

# Robust and bias-corrected estimation of the probability of extreme failure sets

Christophe Dutang \*

Yuri Goegebeur †

Armelle Guillou ‡

August 11, 2015

## 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.

**Keywords:** failure set, bias-correction, tail dependence, robustness, tail quantile process.

**AMS 2000 Subject classifications.** 62G32, 62H12, 62G20.

## 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

---

\*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).

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 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  $\ell$  is a bivariate slowly varying function, i.e.

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

and the function  $\varsigma$  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  $\eta$  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  $\eta$  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  $\eta$ , 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

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

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

Recently, several papers have addressed bias-corrected estimation of the coefficient of tail dependence  $\eta$ , but not too much attention has been paid to the estimation of the probability of extreme failure sets. Concerning  $\eta$ , 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  $(SO)$  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  $(SO)$ , 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

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

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  $\theta$ , and let  $f$  be

the true density function of the random variable under consideration. The idea is then to estimate  $\theta$  by minimizing an empirical version of  $\Delta_\alpha$  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  $\alpha$  controls the trade-off between efficiency and robustness of the MDPD estimator: the estimator becomes more efficient but less robust against outliers as  $\alpha$  gets closer to zero, whereas for increasing  $\alpha$  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

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

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  $\delta$  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  $\xi$  such that

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

for all  $x, y \geq 0$ . We assume additionally that  $\xi$  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 (SO) with*

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

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

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

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

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

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, 1998b) the classical estimators of extreme value index have larger bias for Fréchet distributions than for Pareto distributions. Model (7) 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 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

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

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  $Z_\omega/u$ , conditionally on  $Z_\omega > u$ , for large  $u$ , by the extended Pareto distribution. Formally, using the common reparametrization  $\rho = -\tau\eta$ , for  $u$  large

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

where  $H$  has the density function

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

( $\eta \in (0, 1)$ ,  $\rho < 0$ , and  $\delta_\omega(u) > \max\{-1, \eta/\rho\}$ ). Using this property, one can estimate  $\eta$  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  $(X_1, Y_1), \dots, (X_n, Y_n)$  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

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

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  $\eta$  and  $\delta_\omega$  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. Remember that the parameter  $\delta_\omega$  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. The MDPD estimator for  $\eta$  and  $\delta_\omega$  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  $\eta$  and  $\delta_\omega$  are estimated with the MDPD method. The parameter  $\rho$  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  $\rho$  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  $\rho$  (or  $\tau$ ) is a challenging problem. Goegebeur and Guillou (2013) introduced a consistent estimator for  $\tau$ , 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  $z_n \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $y_n = \tilde{\omega} z_n$  for some  $\tilde{\omega} > 0$ , i.e.

we estimate a tail probability along a ray. We have thus that  $p_n := \bar{F}_{Z_\omega}(z_n)$ , where  $\omega = (1 + \tilde{\omega})^{-1}$ . To this aim, let  $m$  be an intermediate sequence, i.e.  $m \rightarrow \infty$  as  $n \rightarrow \infty$  with  $m = o(n)$ . Assume that  $p_n$  satisfies

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

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

$$\hat{p}_n := \frac{m}{n} \bar{H} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \hat{\eta}_n, \hat{\delta}_{\omega,n}, \tilde{\rho} \right) \quad (12)$$

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

Note that, at first sight, once the minima  $Z_\omega$  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  $\tilde{Z}_{\omega,i}, i = 1, \dots, n$ , 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  $\tilde{Z}_{\omega,1}, \dots, \tilde{Z}_{\omega,n}$ , with order statistics  $\tilde{Z}_{\omega,1,n} \leq \dots \leq \tilde{Z}_{\omega,n,n}$ . In order to obtain the limiting distribution of  $\hat{p}_n$ , 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  $Q_n(t) := \tilde{Z}_{\omega,n-[nt],n}$ ,  $0 < t < n/m$ . 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  $k \rightarrow \infty$ .

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

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

with  $s_1, s_2, s_3 \leq 0$  whose joint asymptotic normality follows from Theorem A (cf. Appendix). However, in order to derive the limiting distribution of  $\hat{p}_n$  defined in (12), we need to add an additional

component,  $Q_n(1) = \tilde{Z}_{\omega, n-m, n}$ , 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  $(X_1, Y_1), \dots, (X_n, Y_n)$  be independent copies of the random vector  $(X, Y)$  which has a joint distribution satisfying (2) such that the function  $c$  given in (4) 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*

$$\sqrt{m} \begin{bmatrix} \frac{k}{n} Q_n(1) - c_\omega^\eta \\ A_{m,n}^{(1)}(s_1) - \frac{1}{1-\eta s_1} \\ A_{m,n}^{(1)}(s'_1) - \frac{1}{1-\eta s'_1} \\ A_{m,n}^{(2)}(s_2) - \frac{\eta}{(1-\eta s_2)^2} \\ A_{m,n}^{(3)}(s_3) - \frac{2\eta^2}{(1-\eta s_3)^3} \end{bmatrix} \rightsquigarrow N_5(\lambda \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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  $\Sigma$  has elements given by

$$\begin{aligned}
\sigma_{11} &:= \eta^2 c_\omega^{2\eta} \\
\sigma_{22} &:= \frac{\eta^2 s_1^2}{(1-s_1\eta)^2(1-2s_1\eta)} \\
\sigma_{33} &:= \frac{\eta^2 s_1'^2}{(1-s_1'\eta)^2(1-2s_1'\eta)} \\
\sigma_{44} &:= \eta^2 \left\{ \frac{2}{(1-2s_2\eta)^3} - \frac{1}{(1-s_2\eta)^4} \right\} \\
\sigma_{55} &:= \eta^4 \left\{ \frac{4!}{(1-2s_3\eta)^5} - \frac{4}{(1-s_3\eta)^6} \right\} \\
\sigma_{12} &= \sigma_{13} = \sigma_{14} = \sigma_{15} := 0 \\
\sigma_{23} &:= \frac{\eta^2 s_1 s_1'}{(1-s_1\eta)(1-s_1'\eta)[1-(s_1+s_1')\eta]} \\
\sigma_{24} &:= \eta \left\{ \frac{1}{[1-(s_1+s_2)\eta]^2} - \frac{1}{(1-s_1\eta)(1-s_2\eta)^2} \right\} \\
\sigma_{25} &:= 2\eta^2 \left\{ \frac{1}{[1-(s_1+s_3)\eta]^3} - \frac{1}{(1-s_1\eta)(1-s_3\eta)^3} \right\} \\
\sigma_{34} &:= \eta \left\{ \frac{1}{[1-(s_1'+s_2)\eta]^2} - \frac{1}{(1-s_1'\eta)(1-s_2\eta)^2} \right\} \\
\sigma_{35} &:= 2\eta^2 \left\{ \frac{1}{[1-(s_1'+s_3)\eta]^3} - \frac{1}{(1-s_1'\eta)(1-s_3\eta)^3} \right\} \\
\sigma_{45} &:= 2\eta^3 \left\{ \frac{3}{[1-(s_2+s_3)\eta]^4} - \frac{1}{(1-s_2\eta)^2(1-s_3\eta)^3} \right\}.
\end{aligned}$$

The proof of this Theorem 1 is an adaptation of the proof of Theorem 2 in Dutang *et al.* (2014) combined with Theorem A. 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  $U_1, \dots, U_n$  i.i.d.  $U(0, 1)$  random variables, with order statistics  $U_{1,n} \leq \dots \leq U_{n,n}$ , one has that  $V_{j,k} := U_{j,n}/U_{k+1,n}$ ,  $j = 1, \dots, k$ , 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  $\tilde{Z}_{\omega,i}$ ,  $i = 1, \dots, n$ . 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  $\left( \frac{\tilde{Z}_{\omega,n,n}}{\tilde{Z}_{\omega,n-m,n}}, \dots, \frac{\tilde{Z}_{\omega,n-m+1,n}}{\tilde{Z}_{\omega,n-m,n}} \right)$  whereas the asymptotic distribution of the first component in our Theorem 1,  $Q_n(1)$ , is totally determined by that of  $\tilde{Z}_{\omega,n-m,n}$ .

From now on, we will denote the true value of  $\eta$  and  $\rho$  by  $\eta_0$  and  $\rho_0$ , respectively. Let  $\delta_{\omega,n} := \delta_\omega(\tilde{Z}_{\omega,n-m,n})$ . In the next theorem, we establish the joint limiting distribution of  $Q_n(1)$ ,  $\hat{\eta}_m$  and  $\hat{\delta}_{\omega,n}$ , when properly normalised. The parameter  $\rho$  in the estimating equations (9) and (10) is denoted in the sequel as  $\tilde{\rho}$  and is either the true value  $\rho_0$  or a mis-specified one.

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

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

where

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

with

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

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

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

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

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

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

$$\begin{aligned}
d_{11}(\tilde{\rho}) &:= \eta_0^2 c_\omega^{2\eta_0} \\
d_{12}(\tilde{\rho}) &= d_{13}(\tilde{\rho}) = d_{14}(\tilde{\rho}) := 0 \\
d_{22}(\tilde{\rho}) &:= \frac{\alpha^2(1+\eta_0)^2}{[1+\alpha(1+\eta_0)]^2[1+2\alpha(1+\eta_0)]} \\
d_{23}(\tilde{\rho}) &:= \frac{\alpha(1+\eta_0)[\alpha(1+\eta_0)-\tilde{\rho}]}{[1+\alpha(1+\eta_0)][1-\tilde{\rho}+\alpha(1+\eta_0)][1-\tilde{\rho}+2\alpha(1+\eta_0)]} \\
d_{33}(\tilde{\rho}) &:= \frac{[\alpha(1+\eta_0)-\tilde{\rho}]^2}{[1-\tilde{\rho}+\alpha(1+\eta_0)]^2[1-2\tilde{\rho}+2\alpha(1+\eta_0)]} \\
d_{24}(\tilde{\rho}) &:= \eta_0 \left( \frac{1}{[1+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^3} \right) \\
d_{34}(\tilde{\rho}) &:= \eta_0 \left( \frac{1}{[1-\tilde{\rho}+2\alpha(1+\eta_0)]^2} - \frac{1}{[1+\alpha(1+\eta_0)]^2[1-\tilde{\rho}+\alpha(1+\eta_0)]} \right) \\
d_{44}(\tilde{\rho}) &:= \eta_0^2 \left( \frac{2}{[1+2\alpha(1+\eta_0)]^3} - \frac{1}{[1+\alpha(1+\eta_0)]^4} \right).
\end{aligned}$$

Note that if  $\tilde{\rho} = \rho_0$  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  $\hat{\eta}_n$  is asymptotically unbiased in the sense that the mean of the limiting distribution is zero, whatever the value of  $\lambda$ . Obviously, if  $\tilde{\rho}$  is not the true value, this property of unbiasedness is lost. Nevertheless, despite the loss of asymptotic unbiasedness when fixing  $\tilde{\rho}$  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  $\eta$ . 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 A, and satisfies  $\overline{W}(1/c_\omega) \sim N(0, 1)$ .

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

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

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

The result of Theorem 3 indicates that the tail probability estimator is asymptotically unbiased if one uses the correct value for  $\rho$ . We now consider the case where  $np_n/m \rightarrow 0$ , 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*

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

## 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  $\alpha$  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

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

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

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

We take two different values of  $\zeta$ :  $\zeta = 1$  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  $p_{z,\tilde{\omega}} = \mathbb{P}(X > z, Y > \tilde{\omega}z)$  is equal to 2%, 0.1% and 0.01%, see Table 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				
$p_{z,\tilde{\omega}}$	2%			0.1%	0.01%

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

The data are generated as follows. First, we simulate  $n$  pairs  $(X_i, Y_i), i = 1, \dots, n$ , 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  $\tilde{Z}_{\omega,i}$ ,  $i = 1, \dots, n$ . Finally, we minimize the empirical density power divergence  $\hat{\Delta}_\alpha$ .

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

- we simulate independently  $\check{X}_j$  and  $\check{Y}_j$ ,  $j = 1, \dots, n_0 := \lfloor n\varepsilon \rfloor$ , from a unit Fréchet distribution;
- all our above methodology is applied to our new  $n + n_0$  pairs  $(X_1, Y_1), \dots, (X_n, Y_n), (\check{X}_1 + X_{n,n}, \check{Y}_1 + Y_{n,n}), \dots, (\check{X}_{n_0} + X_{n,n}, \check{Y}_{n_0} + Y_{n,n})$ , 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  $\hat{p}_{n,\alpha}$  of  $p_{z,1} = 2\%$  and the right panel the mean squared error of the ratio  $\hat{p}_{n,\alpha}/p_{z,1}$  as a function of  $m$  for the FGM copula with  $\zeta = 1$  and different values of  $\alpha$ : 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  $\rho$  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,  $\tilde{\omega}$  is set to the value 1, i.e.  $\omega$  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  $\alpha$  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  $\hat{p}_{n,\alpha}$  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  $\hat{p}_{n,\alpha}$  on  $\hat{p}_{n,0}$  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;
- when a proper choice of  $\alpha$  is used (all considered values of  $\alpha$  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.

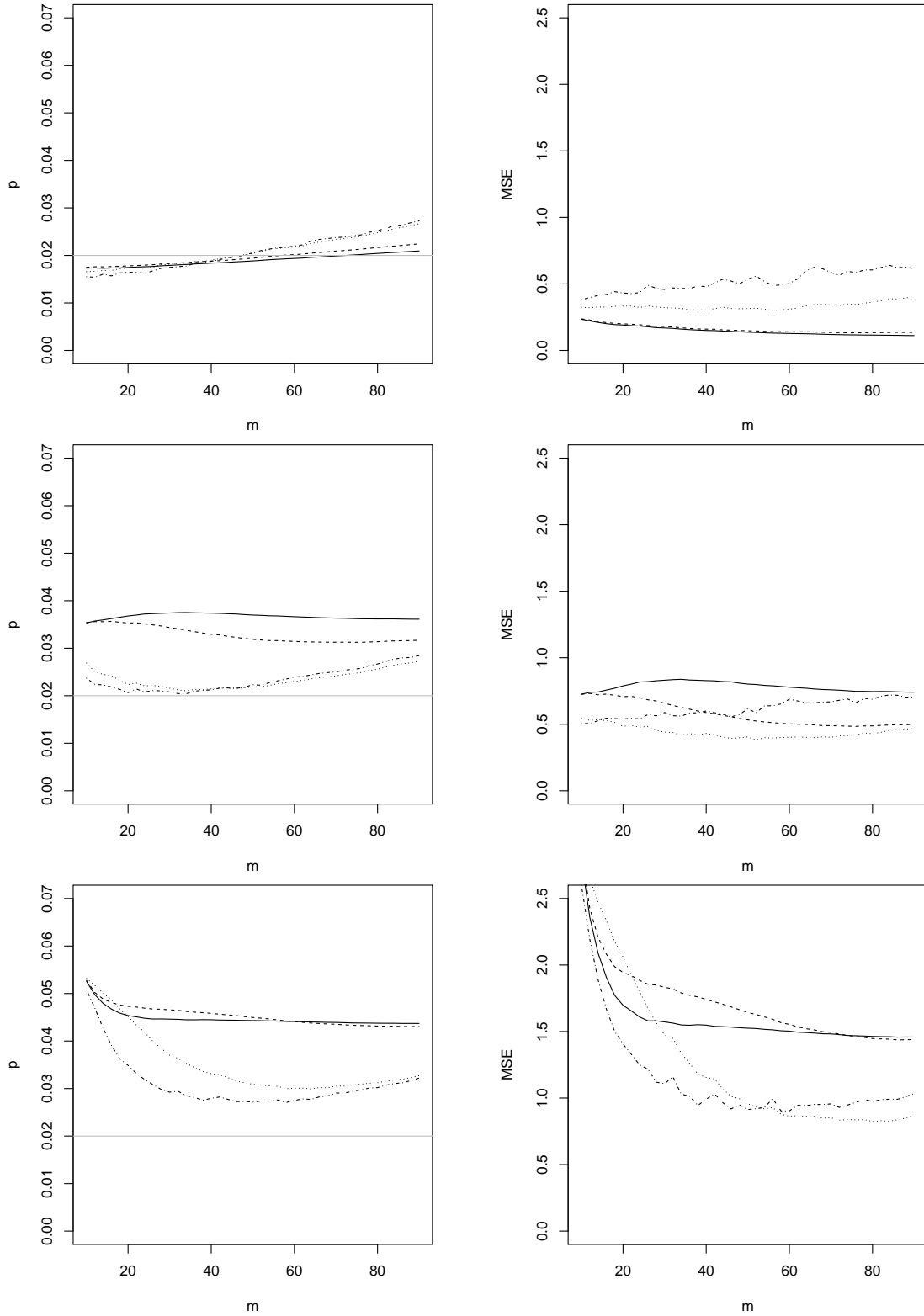


Figure 1: Farlie Gumbel Morgenstern copula with  $\zeta = 1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 100, p_{z,1} = \frac{2}{n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with  $\varepsilon = 2\%$ , and  $\varepsilon = 5\%$ . The parameter  $\rho$  is mis-specified to the value  $-1$ . The horizontal reference line in the left panels of the figure corresponds to the true value of  $p_{z,1}$ .

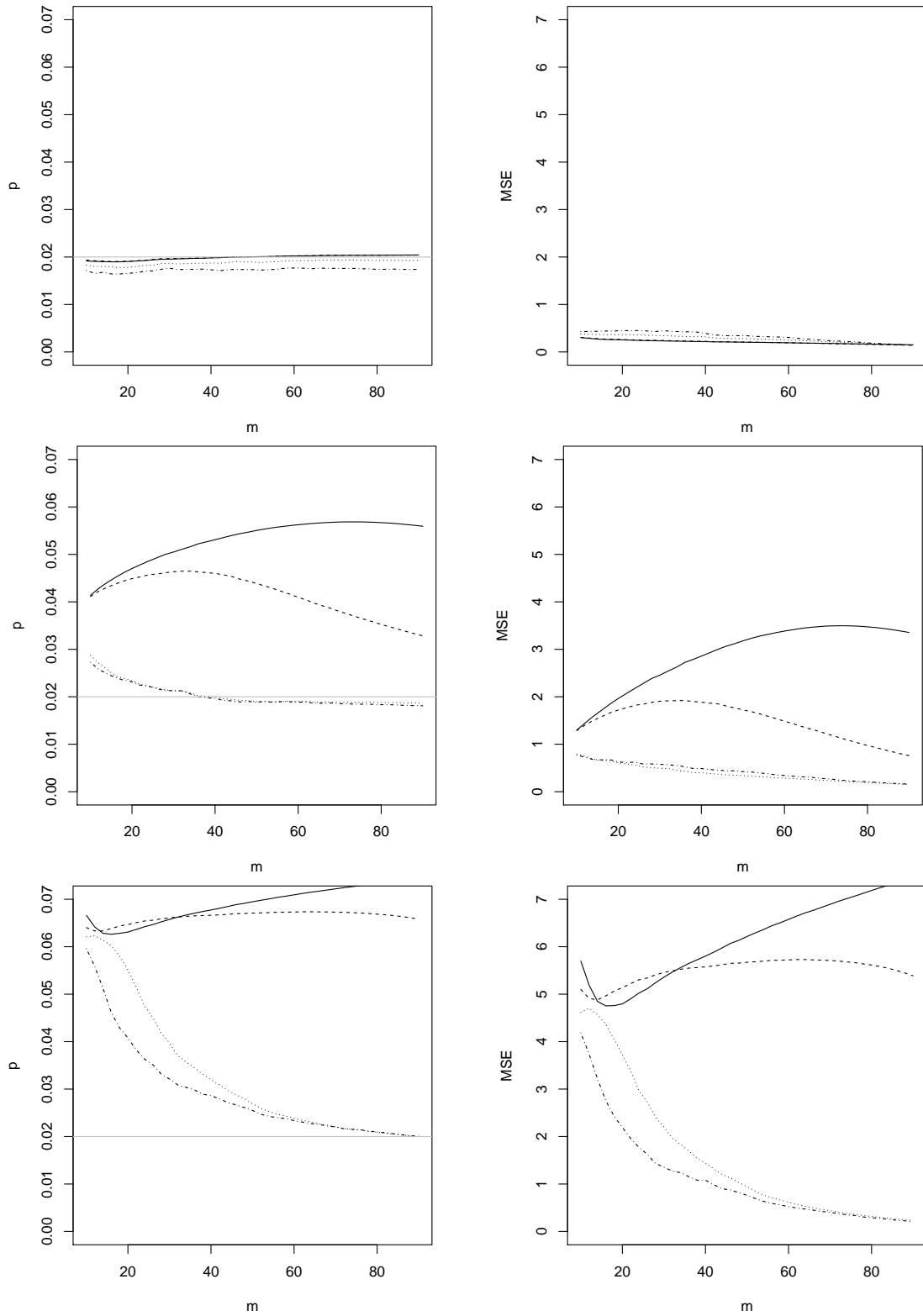


Figure 2: Farlie Gumbel Morgenstern copula with  $\zeta = -1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 100, p_{z,1} = \frac{2}{n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with  $\varepsilon = 2\%$ , and  $\varepsilon = 5\%$ . The parameter  $\rho$  is mis-specified to the value  $-1$ . The horizontal reference line in the left panels of the figure corresponds to the true value of  $p_{z,1}$ .

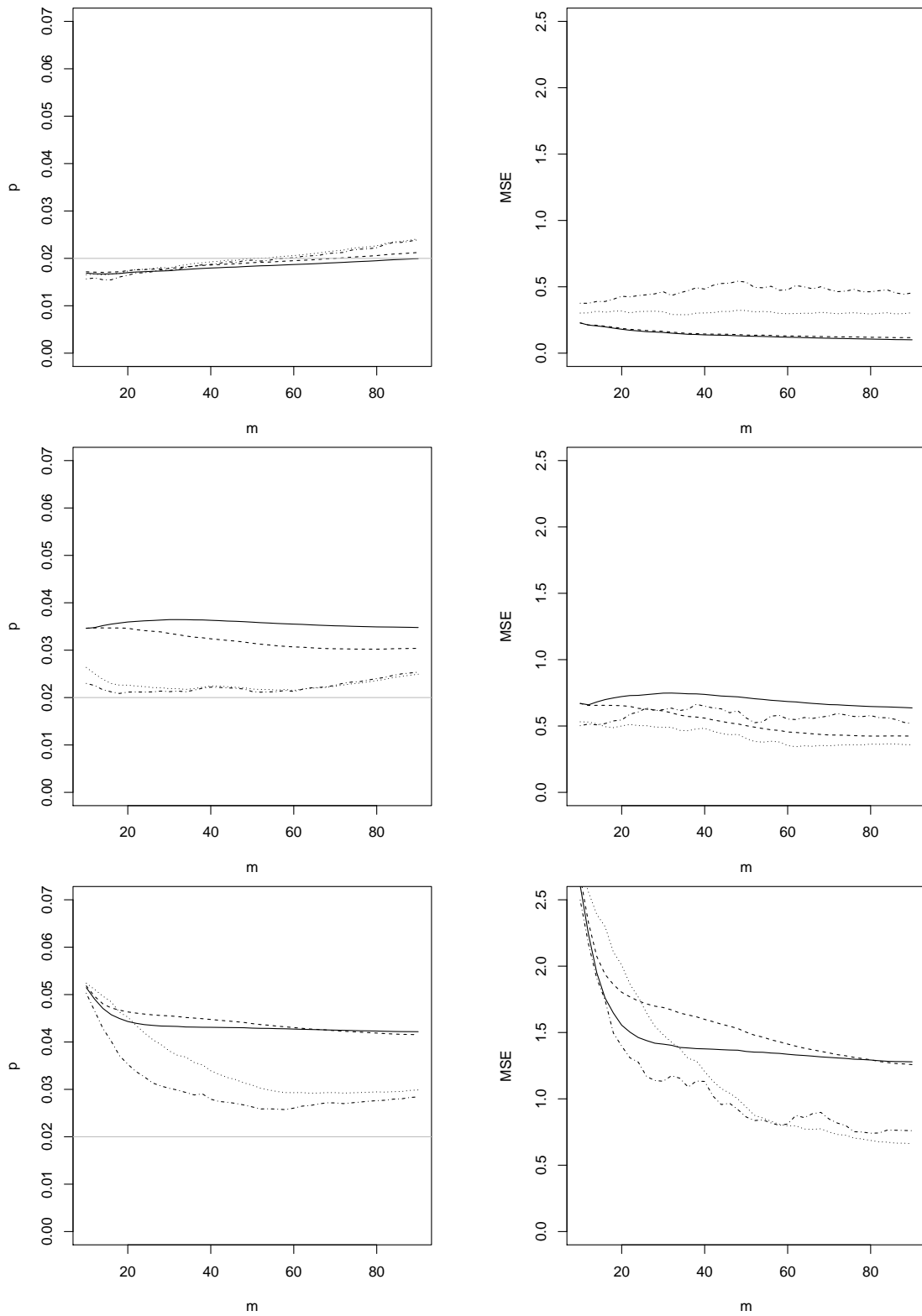


Figure 3: Frank copula: mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 100$ ,  $p_{z,1} = \frac{2}{n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with  $\varepsilon = 2\%$ , and  $\varepsilon = 5\%$ . The parameter  $\rho$  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 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 to 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  $\hat{p}_{n,\alpha}$  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:

- we simulate 1024 samples of size  $n$  (100 and 1000) and take a percentage of contamination  $\varepsilon = 0, 1\%, \dots, 10\%$ ;
- for each value of  $\varepsilon$  and each sample size  $n$ , we compute our estimator  $\hat{p}_{n,\alpha}$  of  $p_{z,1} = 2\%$  and the mean squared error of the ratio  $\hat{p}_{n,\alpha}/p_{z,1}$  as a function of  $m$ ;
- for a fixed value of  $m$ , the breakdown point is defined as the smallest value of  $\varepsilon$  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  $\alpha$  (0, 0.1, 0.5 and 1). For both sample sizes and whatever the distribution, the higher the value of  $\alpha$ , 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  $\hat{p}_{n,\alpha}$ , 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 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  $p_n \rightarrow 0$  when  $n \rightarrow \infty$ , together with some additional assumptions on the involved sequences.

Furthermore, we consider the estimation of our probability  $p_{z,\tilde{\omega}}$ , but this time outside the diagonal, that is for  $\tilde{\omega} \neq 1$ . Figure 7 illustrates the behaviour of our estimator  $\hat{p}_{n,\alpha}$  in the uncontaminated case

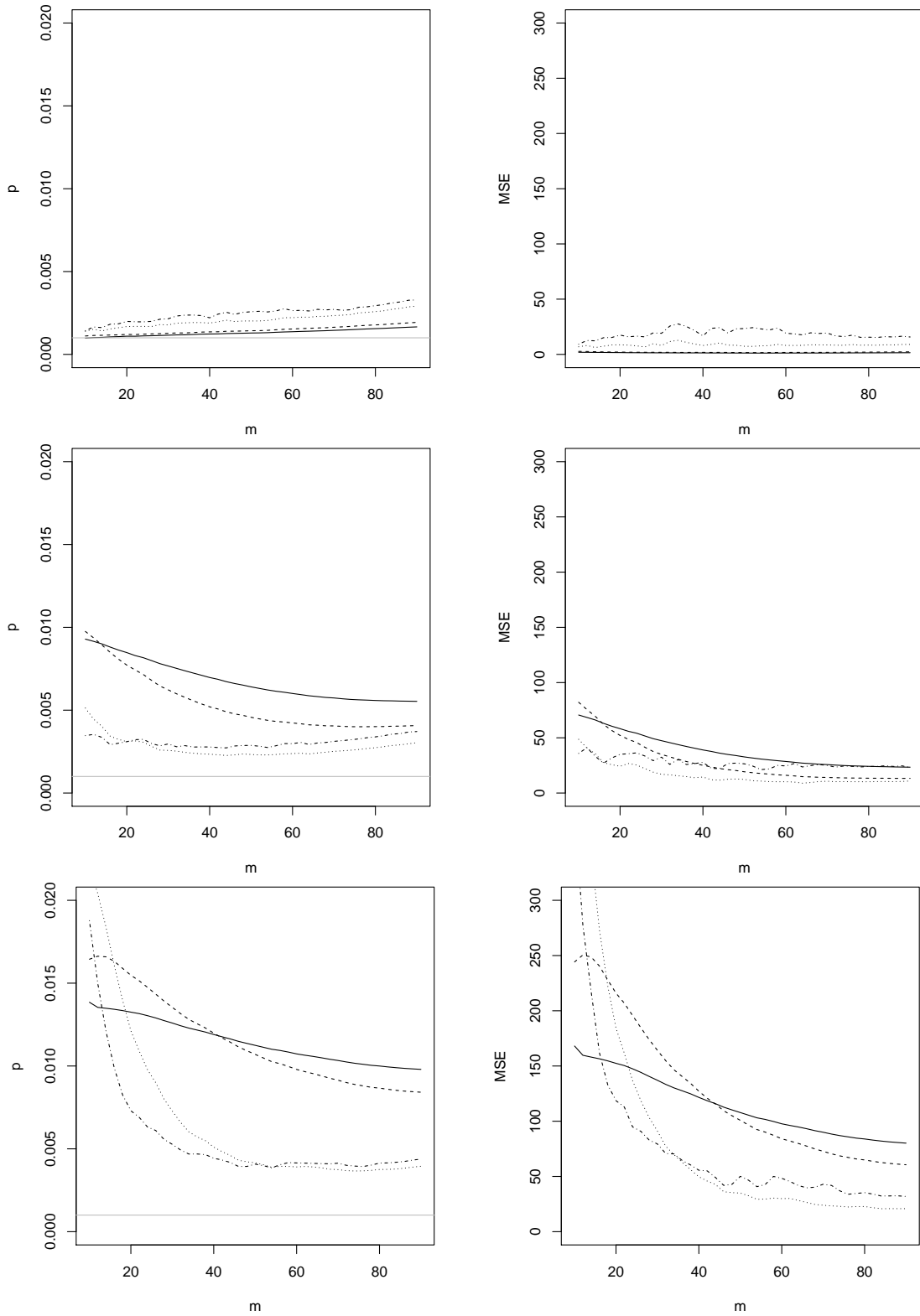


Figure 4: Farlie Gumbel Morgenstern copula with  $\zeta = 1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 100, p_{z,1} = \frac{1}{10n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with  $\varepsilon = 2\%$ , and  $\varepsilon = 5\%$ . The parameter  $\rho$  is mis-specified to the value  $-1$ . The horizontal reference line in the left panels of the figure corresponds to the true value of  $p_{z,1}$ .

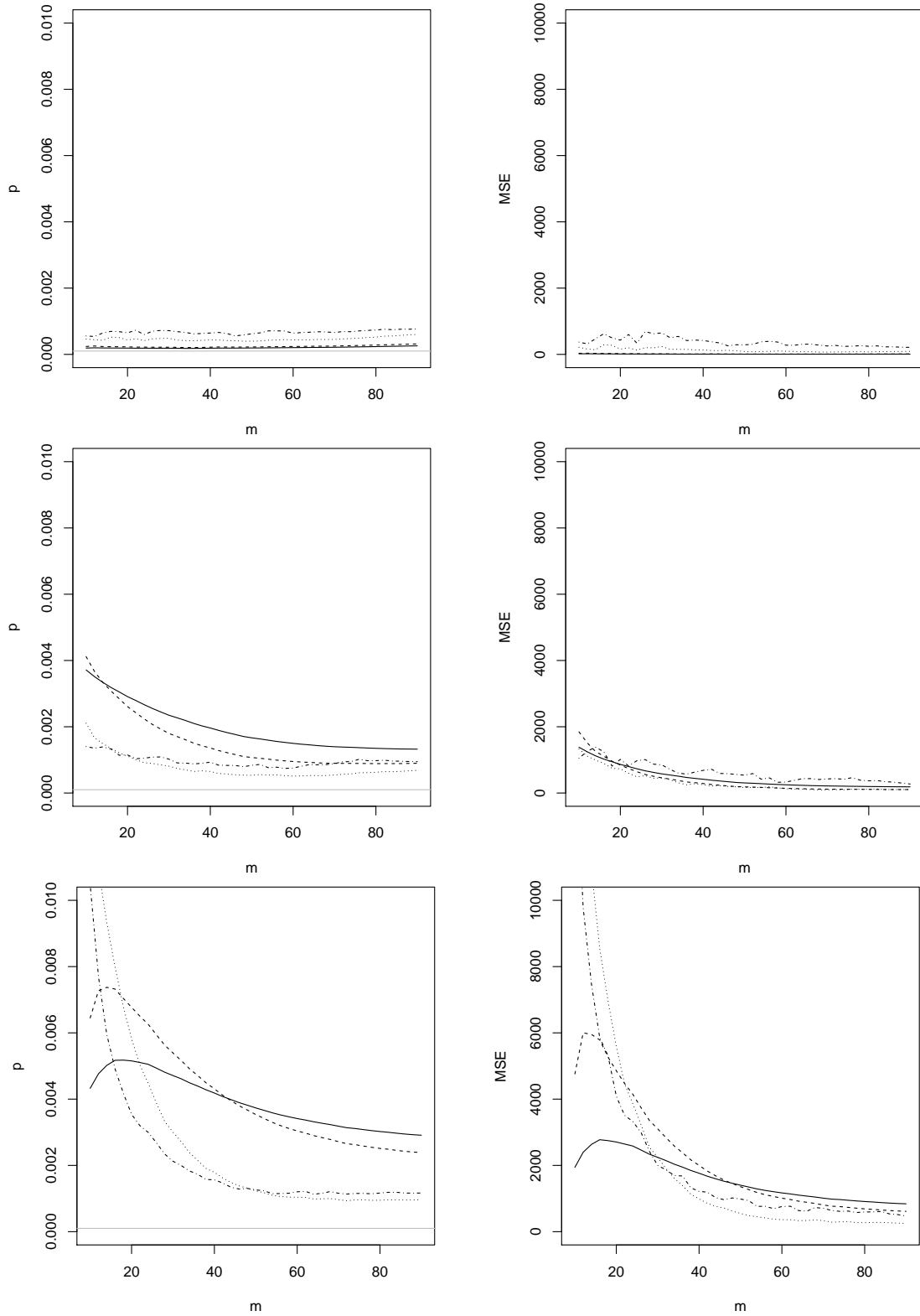


Figure 5: Farlie Gumbel Morgenstern copula with  $\zeta = 1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 100, p_{z,1} = \frac{1}{100n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with  $\varepsilon = 2\%$ , and  $\varepsilon = 5\%$ . The parameter  $\rho$  is mis-specified to the value  $-1$ . The horizontal reference line in the left panels of the figure corresponds to the true value of  $p_{z,1}$ .

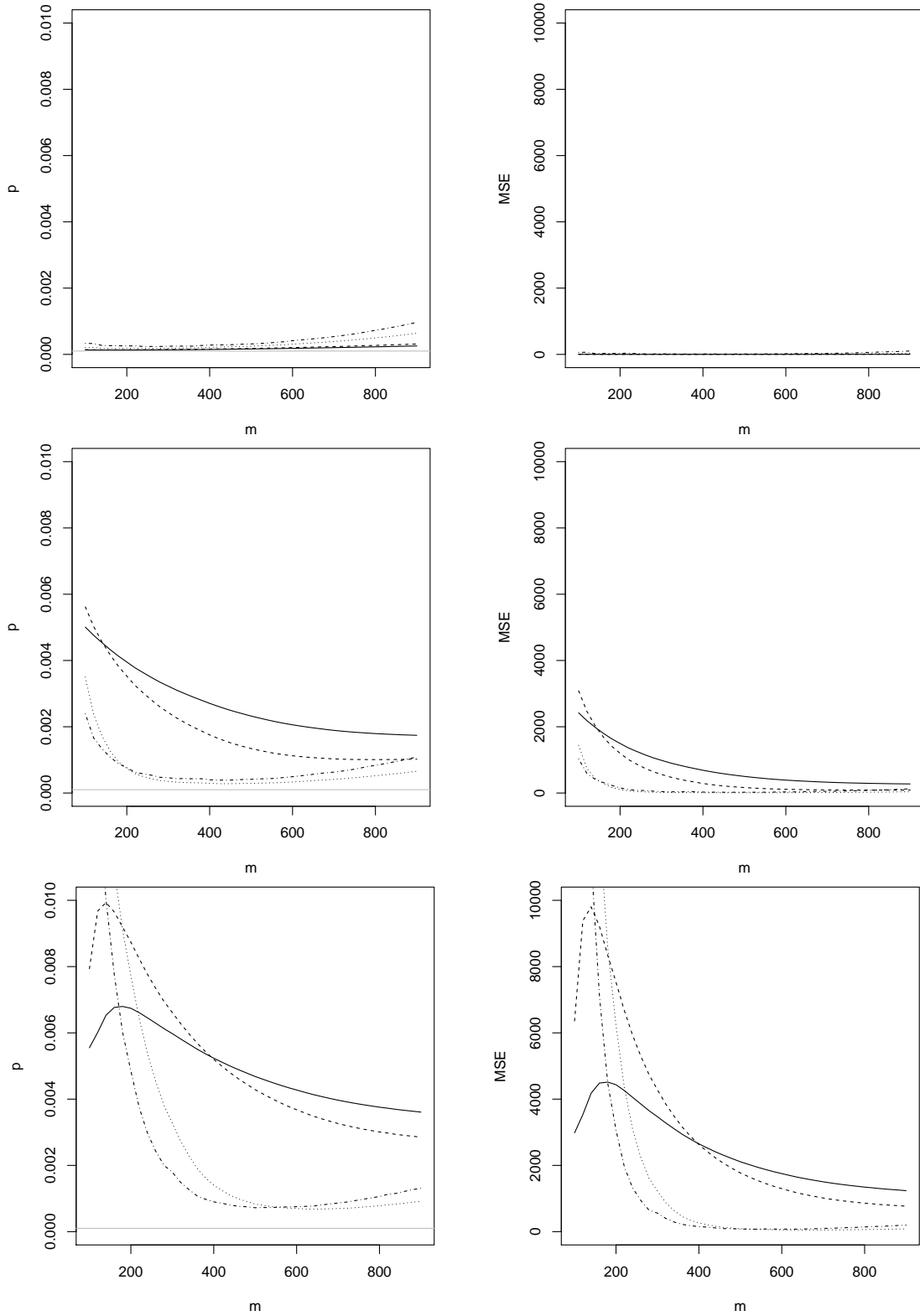


Figure 6: Farlie Gumbel Morgenstern copula with  $\zeta = 1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 1000$ ,  $p_{z,1} = \frac{1}{10n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations; from the top to the bottom: no contamination, contamination with  $\varepsilon = 2\%$ , and  $\varepsilon = 5\%$ . The parameter  $\rho$  is mis-specified to the value  $-1$ . The horizontal reference line in the left panels of the figure corresponds to the true value of  $p_{z,1}$ .

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

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

	$m$	100	200	300	400	500	600	700	800	900
FGM $\zeta = 1$	$\alpha = 0$	3	3	2	2	2	2	2	2	2
	$\alpha = 0.1$	3	3	3	3	3	3	3	3	3
	$\alpha = 0.5$	3	4	5	6	7	7	8	8	8
	$\alpha = 1$	3	5	6	7	8	9	9	8	6
FGM $\zeta = -1$	$\alpha = 0$	2	1	1	1	1	1	1	1	1
	$\alpha = 0.1$	2	2	2	2	2	2	2	2	3
	$\alpha = 0.5$	3	4	5	6	7	8	10	> 10	> 10
	$\alpha = 1$	3	4	5	6	7	9	> 10	> 10	> 10
Frank	$\alpha = 0$	3	3	2	2	2	2	2	2	2
	$\alpha = 0.1$	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 3: Values of the breakdown points, respectively for FGM  $\zeta = 1, -1$ , and Frank ( $n = 1000$ ).

for the FGM copula with  $\zeta = 1$ , the same values of  $\alpha$ , and two values of  $\omega$ : 0.45 (first row) and 0.55 (second row). In both cases, we can observe the almost best behaviour of  $\hat{p}_{n,0}$  compared to  $\hat{p}_{n,\alpha}$ ,  $\alpha > 0$ , 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  $\hat{p}_{n,\alpha}$ , we have to compute

$m$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$
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

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

the expectations of  $\Gamma$  and  $\Delta$  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  $\omega$ . As the sense of variation of the function  $\omega \mapsto \xi_\omega c_\omega^{\rho_0}$  depends if  $\omega$  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  $\rho$  is not crucial, we show in Figure 8 the behaviour of our estimator  $\hat{p}_{n,\alpha}$  for the FGM copula with  $\zeta = 1$  in the uncontaminated framework in case  $\rho$  is replaced by the true  $\rho_0 = -1/2$ . The same values of  $\alpha$  have been used and  $\omega$  is set to  $1/2$ . This figure is indeed very similar to the first row of Figure 1 where  $\rho$  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 <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  $\tilde{Z}_{\frac{1}{2},i}$ ,  $i = 1, \dots, n$ , is depicted in Figure 9 (c). The three points labeled with crosses correspond to possible outliers in the dataset. Their values 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  $\tilde{X}$  and  $\tilde{Y}$  the random variables on the unit Pareto scale, as obtained from the inverse probability integral transform.

Now, in order to detect if these three points listed in Table 5 are outliers or not, we consider the

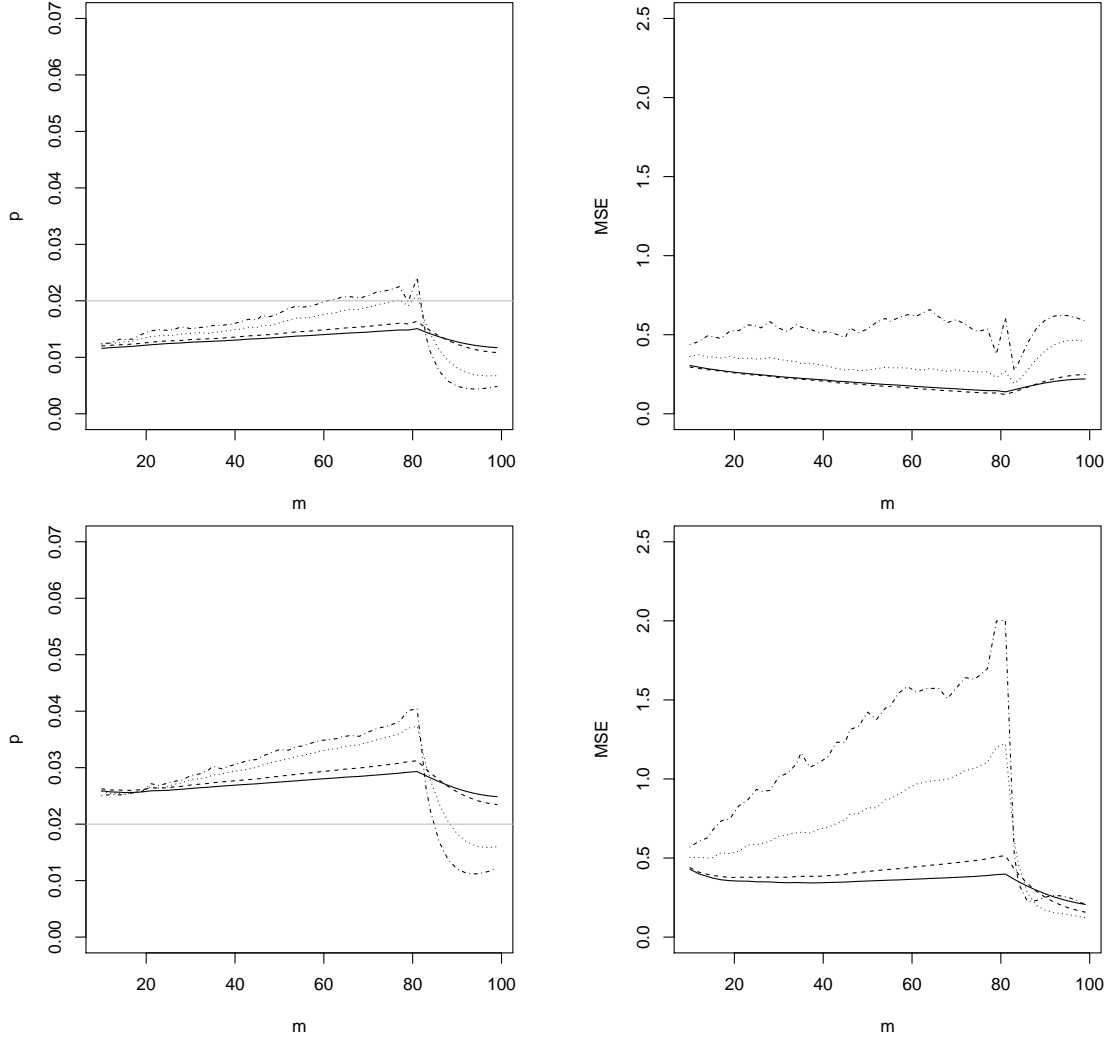


Figure 7: Farlie Gumbel Morgenstern copula with  $\zeta = 1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,\tilde{\omega}}$  for  $n = 100, p_{z,1} = \frac{2}{n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted 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  $\omega = (1 + \tilde{\omega})^{-1}$ . The parameter  $\rho$  is mis-specified to the value  $-1$ . The horizontal reference line in the left panels of the figure corresponds to the true value of  $p_{z,\tilde{\omega}}$ .

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

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

estimation of the probability  $\mathbb{P}(\tilde{X} > 10, \tilde{Y} > 10)$ , with respect to the presence or the removal of these three points. The choices  $z = 10$  and  $\tilde{\omega} = 1$  can be justified by the fact that  $\mathbb{P}(\tilde{X} > 10, \tilde{Y} > 10)$  is a possible indicator of bankruptcy since it is unlikely that an insurer can deal with both high losses

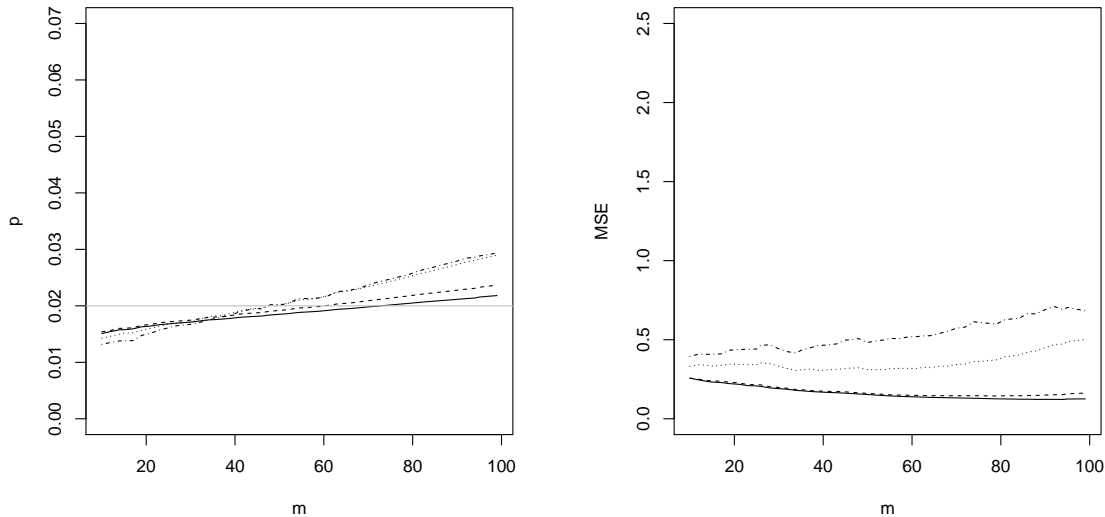


Figure 8: Farlie Gumbel Morgenstern copula with  $\zeta = 1$ : mean (left) of  $\hat{p}_{n,\alpha}$ , MSE (right) of  $\hat{p}_{n,\alpha}/p_{z,1}$  for  $n = 100, p_{z,1} = \frac{2}{n}$  and different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line), based on 1024 simulations in the uncontaminated case. The parameter  $\rho$  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  $p_{z,1}$ .

and high expenses. Figure 10 shows the estimators  $\hat{p}_{n,\alpha}$  as a function of  $m$  for different values of  $\alpha$ : 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  $i \in N := \{1, \dots, n\}$  to estimate  $\hat{p}_{n,\alpha}$ , 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  $\alpha$  is. This corroborates our feeling that the three points are outliers. Indeed, as observed in Figure 10 (a) the estimator  $\hat{p}_{n,\alpha}$  differs considerably depending on the value of  $\alpha$  which is expected in the presence of outliers. As in the simulation study, the estimator  $\hat{p}_{n,0.5}$  remains stable both in the situation with and without outliers.

Finally, Figure 11 shows the median over  $m = 50, \dots, 80$  of our probability estimates  $\hat{p}_{n,\alpha}$  of  $\mathbb{P}(\tilde{X} > z, \tilde{Y} > z)$  for increasing values of  $z$  and different values of  $\alpha$ . 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  $\hat{p}_{n,0.5}$  exhibits a more stable pattern than  $\hat{p}_{n,\alpha}$  for other values of  $\alpha$  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.



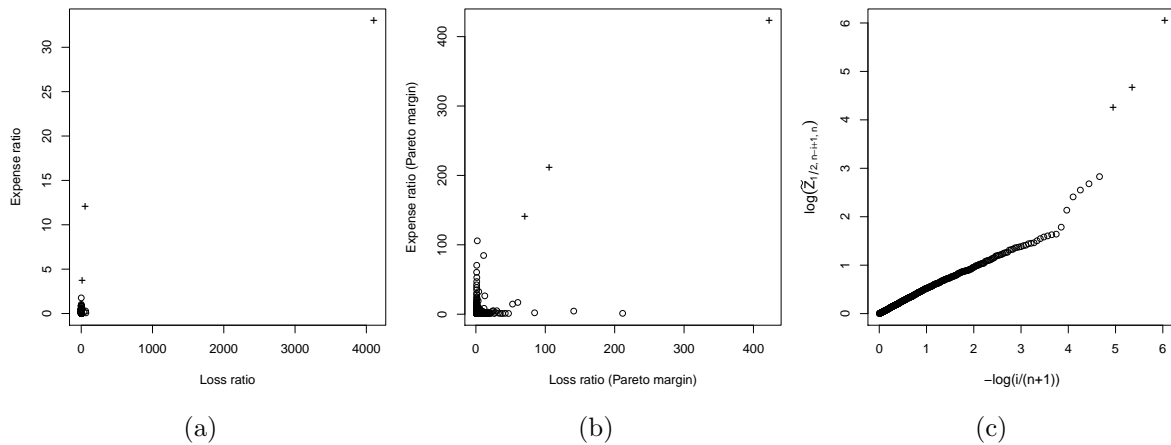


Figure 9: Loss/expense ratios dataset: (a) Scatterplot of the original data; (b) Scatterplot of the data transformed into unit Pareto margins; (c) Pareto quantile plot of the  $\tilde{Z}_{\frac{1}{2}, i}, i = 1, \dots, n$ .

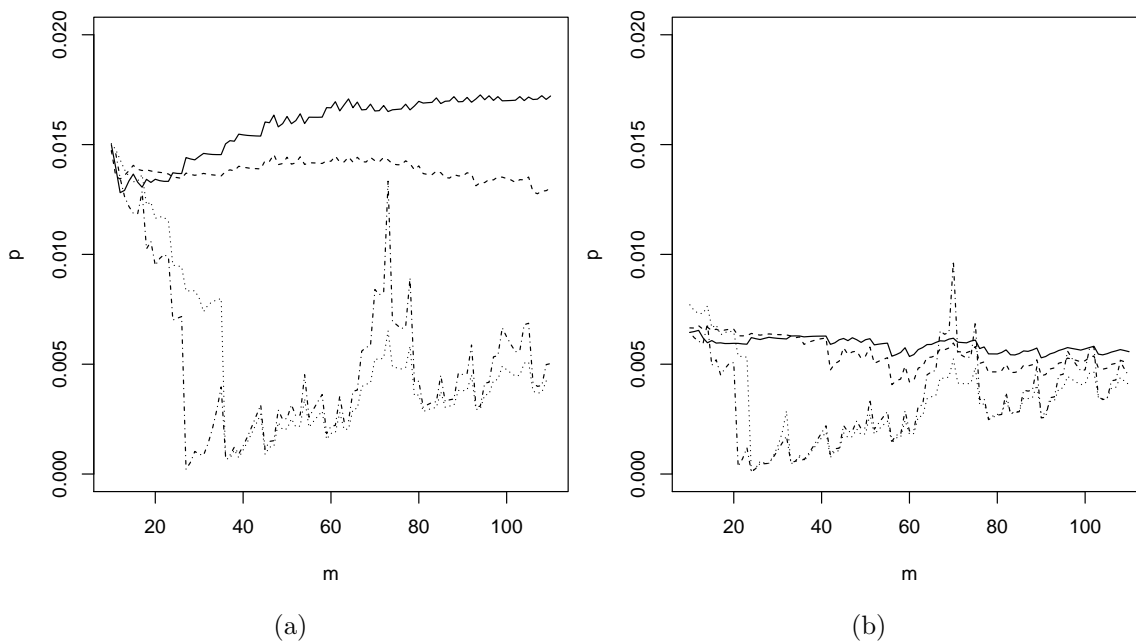


Figure 10: Loss/expense ratios dataset: Estimator  $\hat{p}_{n, \alpha}$  of  $\mathbb{P}(\tilde{X} > 10, \tilde{Y} > 10)$  for different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line). (a) Whole dataset; (b) Dataset without the three points.

## Acknowledgement

This work was supported by a research grant (VKR023480) from VILLUM FONDEN and a PICS project (PICS-6416). The authors thank the editor and two referees for their helpful comments which led to significant improvements of this article.

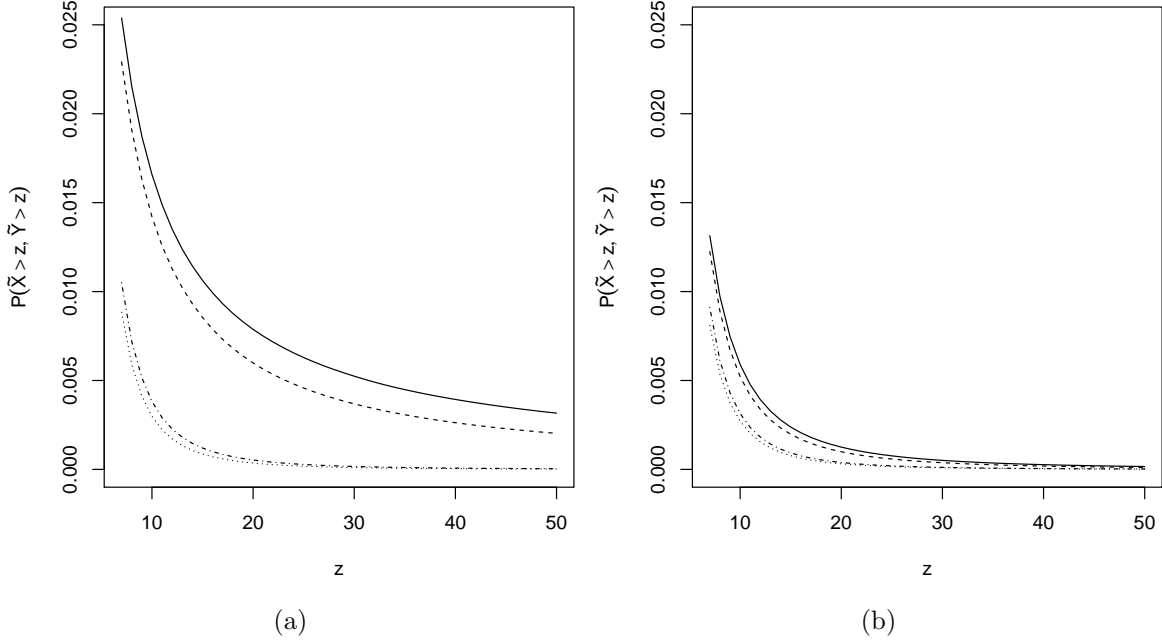


Figure 11: Median of the estimator  $\hat{p}_{n,\alpha}$  of  $\mathbb{P}(\tilde{X} > z, \tilde{Y} > z)$  over  $m = 50, \dots, 80$  for different values of  $\alpha$ : 0 (solid line), 0.1 (dashed line), 0.5 (dotted line) and 1 (dash-dotted line). (a) Whole dataset; (b) Dataset without the three points.

## Appendix

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

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

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

**Proof of Theorem 2** We follow the lines of proof of Theorem 2 in Dierckx *et al.* (2013). We give here only the adjustments needed. Define

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

We can easily infer from Theorem 1 that  $\overline{A}_{m,n}(\tilde{\rho}) \rightsquigarrow A(\tilde{\rho}) \sim N_4(\lambda \tilde{\boldsymbol{\mu}}, \mathbb{D}(\tilde{\rho}))$ . From the proof of Theorem

2 in Dierckx *et al.* (2013), we deduce that

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

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

$$\begin{aligned} \ln \frac{\hat{p}_n}{p_n} &= \ln \frac{\frac{m}{n} \bar{H} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \hat{\eta}_n, \hat{\delta}_{\omega,n}, \tilde{\rho} \right)}{\bar{F}_{Z_\omega}(z_n)} \\ &= -\ln \left( \frac{n}{m} \bar{F}_{Z_\omega}(\tilde{Z}_{\omega,n-m,n}) \right) + \left[ \ln \bar{H} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \eta_0, \delta_{\omega,n}, \rho_0 \right) - \ln \frac{\bar{F}_{Z_\omega} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}} \tilde{Z}_{\omega,n-m,n} \right)}{\bar{F}_{Z_\omega}(\tilde{Z}_{\omega,n-m,n})} \right] \\ &\quad + \left[ \ln \bar{H} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \hat{\eta}_n, \hat{\delta}_{\omega,n}, \tilde{\rho} \right) - \ln \bar{H} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \eta_0, \delta_{\omega,n}, \rho_0 \right) \right] \\ &=: T_1 + T_2 + T_3. \end{aligned}$$

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

**Concerning  $T_1$ .** This term can be rewritten as

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

From Theorem A, we easily infer that

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

Thus, if we denote by  $Q_{Z_\omega}$  the quantile function of the variable  $Z_\omega$  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

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

Recall now that

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

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

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

we deduce that

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

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

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

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,  $T_{12}$ :

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

by (15) and (16).

Thus

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

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

$$\begin{aligned}
T_{13} &\stackrel{d}{=} -\ln \left( \frac{n}{m} \bar{F}_{Z_{\omega}} (Q_{Z_{\omega}}(U_{n-m,n})) \right) \\
&\stackrel{d}{=} -\left( \frac{n}{m} U_{m+1,n} - 1 \right) (1 + o_{\mathbb{P}}(1)) \\
&= -\frac{\Omega}{\sqrt{m}} + o_{\mathbb{P}} \left( \frac{1}{\sqrt{m}} \right)
\end{aligned}$$

by (17). Consequently

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

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

$$\begin{aligned}
\frac{n}{m} \bar{F}_{Z_{\omega}}(z_n) &= \frac{np_n}{m} \longrightarrow \beta, \\
\frac{n}{m} \bar{F}_{Z_{\omega}}(\tilde{Z}_{\omega,n-m,n}) &= \frac{n}{m} C_{\omega} \tilde{Z}_{\omega,n-m,n}^{-1/\eta_0} (1 + o_{\mathbb{P}}(1)) = \frac{n}{m} \frac{C_{\omega}}{c_{\omega}} \left( q^{\leftarrow} \left( \frac{m}{n} \right) \right)^{1/\eta_0} (1 + o_{\mathbb{P}}(1)) = 1 + o_{\mathbb{P}}(1),
\end{aligned}$$

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

$$\begin{aligned}
T_2 &= \frac{\bar{H} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}}; \eta_0, \delta_{\omega,n}, \rho_0 \right) - \frac{\bar{F}_{Z_{\omega}} \left( \frac{z_n}{\tilde{Z}_{\omega,n-m,n}} \tilde{Z}_{\omega,n-m,n} \right)}{\bar{F}_{Z_{\omega}}(\tilde{Z}_{\omega,n-m,n})}}{\frac{\bar{F}_{Z_{\omega}}(z_n)}{\bar{F}_{Z_{\omega}}(\tilde{Z}_{\omega,n-m,n})}} (1 + o_{\mathbb{P}}(1)) \\
&= o_{\mathbb{P}}(\delta_{\omega,n}) \\
&= o_{\mathbb{P}} \left( \frac{1}{\sqrt{m}} \right).
\end{aligned} \quad (19)$$

**Concerning  $T_3$ .** Remark that

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

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

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

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

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

For  $T_4$ , note that

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

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

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

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  $\hat{\delta}_{\omega,n} = O_{\mathbb{P}}(1/\sqrt{m})$  and  $z_n/\tilde{Z}_{\omega,n-m,n} \rightarrow \infty$ ). Collecting all the terms establishes the result.

## References

- [1] 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.
- [2] Beirlant, J., Dierckx, G., Guillou, A., 2011. Bias-reduced estimators for bivariate tail modelling. *Insurance: Mathematics and Economics*, 49, 18–26.
- [3] 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.
- [4] 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.
- [5] Charpentier, A., Juri, A., 2006. Limiting dependence structures for tail events, with applications to credit derivatives. *Journal of Applied Probability*, 43, 563–586.
- [6] Ciuperca, G., Mercadier, C., 2010. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13, 55–87.
- [7] de Haan, L., de Ronde, J., 1998. Sea and wind: multivariate extremes at work. *Extremes*, 1, 7–45.
- [8] 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.
- [9] 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.
- [10] Dierckx, G., Goegebeur, Y., Guillou, A., 2014. Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. *Test*, 23, 330–355.
- [11] Draisma, G., Drees, H., Ferreira, A., de Haan, L., 2004. Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10, 251–280.
- [12] Drees, H., 1998a. On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25, 187–210.
- [13] Drees, H., 1998b. A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference*, 66, 95–112.
- [14] Dutang, C., Goegebeur, Y., Guillou, A., 2014. Robust and bias-corrected estimation of the coefficient of tail dependence. *Insurance: Mathematics and Economics*, 57, 46–57.

- [15] Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, 27, 760–781.
- [16] 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.
- [17] 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.
- [18] Goegebeur, Y., Guillou, A., 2013. Asymptotically unbiased estimation of the coefficient of tail dependence. *Scandinavian Journal of Statistics*, 40, 174–189.
- [19] Gomes, M.I., Martins, M.J., 2004. Bias reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference*, 124, 361–378.
- [20] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.
- [21] Kim, M., Lee, S., 2008. Estimation of a tail index based on minimum density power divergence. *Journal of Multivariate Analysis*, 99, 2453–2471.
- [22] Ledford, A.W., Tawn, J.A., 1997. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society Series B*, 59, 475–499.
- [23] Peng, L., 1999. Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters*, 43, 399–409.
- [24] Pickands, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.
- [25] Ramos, A., Ledford, A.W., 2009. A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society Series B*, 71, 219–241.
- [26] Weissman, I., 1978. Estimation of parameters and larger quantiles based on the  $k$  largest observations. *Journal of the American Statistical Association*, 73, 812–815.