^{\eta_0}U_{m+1,n}^{-\eta_0})(1 + o_{\mathbb{P}}(1)))} \\
&= \ \ln \frac{\frac{n}{k} c_\omega^{\eta_0} \left(1 + \frac{\eta_0}{\sqrt{m}}\overline{W}\left(\frac{1}{c_\omega}\right) - \frac{\lambda}{\sqrt{m}}\eta_0^2 \frac{c_\omega^{\rho_0}\xi_\omega - 1}{\rho_0} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)\right)}{C_\omega^{\eta_0}\left(q(\frac{k}{n})\right)^{-\eta_0} \left(1 + [\frac{n}{m}U_{m+1,n} - 1]\right)^{-\eta_0}(1 + \delta_\omega(C_\omega^{\eta_0}U_{m+1,n}^{-\eta_0})(1 + o_{\mathbb{P}}(1)))}.
\end{aligned}
$$

Recall now that

$$
\begin{aligned}
q\left(\frac{k}{n}\right) &= \mathbb{P}\left(1 - F_X(X) < \frac{k}{n}, 1 - F_Y(Y) < \frac{k}{n}\right) = \mathbb{P}\left(X > \frac{n}{k}, Y > \frac{n}{k}\right) \\
&= C_{\frac{1}{2}}\left(\frac{n}{k}\right)^{-1/\eta_0}\left(1 + \frac{1}{\eta_0}\delta_{\frac{1}{2}}\left(\frac{n}{k}\right)\right) = g(1,1)\left(\frac{n}{k}\right)^{-1/\eta_0}\left(1 + \frac{1}{\eta_0}\delta_{\frac{1}{2}}\left(\frac{n}{k}\right)\right) \quad (15)
\end{aligned}
$$

and by Lemma 1

$$
c_\omega = \left(\frac{\omega}{1-\omega}\right)^{d_2} \frac{g(1,\frac{\omega}{1-\omega})}{g(1,1)} = \frac{C_\omega}{g(1,1)}. \quad (16)
$$

Thus using the convergence

$$\sqrt{m}\left(\frac{n}{m}U_{m+1,n}-1\right)\rightsquigarrow \Omega \sim N(0,1), \tag{17}$$

we deduce that

$$
\begin{aligned}
\ln\frac{\widetilde{Z}_{\omega,n-m,n}}{Z_{\omega,n-m,n}} &= \ln\frac{1+\frac{\eta_0}{\sqrt{m}}\overline{W}\left(\frac{1}{c_\omega}\right)-q_1(\frac{k}{n})\,\eta_0^2\,\frac{c_\omega^{\rho_0}\xi_\omega-1}{\rho_0}+o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)}{\left(1-\delta_{\frac{1}{2}}\left(\frac{n}{k}\right)(1+o(1))\right)\left(1-\frac{\eta_0}{\sqrt{m}}\Omega+o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)\right)(1+\delta_\omega(C_\omega^{\eta_0}U_{m+1,n}^{-\eta_0})(1+o_{\mathbb{P}}(1)))} \\[2mm]
&= \ln\frac{1+\frac{\eta_0}{\sqrt{m}}\overline{W}\left(\frac{1}{c_\omega}\right)-q_1(\frac{k}{n})\,\eta_0^2\,\frac{c_\omega^{\rho_0}\xi_\omega-1}{\rho_0}+o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)}{1-\frac{\eta_0}{\sqrt{m}}\Omega-\delta_{\frac{1}{2}}\left(\frac{n}{k}\right)(1+o(1))+\delta_\omega(C_\omega^{\eta_0}U_{m+1,n}^{-\eta_0})(1+o_{\mathbb{P}}(1))+o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)}.
\end{aligned}
$$

Combining Lemma 1 with the fact that $|q_1|$ is regularly varying with index τ and (3), we can infer that

$$
\begin{aligned}
\delta_\omega(C_\omega^{\eta_0}U_{m+1,n}^{-\eta_0})) &= -\xi_\omega\frac{\eta_0^2}{\rho_0}q_1\left(C_\omega^{-\eta_0}U_{m+1,n}^{\eta_0}\right)(1+o_{\mathbb{P}}(1)) \\[2mm]
&= -\xi_\omega\frac{\eta_0^2}{\rho_0}C_\omega^{\rho_0}q_1\left(\left(\frac{m}{n}\right)^{\eta_0}\right)(1+o_{\mathbb{P}}(1)) \\[2mm]
&= -\xi_\omega\frac{\eta_0^2}{\rho_0}C_\omega^{\rho_0}q_1\left(q^{\eta_0}\left(\frac{k}{n}\right)\right)(1+o_{\mathbb{P}}(1)) \\[2mm]
&= -\xi_\omega\frac{\eta_0^2}{\rho_0}C_\omega^{\rho_0}q_1\left(\frac{k}{n}g^{\eta_0}(1,1)\right)(1+o_{\mathbb{P}}(1)) \\[2mm]
&= -\xi_\omega\frac{\eta_0^2}{\rho_0}c_\omega^{\rho_0}q_1\left(\frac{k}{n}\right)(1+o_{\mathbb{P}}(1)),
\end{aligned}
$$

the last equality coming from (16).

By an application of Taylor's theorem we finally get that

$$T_{11}=\frac{1}{\sqrt{m}}\overline{W}\left(\frac{1}{c_\omega}\right)+\frac{1}{\sqrt{m}}\Omega+o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right).$$

We have now to study the second term, T_{12}:

$$
\begin{aligned}
T_{12} &= -\frac{1}{\eta_0}\left(\delta_{\omega,n}(1+o_{\mathbb{P}}(1))-\delta_\omega(Z_{\omega,n-m,n})(1+o_{\mathbb{P}}(1))\right) \\[2mm]
&= -\frac{1}{\eta_0}\left[\delta_\omega\left(\frac{n}{k}c_\omega^{\eta_0}\right)(1+o_{\mathbb{P}}(1))-\delta_\omega\left(C_\omega^{\eta_0}U_{m+1,n}^{-\eta_0}\right)(1+o_{\mathbb{P}}(1))\right] \\[2mm]
&= \frac{1}{\eta_0}\left[\xi_\omega\frac{\eta_0^2}{\rho_0}q_1\left(\frac{k}{n}c_\omega^{-\eta_0}\right)(1+o_{\mathbb{P}}(1))-\xi_\omega\frac{\eta_0^2}{\rho_0}q_1\left(C_\omega^{-\eta_0}U_{m+1,n}^{\eta_0}\right)(1+o_{\mathbb{P}}(1))\right] \\[2mm]
&= \frac{\xi_\omega\eta_0}{\rho_0}\left[q_1\left(\frac{k}{n}c_\omega^{-\eta_0}\right)(1+o_{\mathbb{P}}(1))-q_1\left(C_\omega^{-\eta_0}q^{\eta_0}\left(\frac{k}{n}\right)\right)(1+o_{\mathbb{P}}(1))\right]
\end{aligned}
$$

$$= \quad o_{\mathbb{P}} \left(q_1 \left(\frac{k}{n} \right) \right),$$

by (15) and (16).

Thus

$$T_{12} \quad = \quad o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right).$$

Now the last term T_{13} can be treated as follows

$$T_{13} \quad \overset{d}{=} \quad -\ln \left(\frac{n}{m} \overline{F}_{Z_\omega} \left(Q_{Z_\omega}(U_{n-m,n}) \right) \right)$$

$$\overset{d}{=} \quad -\left(\frac{n}{m} U_{m+1,n} - 1 \right) (1 + o_{\mathbb{P}}(1))$$

$$= \quad -\frac{\Omega}{\sqrt{m}} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right)$$

by (17). Consequently

$$T_1 = \frac{1}{\sqrt{m}} \overline{W} \left(\frac{1}{c_\omega} \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right). \tag{18}$$

Concerning T_2. Using (5), (11) and (16), we have the two convergences

$$\frac{n}{m} \overline{F}_{Z_\omega}(z_n) \quad = \quad \frac{np_n}{m} \longrightarrow \beta,$$

$$\frac{n}{m} \overline{F}_{Z_\omega}(\widetilde{Z}_{\omega,n-m,n}) \quad = \quad \frac{n}{m} C_\omega \widetilde{Z}_{\omega,n-m,n}^{-1/\eta_0}(1 + o_{\mathbb{P}}(1))$$

$$= \quad \frac{n}{m} \frac{C_\omega}{c_\omega} \left(q^\leftarrow \left(\frac{m}{n} \right) \right)^{1/\eta_0} (1 + o_{\mathbb{P}}(1)) = 1 + o_{\mathbb{P}}(1),$$

from which we deduce, by using Proposition 2.3 from Beirlant et al. (2009), that

$$T_2 \quad = \quad \frac{\overline{H} \left(\dfrac{z_n}{\widetilde{Z}_{\omega,n-m,n}}; \eta_0, \delta_{\omega,n}, \rho_0 \right) - \dfrac{\overline{F}_{Z_\omega} \left(\dfrac{z_n}{\widetilde{Z}_{\omega,n-m,n}} \widetilde{Z}_{\omega,n-m,n} \right)}{\overline{F}_{Z_\omega}(\widetilde{Z}_{\omega,n-m,n})}}{\dfrac{\overline{F}_{Z_\omega}(z_n)}{\overline{F}_{Z_\omega}(\widetilde{Z}_{\omega,n-m,n})}} (1 + o_{\mathbb{P}}(1))$$

$$= \quad o_{\mathbb{P}} \left(\delta_{\omega,n} \right)$$

$$= \quad o_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right). \tag{19}$$

Concerning T_3. Remark that

$$
\begin{aligned}
T_3 &= \ln\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\left(1+\widehat{\delta}_{\omega,n}-\widehat{\delta}_{\omega,n}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\right)^{\widetilde{\rho}/\widehat{\eta}_n}\right)\right)^{-1/\widehat{\eta}_n}\\
&\quad -\ln\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\left(1+\delta_{\omega,n}-\delta_{\omega,n}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\right)^{\rho_0/\eta_0}\right)\right)^{-1/\eta_0}\\
&= \left(\frac{1}{\eta_0}-\frac{1}{\widehat{\eta}_n}\right)\ln\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\\
&\quad -\frac{1}{\widehat{\eta}_n}\left[\ln\left(1+\widehat{\delta}_{\omega,n}-\widehat{\delta}_{\omega,n}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\right)^{\widetilde{\rho}/\widehat{\eta}_n}\right)\right.\\
&\qquad\qquad \left.-\ln\left(1+\delta_{\omega,n}-\delta_{\omega,n}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\right)^{\rho_0/\eta_0}\right)\right]\\
&\quad +\left(\frac{1}{\eta_0}-\frac{1}{\widehat{\eta}_n}\right)\ln\left(1+\delta_{\omega,n}-\delta_{\omega,n}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\right)^{\rho_0/\eta_0}\right).
\end{aligned}
$$

Using the fact that $\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}=\frac{C_\omega^{\eta_0}p_n^{-\eta_0}}{\frac{n}{k}c_\omega^{\eta_0}}(1+o_{\mathbb{P}}(1))=\left(\frac{m}{np_n}\right)^{\eta_0}(1+o_{\mathbb{P}}(1))\xrightarrow{\mathbb{P}}\beta^{-\eta_0}$, we can easily deduce that

$$
T_3=-\frac{\ln\beta}{\sqrt{m}}\frac{\Gamma}{\eta_0}-\frac{1}{\eta_0}\frac{1}{\sqrt{m}}\Delta(1-\beta^{-\widetilde{\rho}})-\frac{\delta_{\omega,n}}{\eta_0}(\beta^{-\rho_0}-\beta^{-\widetilde{\rho}})+o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right). \quad (20)
$$

Combining (18)–(20) with the fact that $\sqrt{m}\delta_{\omega,n}\xrightarrow{\mathbb{P}}-\lambda\eta_0^2\xi_\omega c_\omega^{\rho_0}/\rho_0$, our theorem follows.

PROOF OF THEOREM 4.   By straightforward calculations one obtains the following decomposition

$$
\begin{aligned}
\ln\frac{\widehat{p}_n}{p_n} &= \ln\left[\frac{m}{np_n}\left(\frac{z_n}{Z_{\omega,n-m,n}}\right)^{-1/\widehat{\eta}_n}\right]-\frac{1}{\widehat{\eta}_n}\ln\frac{Z_{\omega,n-m,n}}{\widetilde{Z}_{\omega,n-m,n}}\\
&\quad -\frac{1}{\widehat{\eta}_n}\ln\left[1+\widehat{\delta}_{\omega,n}-\widehat{\delta}_{\omega,n}\left(\frac{z_n}{\widetilde{Z}_{\omega,n-m,n}}\right)^{\widetilde{\rho}/\widehat{\eta}_n}\right]\\
&=: T_4+T_5+T_6.
\end{aligned}
$$

For T_4, note that

$$\widetilde{p}_n := \frac{m}{n} \left(\frac{z_n}{Z_{\omega,n-m,n}} \right)^{-1/\widehat{\eta}_n}$$

is in fact the well-known Weissman estimator (Weissman, 1978), and hence

$$T_4 = \frac{\widehat{\eta}_n - \eta_0}{\widehat{\eta}_n} \ln d_n + O_{\mathbb{P}} \left(\frac{1}{\sqrt{m}} \right).$$

From the proof of Theorem 3 we have that $T_5 = O_{\mathbb{P}}(1/\sqrt{m})$ (see the analysis of the term T_{11}), and $T_6 = O_{\mathbb{P}}(1/\sqrt{m})$ (since $\widehat{\delta}_{\omega,n} = O_{\mathbb{P}}(1/\sqrt{m})$ and $z_n/\widetilde{Z}_{\omega,n-m,n} \to \infty$). Collecting all the terms establishes the result.

References

BASU, A., HARRIS, I.R., HJORT, N.L. and JONES, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika,* **85**, 549–559.

BEIRLANT, J., DIERCKX, G. and GUILLOU, A. (2011). Bias-reduced estimators for bivariate tail modelling. *Insurance Math. Econom.,* **49**, 18–26.

BEIRLANT, J., JOOSSENS, E. and SEGERS, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *J. Statist. Plann. Inference,* **139**, 2800–2815.

BEIRLANT, J. and VANDEWALLE, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value in statistics. *Statist. Probab. Lett.,* **60**, 265–278.

CHARPENTIER, A. and JURI, A. (2006). Limiting dependence structures for tail events, with applications to credit derivatives. *J. Appl. Probab.,* **43**, 563–586.

CIUPERCA, G. and MERCADIER, C. (2010). Semi-parametric estimation for heavy tailed distributions. *Extremes,* **13**, 55–87.

DE HAAN, L. and DE RONDE, J. (1998). Sea and wind: multivariate extremes at work. *Extremes,* **1**, 7–45.

DEKKERS, A.L.M., EINMAHL, J.H.J and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.,* **17**, 1833–1855.

DIERCKX, G., GOEGEBEUR, Y. and GUILLOU, A. (2013). An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *J. Multivariate Anal.,* **121**, 70–86.

DIERCKX, G., GOEGEBEUR, Y. and GUILLOU, A. (2014). Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. *Test,* **23**, 330–355.

DRAISMA, G., DREES, H., FERREIRA, A. and DE HAAN, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli,* **10**, 251–280.

DREES, H. (1998a). On smooth statistical tail functionals. *Scand. J. Stat.,* **25**, 187–210.

DREES, H. (1998b). A general class of estimators of the extreme value index. *J. Statist. Plann. Inference,* **66**, 95–112.

DUTANG, C., GOEGEBEUR, Y. and GUILLOU, A. (2014). Robust and bias-corrected estimation of the coefficient of tail dependence. *Insurance Math. Econom.,* **57**, 46–57.

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.,* **27**, 760–781.

Fraga Alves, M.I., Gomes, M.I and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Port. Math.,* **60**, 193–213.

Goegebeur, Y., Beirlant, J and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics. *J. Statist. Plann. Inference,* **140**, 2632–2652.

Goegebeur, Y. and Guillou, A. (2013). Asymptotically unbiased estimation of the coefficient of tail dependence. *Scand. J. Stat.,* **40**, 174–189.

Gomes, M.I. and Martins, M.J. (2004). Bias reduction and explicit semi-parametric estimation of the tail index. *J. Statist. Plann. Inference,* **124**, 361–378.

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

Kim, M. and Lee, S. (2008). Estimation of a tail index based on minimum density power divergence. *J. Multivariate Anal.,* **99**, 2453–2471.

Ledford, A.W. and Tawn, J.A. (1997). Modelling dependence within joint tail regions. *J. R. Stat. Soc. Ser. B,* **59**, 475–499.

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.,* **43**, 399–409.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.,* **3**, 119–131.

Ramos, A. and Ledford, A.W. (2009). A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B,* **71**, 219–241.

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

Christophe Dutang
Laboratoire Manceau de
Mathématique,
Université du Maine,
Avenue Olivier Messiaen,
72000 Le Mans, France
E-mail: christophe.dutang@univ-lemans.fr

Yuri Goegebeur
Department of Mathematics and
Computer Science,
University of Southern Denmark,
Campusvej 55, 5230 Odense M,
Denmark
E-mail: yuri.goegebeur@imada.sdu.dk

ARMELLE GUILLOU
INSTITUT RECHERCHE MATHÉMATIQUE
AVANCÉE, UMR 7501,
UNIVERSITÉ DE STRASBOURG ET CNRS,
7 RUE RENÉ DESCARTES, 67084
STRASBOURG CEDEX, FRANCE
E-mail: armelle.guillou@math.unistra.fr