



HAL
open science

Semi-parametric estimation of multivariate extreme expectiles

Nicholas Beck, Elena Di Bernardino, Mélina Mailhot

► **To cite this version:**

Nicholas Beck, Elena Di Bernardino, Mélina Mailhot. Semi-parametric estimation of multivariate extreme expectiles. *Journal of Multivariate Analysis*, 2021, 184, 10.1016/j.jmva.2021.104758. hal-03347188

HAL Id: hal-03347188

<https://hal.science/hal-03347188>

Submitted on 17 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semi-parametric Estimation of Multivariate Extreme Expectiles

Nicholas Beck^{a,*}, Elena Di Bernardino^b, Mélina Mailhot^c

^a*Department of Decision Sciences, HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal (QC) Canada H3T 2A7*

^b*Laboratoire J.A. Dieudonné, UMR CNRS 7351, Université Côte d'Azur, Parc Valrose, 06108 Nice, Cedex 2, France*

^c*Department of Mathematics and Statistics, Concordia University, 1455 De Maisonneuve Blvd. W., Montréal (QC) Canada H3G 1M8*

Abstract

This paper focuses on semi-parametric estimation of multivariate expectiles for extreme levels of risk. Multivariate expectiles and their extremes have been the focus of plentiful research in recent years. In particular, it has been noted that due to the difficulty in estimating these values for elevated levels of risk, an alternative formulation of the underlying optimization problem would be necessary. However, in such a scenario, estimators have only been provided for the limiting cases of tail dependence: independence and comonotonicity. In this paper, we extend the estimation of multivariate extreme expectiles (MEEs) by providing a consistent estimation scheme for random vectors with any arbitrary dependence structure. Specifically, we show that if the upper tail dependence function, tail index, and tail ratio can be consistently estimated, then one would be able to accurately estimate MEEs. The finite-sample performance of this methodology is illustrated using both simulated and real data.

Keywords: Dependence, extreme value theory, multivariate risk measures, optimization, semi-parametric estimation
2010 MSC: 60G70, 62H12, 90C53, 91G70

1. Introduction

1.1. Multivariate risk measures

Risk measures are viewed as a crucial tool in a multitude of fields relating to mathematics and statistics. Insurance, finance and economics must all use these measures in evaluating their prospective positions in the face of uncertain outcomes. Moreover, these measures are often required by regulatory organizations so that companies operate in good faith of the interests of stakeholders. To this end, it is easy to see why risk measures are constantly evolving. There have been numerous developments in this field, whether it be through establishing ideal properties [3, 21, 45, 46], extensions of univariate measures to higher dimension [11–13, 22], estimating these measures non-parametrically [4, 19], or even the development of new measures [31, 32, 35, 47].

Recently, the concept of elicibility has been developed for risk measures. Gneiting [28] defines elicibility as the ability to express a risk measure in the form of an optimization problem. Specifically, a risk measure T_α of a random variable (r.v.) X is considered elicitable if it takes the form

$$T_\alpha(X) = \arg \min_{x \in \mathbb{R}} \mathbb{E} \{S(x, \alpha)\},$$

*Corresponding author. Email address: nicholas.beck@hec.ca

for a risk level $\alpha \in (0, 1)$, where S is the score function associated to the risk measure T_α . Elicitability is a desirable property as it allows for a natural backtesting methodology; see Bellini and Bignozzi [5] for more information. Two of the most well-known elicitable risk measures are value-at-risk (VaR) and expectiles. From the usual definition in the univariate setting, VaR is the smallest such value a r.v. can take for which the probability of failing to exceed it is at least α . Thus, at level α , VaR is classically written $\text{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$ but can be written in an elicitable form as

$$\text{VaR}_\alpha(X) = \arg \min_{x \in \mathbb{R}} \mathbb{E}\{\alpha(X - x)_+ + (1 - \alpha)(X - x)_-\},$$

where $x_+ = \max\{0, x\}$ and $x_- = \max\{0, -x\}$. While elicitable, it is well documented that VaR is not a coherent risk measure in the sense of Artzner et al. [3] as it does not respect the crucial sub-additivity property. The second measure, and the focus of this paper, is the α -level expectile. Expectiles $e_\alpha(X)$ have been introduced by Newey and Powell [43] as the minimizers of an asymmetric quadratic loss. Indeed, at level $\alpha \in (0, 1)$ the expectile of a r.v. X is defined by

$$e_\alpha(X) = \arg \min_{x \in \mathbb{R}} \mathbb{E}\{\alpha(X - x)_+^2 + (1 - \alpha)(X - x)_-^2\}. \quad (1)$$

Note that the uniqueness of the minimum in (1) is guaranteed by the strict convexity of the scoring function. When $\alpha = 1/2$, it is well known that $e_\alpha(X) = \mathbb{E}[X]$, thus expectiles can be seen as an asymmetric generalization of the mean.

20 The term ‘‘expectiles’’ has probably been suggested as a combination of ‘‘expectation’’ and ‘‘quantiles’’. For $\alpha < 1/2$, expectiles are super-additive and therefore not coherent. However, counter to VaR, expectiles are coherent when $\alpha \geq 0.5$; see Ziegel [56]. As risk managers are often interested in elevated levels of risk α , this is quite advantageous.

Thus, the potential coherence of expectiles makes them an alternative risk measure to the VaR for practical applications. It should be noted that, prior to the recent discussions on expectiles, a third measure of risk had been discussed as a potential replacement for VaR in practical applications: expected shortfall (ES) [1, 52, 53]. The expected shortfall at risk level α for a r.v. X is defined by

$$\text{ES}_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_u(X) du,$$

which can be interpreted as the expected value of X given that it surpasses $\text{VaR}_\alpha(X)$. While expected shortfall is a coherent risk measure [2], it is not elicitable. Given that expectiles can be both elicitable and coherent, they present themselves as a potential improvement over both VaR and ES. In fact, for $\alpha > 1/2$, expectiles are the only risk measure that is coherent, elicitable and law-invariant [56].

Expectiles are uniquely identified by the first-order condition,

$$\alpha \mathbb{E}[\{X - e_\alpha(X)\}_+] = (1 - \alpha) \mathbb{E}[\{X - e_\alpha(X)\}_-].$$

The above equation can also be written as

$$\frac{1 - \alpha}{\alpha} = \frac{\mathbb{E}[\{X - e_\alpha(X)\}_+]}{\mathbb{E}[\{X - e_\alpha(X)\}_-]}.$$

This makes the economic interpretation of expectiles as risk measures clearer. Expectiles can be seen as the value of X that provides a profits/loss ratio of $(1 - \alpha)/\alpha$. A comparison between expectiles and more common financial risk measures, including VaR and ES can be found in Bellini and Di Bernardino [6]. Note that both expectiles and VaR fall into the family of generalized quantiles, as described in Bellini et al. [7], defined by

$$q_\alpha(X) = \arg \min_{x \in \mathbb{R}} (\alpha \mathbb{E}[\Phi_1\{(X - x)_+\}] + (1 - \alpha) \mathbb{E}[\Phi_2\{(X - x)_-\}]),$$

30 where $\Phi_1, \Phi_2 : [0, \infty) \mapsto [0, \infty)$ are strictly increasing convex functions satisfying $\Phi_i(0) = 0$ and $\Phi_i(1) = 1$ for $i \in \{1, 2\}$.

While univariate risk measures are well researched and developed, they limit one's ability for sound inference in real life situations as they often ignore any potential dependence that risks can have on one another. Ignoring such relationships can provide inaccurate inference and induce prohibitive losses for businesses. As such, our interest lies in exploring multivariate expectiles as these dependence structures can be incorporated directly into the measure.

First, the notion of elicibility is extended to the multivariate context. Here, for any d -dimensional random vector $\mathbf{X} \in \mathbb{R}^d$ an associated risk measure T_α is elicitable if we can write it as

$$T_\alpha(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \{S(\mathbf{x}, \alpha)\}.$$

Now consider the multivariate extensions of expectiles. Herrmann et al. [31] present a multivariate geometric definition of expectiles, which considers the possibility of selecting different levels of risk α for the components of the vector \mathbf{X} . Maume-Deschamps et al. [41] define two notions of multivariate expectiles: L^p -expectiles and Σ -expectiles. The focus of this paper is on L^p -expectiles. For any $p \in [1, \infty]$, define the L^p -expectile of a random vector \mathbf{X} by

$$\mathbf{e}_\alpha^p(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left\{ \alpha \|(X - \mathbf{x})_+\|_p^2 + (1 - \alpha) \|(X - \mathbf{x})_-\|_p^2 \right\}.$$

Specifically, this paper explores semi-parametric estimation of the L^1 -expectile, *i.e.*,

$$\mathbf{e}_\alpha(\mathbf{X}) := \mathbf{e}_\alpha^1(\mathbf{X}) = \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left\{ \alpha \left(\sum_{i=1}^d |X_i - x_{i+}| \right)^2 + (1 - \alpha) \left(\sum_{i=1}^d |X_i - x_{i-}| \right)^2 \right\}. \quad (2)$$

Analogously to the univariate case, the L^1 -expectile in (2) is the unique solution in \mathbb{R}^d of

$$\frac{1 - \alpha}{\alpha} = \frac{\mathbb{E}[\|(X - \mathbf{x})_+\|_1 \mathbb{1}\{X_k > x_k\}]}{\mathbb{E}[\|(X - \mathbf{x})_-\|_1 \mathbb{1}\{X_k < x_k\}]}, \quad k \in \{1, \dots, d\}.$$

35 Thus, it can be interpreted as a ratio of expected positive scenarios over negative ones.

In terms of estimation, we provide a consistent estimator for $\mathbf{e}_\alpha(\mathbf{X})$ in (2) for elevated risk levels $\alpha \approx 1$ when the underlying dependence structure and marginal distributions are unknown, through the approximated optimization problem

$$\arg \min_{\Theta \in \mathbb{R}^d} L_{\hat{\Lambda}}(\Theta),$$

for some loss function L and consistently estimated parameter set $\hat{\Lambda}$. In Maume-Deschamps et al. [41], it was shown that multivariate expectiles could be consistently estimated using Robbins-Monro's stochastic optimization for moderate levels of α (see discussion in Section 4.3). However, for elevated levels of α their approach fails. As such, extra care must be given when estimating $\mathbf{e}_\alpha(\mathbf{X})$ at extreme levels of risk.

40 The outline of the paper is as follows: In the remainder of Section 1, multivariate extreme expectiles (MEEs) and some important related concepts are introduced. In Section 2, a general optimization framework is provided for MEEs. The main contribution of this paper is in Section 3, where the consistency of the proposed approximate loss function and its capability to accurately produce MEEs for extreme levels of risk α is provided (Theorem 1 and Corollary 3, respectively). In Section 4, the finite-sample performance of our procedure is illustrated with several simulation studies. In Section 5, possible improvements of our estimation procedure are discussed. Section 6 provides concluding remarks. Finally, in the supplementary material Section 7 we provide additional simulation, exploration, and discussion for our procedure (see Section 7.1), an application of our procedure for real data (see Section 7.2) and the gradient calculation of the considered loss function (see Section 7.3).

1.2. Multivariate Extreme Expectiles

The notion of extreme expectiles has been explored in the univariate case by Bellini and Bigozzi [5], Mao et al. [39], Bellini and Di Bernardino [6], Daouia et al. [16, 17] and in the multivariate case by Maume-Deschamps et al. [41, 42]

and Herrmann et al. [31], for example. In the latter works, the authors discuss situations for which expectiles can be solved for explicitly or for moderate risk levels α . We recall several asymptotic and multivariate concepts that will be crucial to this work. The first of which is the notion of multivariate regular variation (MRV). Recall in one dimension a function f is considered to be regularly varying at $a \in \{0, +\infty\}$ with index $\theta \in \mathbb{R}$, denoted $f \in RV_\theta(a)$, if

$$\lim_{t \rightarrow a} \frac{f(tx)}{f(t)} = x^\theta$$

50 for all $x > 0$. Extending this definition to random vectors, one has the following equivalent definitions:

Definition 1 (MRV equivalent definitions). Let \mathbf{X} be a random vector on \mathbb{R}^d , the following definitions are equivalent:

- The vector \mathbf{X} has regularly varying tail of index θ .
- There exists for all $x > 0$ a finite measure μ on the unit sphere \mathbb{S}^{d-1} , a normalizing function $b : (0, \infty) \mapsto (0, \infty)$ such that

$$\lim_{t \rightarrow +\infty} \Pr \left\{ \|\mathbf{X}\| > xb(t), \frac{\mathbf{X}}{\|\mathbf{X}\|} \in \cdot \right\} = x^{-\theta} \mu(\cdot).$$

The measure μ depends on the chosen norm, it is called the *spectral measure* of \mathbf{X} .

- There exists a finite measure μ on the unit sphere \mathbb{S}^{d-1} , a slowly varying function L , and a positive real $\theta > 0$ such that

$$\lim_{x \rightarrow +\infty} \frac{x^\theta}{L(x)} \Pr \left(\|\mathbf{X}\| > x, \frac{\mathbf{X}}{\|\mathbf{X}\|} \in B \right) = \mu(B)$$

for all $B \in \mathcal{B}(\mathbb{S}^{d-1})$ with $\mu(\partial B) = 0$, where \mathcal{B} represents the Borel σ -algebra and ∂B the boundary of the set B .

55 For more information on the concept of regular variation and how it relates to MEEs, see Section 1 of Maume-Deschamps et al. [42]. Drawing further from Maume-Deschamps et al. [42], the focus here is on marginal distributions with asymptotically equivalent tails. This can be stated in the following assumption.

Assumption 1 (Equivalent Tails). Assume that \mathbf{X} has equivalent regularly varying marginal tails, which means that both (i) and (ii) below are satisfied:

60 (i) $\bar{F}_1 \in RV_{-\theta(+\infty)}$ with $\theta > 0$;

(ii) The tails of X_1, \dots, X_d are equivalent. That is for all $i \in \{2, \dots, d\}$, there is a positive constant c_i such that

$$\lim_{x \rightarrow +\infty} \frac{\bar{F}_i(x)}{\bar{F}_1(x)} = c_i.$$

Conditions (i) and (ii) imply that all margins are $RV_{-\theta(+\infty)}$. Denote c_i the tail ratio. Obviously $c_1 = 1$. Extreme asymptotic relationships within a random vector \mathbf{X} on \mathbb{R}^d are further quantified by the upper tail dependence function defined by

$$\lambda(\mathbf{x}) = \lim_{t \rightarrow 0^+} \frac{\Pr \{F_1(X_1) \geq 1 - tx_1, \dots, F_d(X_d) \geq 1 - tx_d\}}{t} = \lim_{t \rightarrow 0^+} \frac{\bar{C}(tx_1, \dots, tx_d)}{t}, \quad (3)$$

where \bar{C} is the survival copula associated to \mathbf{X} (see, for instance, Equation (1.4) or (1.5) in Maume-Deschamps et al. [42]). Pairwise upper tail dependencies will be of particular importance in this paper, so we denote by $\lambda^{ik}(x_i, x_k)$, the upper tail dependence function for the random pair (X_i, X_k) . It is also beneficial to define the stable tail dependence function (stdf)

$$\mathcal{L}(\mathbf{x}) = \lim_{t \rightarrow 0^+} \frac{1}{t} \Pr \{F(X_1) > 1 - tx_1 \text{ or } \dots \text{ or } F(X_d) > 1 - tx_d\}$$

and the bivariate relationship $\mathcal{L}(x, y) = x + y - \lambda(x, y)$. The following proposition from Maume-Deschamps et al. [42] connects all these concepts to MEEs by providing an alternative system of equations when $\alpha \rightarrow 1$.

Proposition 1 (Maume-Deschamps et al. [42], Proposition 3.3). Assume that X has regular varying multivariate distribution in the sense of Definition 1 and that Assumption 1 holds. Consider the L_1 -expectile $e_\alpha(X) = (e_\alpha^i(X))_{i \in \{1, \dots, d\}}$. Then any limit vector $\Theta = (\eta, \beta_2, \dots, \beta_d)$ of $((1 - \alpha)/\bar{F}_1\{e_\alpha^1(X)\}, e_\alpha^2(X)/e_\alpha^1(X), \dots, e_\alpha^d(X)/e_\alpha^1(X))$ for $\alpha \rightarrow 1$ satisfies the following system of equations

$$\frac{1}{\theta - 1} - \eta \frac{\beta_k^\theta}{c_k} = - \sum_{i=1, i \neq k}^d \left\{ \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt - \eta \frac{\beta_k^{\theta-1}}{c_k} \beta_i \right\}, \quad k \in \{1, \dots, d\}. \quad (4)$$

In particular, Maume-Deschamps et al. [42] show that the system solutions are

$$\eta^\perp = \frac{1}{(\theta - 1) \left\{ 1 + \sum_{j=2}^d c_j^{1/(\theta-1)} \right\}}, \quad \beta_k^\perp = c_k^{1/(\theta-1)}, \quad (5)$$

for asymptotically independent random vectors and

$$\eta^+ = \frac{1}{(\theta - 1)}, \quad \beta_k^+ = c_k^{1/\theta}, \quad (6)$$

for comonotonic random vectors, for all $k \in \{1, \dots, d\}$. Denote $\Theta^\perp = (\eta^\perp, \beta_2^\perp, \dots, \beta_d^\perp)$ and $\Theta^+ = (\eta^+, \beta_2^+, \dots, \beta_d^+)$. In the rest of this paper we are interested in using Proposition 1 towards estimating Θ and using this result to calculate MEEs. Given these definitions, it is easy to see that estimating $e_\alpha(X)$ for extreme risk levels α will require one to first consistently estimate θ , $c = (c_2, \dots, c_d)$ and λ^{ik} .

2. Optimization problem for MEE

As mentioned in Section 1, for moderate levels of α , effective calculation of $e_\alpha(X)$ is possible through Robbins-Monro's stochastic optimization. However, such a method does not perform well for $\alpha \approx 1$ (see Section 4.3 for further details). An alternative methodology will be required. In Definition 2 below, we use Proposition 1 as inspiration to define an alternative optimization problem for MEEs.

Definition 2. Consider a d -dimensional random vector X with regularly varying multivariate distribution in the sense of Definition 1 with margins possessing equivalent tails as given in Assumption 1. Let $\beta^* = (\beta_1, \dots, \beta_d)$ with $\beta_1 = 1$ and $\Theta = (\eta, \beta_2, \dots, \beta_d)$. Let $c_1 = 1$ and $\Lambda = (\theta, c_2, \dots, c_d, \lambda(\cdot))$, where c_i is the tail ratio for $i \in \{2, \dots, d\}$, θ is the tail index and λ is the upper tail dependence function associated to the random vector X in (3). Define the loss function

$$L_\Lambda(\Theta) := \frac{1}{2} \|F_\Lambda(\Theta)\|_2^2, \quad (7)$$

where

$$F_\Lambda(\Theta) = (F_\Lambda^{(1)}(\Theta), \dots, F_\Lambda^{(d)}(\Theta)) = (g_\Lambda^{(1)}(\Theta) + f_\Lambda^{(1)}(\Theta), \dots, g_\Lambda^{(d)}(\Theta) + f_\Lambda^{(d)}(\Theta)),$$

with, for all $k \in \{1, \dots, d\}$,

$$g_\Lambda^{(k)}(\Theta) = \frac{1}{\theta - 1} - \eta \frac{\beta_k^\theta}{c_k} \quad \text{and} \quad f_\Lambda^{(k)}(\Theta) = \sum_{i \neq k}^d \left\{ \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt - \eta \frac{\beta_k^{\theta-1}}{c_k} \beta_i \right\}.$$

Define an optimal vector Θ^* , obtained by optimizing the loss function L_Λ in (7), i.e.,

$$\Theta^* = \arg \min_{\Theta} L_\Lambda(\Theta). \quad (8)$$

Corollary 1 establishes that the optimization problem in (8) can be used to calculate MEEs.

Corollary 1. Consider a d -dimensional random vector \mathbf{X} with regularly varying multivariate distribution in the sense of Definition 1. Consider the L_1 -expectile $\mathbf{e}_\alpha(\mathbf{X}) = (e_\alpha^i(\mathbf{X}))_{i \in \{1, \dots, d\}}$ as in (2). Under Assumption 1 and the assumption that the vector $((1 - \alpha)/\bar{F}_1\{e_\alpha^1(\mathbf{X})\}, e_\alpha^2(\mathbf{X})/e_\alpha^1(\mathbf{X}), \dots, e_\alpha^d(\mathbf{X})/e_\alpha^1(\mathbf{X}))$ has a unique limit point $\Theta = (\eta, \beta_2, \dots, \beta_d)$ for $\alpha \rightarrow 1$, then Θ satisfies the optimization problem (8) in Definition 2. Moreover, the optimization problems (2) and (8) are equivalent for $\alpha \rightarrow 1$, i.e.,

$$\lim_{\alpha \rightarrow 1} \frac{e_\alpha^i(\mathbf{X})}{\bar{e}_\alpha^i(\mathbf{X})} = 1, \text{ for } i \in \{1, \dots, d\}, \quad \bar{\mathbf{e}}_\alpha(\mathbf{X}) := \text{VaR}_\alpha(X_1)\eta^{1/\theta}(1, \beta_2, \dots, \beta_d)^\top.$$

Proof. From Proposition 1 and the uniqueness of the limit $\Theta = (\eta, \beta_2, \dots, \beta_d)$ we get that Θ satisfies the optimization problem in (8). Furthermore, from Proposition 5.1 in Maume-Deschamps et al. [42], we know that $\mathbf{e}_\alpha(\mathbf{X}) \sim \text{VaR}_\alpha(X_1)\eta^{1/\theta}(1, \beta_2, \dots, \beta_d)^\top$, for $\alpha \rightarrow 1$. Hence, the result is proved. \square

80 To solve the optimization problem (8) in Definition 2, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) descent algorithm will be used in the present paper. The algorithm is a member of the family of quasi-Newton optimization methods and allows one to avoid calculating second derivatives, a feat which can drastically improve computation time. The BFGS algorithm is given in Algorithm 1.

Remark 1. For convergence results of the BFGS algorithm the interested reader is referred to Nocedal and Wright [44]. It is important to note that, in the general case, the true optimization problem (8) does not satisfy the convexity assumption required for global convergence of the BFGS algorithm. Nevertheless, the choice to use BFGS is motivated by its simplicity of implementation and further validated by the satisfactory numerical results provided in Section 4. More sophisticated algorithms which extend BFGS would indeed be possible, including L-BFGS-B [8] and modified BFGS [36, 37], the latter of which allows one to relax the convexity assumption and still achieve global convergence. These methods are briefly discussed in Section 5.

Algorithm 1 BFGS quasi-Newton optimization algorithm

(Step 0) Put counter $k := 0$ and choose initial values $\Theta^0 \in \mathbb{R}^d$, $H_0 \in \mathbb{R}^{d \times d}$ initial approximation to the inverse of the Hessian matrix of L_Λ , $\sigma \in (0, 1/2)$, $\rho \in (\sigma, 1)$, and $\epsilon \geq 0$.

(Step 1) Let L_Λ as in Definition 2. If $\|\nabla L_\Lambda(\Theta^k)\| \leq \epsilon$: STOP.

(Step 2) Calculate the direction $\mathbf{d}^k = -H_k \nabla L_\Lambda(\Theta^k)$.

(Step 3) Determine the step size $t_k > 0$ such that the Wolfe conditions are satisfied, i.e.,

$$\begin{aligned} L_\Lambda(\Theta^k + t_k \mathbf{d}^k) &\leq L_\Lambda(\Theta^k) + \sigma t_k \nabla L_\Lambda(\Theta^k)^\top \mathbf{d}^k, \\ \nabla L_\Lambda(\Theta^k + t_k \mathbf{d}^k)^\top \mathbf{d}^k &\geq \rho \nabla L_\Lambda(\Theta^k)^\top \mathbf{d}^k. \end{aligned}$$

(Step 4) Update the following:

$$\begin{aligned} \bullet \Theta^{k+1} &:= \Theta^k + t_k \mathbf{d}^k & \bullet \mathbf{y}^k &:= \nabla L_\Lambda(\Theta^{k+1}) - \nabla L_\Lambda(\Theta^k), \\ \bullet \mathbf{s}^k &:= \Theta^{k+1} - \Theta^k & \bullet H_{k+1} &:= (\mathbb{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^\top) H_k (\mathbb{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^\top) + \rho_k \mathbf{s}_k \mathbf{s}_k^\top, \end{aligned}$$

where $\rho_k = 1/\mathbf{y}_k^\top \mathbf{s}_k$.

(Step 5) Set $k \leftarrow k + 1$ and go to **(Step 1)**.

3. Approximated Optimization Problem

Since the underlying distribution is not known in practice, crucial information, including the values of λ , θ , and c_i for $i \in \{2, \dots, d\}$, is not available. As such, direct application of the BFGS algorithm for the optimization problem (8) is not possible. Instead, an approximated version of (8) will be solved, *i.e.*, one can focus on the approximate optimum

$$\arg \min_{\Theta \in \mathbb{R}^d} L_{\hat{\Lambda}}(\Theta) \quad (9)$$

for some vector of estimators $\hat{\Lambda} = (\hat{\theta}, \hat{c}_2, \dots, \hat{c}_d, \hat{\lambda})$. Specifically, convergence of the estimated optimum can be shown in the following way. The first step will be to show that $\hat{\Lambda} \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Lambda$, and subsequently that $L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\text{Pr}} L_{\Lambda}(\Theta)$ and $\nabla L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \nabla L_{\Lambda}(\Theta)$, where $\xrightarrow[n \rightarrow \infty]{\text{Pr}}$ denotes convergence in probability. Then, we will prove the consistency of every iteration of the BFGS algorithm (see Theorem 1)

$$\hat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Theta^k, \quad k \in \{1, 2, \dots\}.$$

Finally, by using a two-step procedure (see Algorithm 2), we can study the behaviour of the proposed method (see Corollary 3).

3.1. Loss function

Establishing the consistency of every iteration of the BFGS algorithm applied to the approximated loss function comes down to establishing the convergence of the approximate loss function $L_{\hat{\Lambda}}$ and its gradient $\nabla L_{\hat{\Lambda}}$. In the end, this will require convergence arguments for estimators $\hat{\lambda}^{ik}$, $\hat{\theta}$, and \hat{c}_i for $i \in \{2, \dots, d\}$, of λ^{ik} , θ , and c_i for $i \in \{2, \dots, d\}$, respectively. Moreover, the convergence

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt \xrightarrow[n \rightarrow \infty]{\text{Pr}} \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt$$

is also required. For n observations of the random vector X , denote the random sample $X_i = (X_{i1}, \dots, X_{id})$, $i \in \{1, \dots, n\}$. Next, denote $X_{j:1,n} \leq X_{j:2,n} \leq \dots \leq X_{j:n,n}$ the order statistics associated to the realizations of the j^{th} random variable of X for $j \in \{1, \dots, d\}$. First, consider the estimated tail index $\hat{\theta}$ via Hill's estimator Hill [33],

$$\hat{\gamma} = \frac{1}{\ell_{\theta}} \sum_{i=1}^{\ell_{\theta}} \ln(X_{1:n-i+1,n}) - \ln(X_{1:n-\ell_{\theta},n}), \quad \hat{\theta} = \frac{1}{\hat{\gamma}}, \quad (10)$$

for an intermediate integer sequence, *i.e.*, $\ell_{\theta} = \ell_{\theta}(n)$ such that $\ell_{\theta}(n) = o(n)$ and $\ell_{\theta}(n) \xrightarrow[n \rightarrow \infty]{} \infty$. The tail ratio, c_i , can be estimated by

$$\hat{c}_i = \left(\frac{X_{i:n-\ell_i+1,n}}{X_{1:n-\ell_i+1,n}} \right)^{-\hat{\theta}}, \quad i \in \{2, \dots, d\}, \quad (11)$$

95 where $\ell_i = \ell_i(n)$ is an intermediate integer sequence and $\hat{\theta}$ is as defined in (10). The consistency of $\hat{\theta}$ and \hat{c}_i is established, for instance, in Deheuvels et al. [18] or de Haan and Ferreira [30] and Maume-Deschamps et al. [42], respectively. This leaves the upper tail dependence function as the sole remaining piece.

3.2. Upper tail dependence function

One group of estimators for the upper tail dependence function take the form

$$\hat{\lambda}(\mathbf{x}) = \frac{n}{\ell_\lambda} \widehat{C} \left(\frac{\ell_\lambda}{n} \mathbf{x} \right),$$

where \widehat{C} is some estimator of the survival copula \overline{C} for the random vector \mathbf{X} and $\ell_\lambda = \ell_\lambda(n)$ is an intermediate integer sequence. Thus, to estimate the upper tail dependence function effectively, a consistent estimator of the copula is required. In this paper, the empirical beta copula was chosen. For a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ with corresponding ranks denoted by $R_{i,j} = \sum_{k=1}^n \mathbb{1}\{X_{kj} \leq X_{ij}\}$, $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, d\}$, the empirical beta copula is defined as

$$C_n^{\text{Beta}}(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{R_{i,j},n}(u_i),$$

where $\mathbf{u} \in [0, 1]^d$ and $F_{r,n}(u)$ is the cumulative distribution function (cdf) of $U_{r,n}$, the r^{th} order statistic of n independent uniform random variables on $[0, 1]$. This also corresponds to $U_{r,n} \sim \text{Beta}(r, n+1-r)$. In two dimensions this gives

$$C_n^{\text{Beta}}(u, v) = \frac{1}{n} \sum_{i=1}^n F_{R_{i,1},n}(u) F_{R_{i,2},n}(v).$$

Moreover, as C_n^{Beta} is a genuine copula, using the bivariate relationship $\overline{C}(u, v) = u + v - 1 + C(1-u, 1-v)$ gives pairwise upper tail dependence function

$$\hat{\lambda}_{\text{Beta}}^{ik}(x_i, x_k) = \frac{n}{\ell_\lambda} \left\{ \frac{\ell_\lambda}{n} x_i + \frac{\ell_\lambda}{n} x_k - 1 + \frac{1}{n} \sum_{j=1}^n F_{R_{j,i},n} \left(1 - \frac{\ell_\lambda}{n} x_i \right) F_{R_{j,k},n} \left(1 - \frac{\ell_\lambda}{n} x_k \right) \right\}. \quad (12)$$

Proposition 2 establishes the consistency of the upper tail dependence function based on the empirical beta copula.

Proposition 2. *For the estimator of the upper tail dependence function based on the empirical beta copula, as defined in (12), it holds that $\hat{\lambda}_{\text{Beta}}^{ik}$ is a consistent estimator for λ^{ik} , i.e.,*

$$\hat{\lambda}_{\text{Beta}}^{ik}(x_i, x_k) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \lambda^{ik}(x_i, x_k).$$

Proof. The proof is straightforward from the work of Kiriliouk et al. [34]. They define the empirical process

$$B_{n,\ell_\lambda}^{\text{Beta}} = \sqrt{\ell_\lambda} (\mathcal{L}_{n,\ell_\lambda}^{\text{Beta}} - \mathcal{L}),$$

where $\mathcal{L}_{n,\ell_\lambda}^{\text{Beta}}(\mathbf{x}) = n \{1 - C_n^{\text{Beta}}(1 - \ell_\lambda \mathbf{x}/n)\} / \ell_\lambda$ is the stdf defined under the empirical beta copula. In particular, they established the weak convergence (as defined in van der Vaart and Wellner [50]) $B_{n,\ell_\lambda}^{\text{Beta}} \rightsquigarrow B$ on $\ell^\infty([0, 1])$, where B is some stochastic process with continuous trajectories. Using this result, and the bivariate relationship that for any stdf $\lambda(x, y) = x + y - \mathcal{L}(x, y)$, the empirical process for upper tail dependence function based on the empirical beta copula will have the same (up to a sign) asymptotic distribution. This implies that

$$\hat{\lambda}_{\text{Beta}}^{ik}(x_i, x_k) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \lambda^{ik}(x_i, x_k),$$

100 establishing the result. □

In particular, Proposition 2 gives

$$\hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right).$$

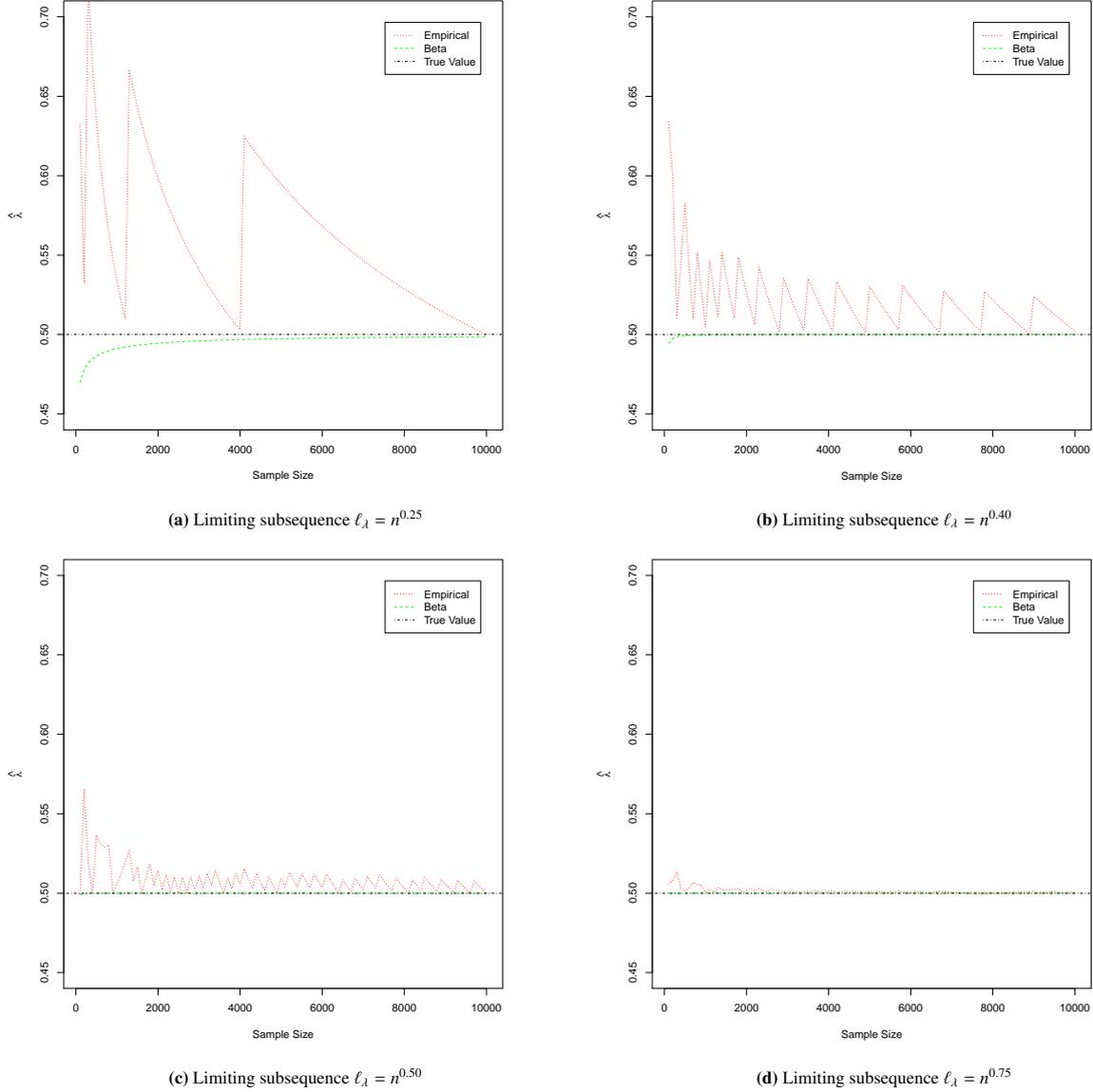


Fig. 1: Comparison between estimates of the upper tail dependence function based the empirical copula, $\hat{\lambda}_n^{23}(0.5, 1.5)$ (red), and the empirical beta $\hat{\lambda}_{\text{Beta}}(0.5, 1.5)$ (green), for varying values of ℓ_λ with sample size ranging from 100 to 10000. The true value, $\lambda^{23}(0.5, 1.5)$ (black), is also included. The margins are Pareto type I. Specifically, $X_i \sim P[2, 1.25(1+i)]$, $i \in \{1, 2, 3\}$.

In Fig. 1, an illustration of the performance of $\hat{\lambda}_{\text{Beta}}$ is provided using independent Pareto type I distributions. Specifically, we consider $X_i \sim P[2, 1.25(1+i)]$, $i \in \{1, 2, 3\}$ with cdf $F_i(x) = 1 - \{1.25(1+i)/x\}^2$. For comparison, the pairwise upper tail dependence function based on the standard empirical copula

$$\hat{\lambda}_n^{ik}(x_i, x_k) = \frac{1}{\ell_\lambda} \sum_{j=1}^n \mathbb{1} \left\{ F_{in}(X_{ji}) > 1 - \frac{\ell_\lambda}{n} x_i, F_{jn}(X_{jk}) > 1 - \frac{\ell_\lambda}{n} x_k \right\},$$

where X_{ji} is the j^{th} realization of the r.v. X_i , is also used. As can be seen, $\hat{\lambda}_{\text{Beta}}$ provides a far more accurate estimate of λ and doesn't exhibit any of the oscillating behavior that is seen when considering $\hat{\lambda}_n$. The performance is also better at smaller sample sizes when a larger limiting subsequence is considered.

3.3. Integral of the upper tail dependence function

The final piece of the estimation procedure is to approximate

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt. \quad (13)$$

105 To our knowledge, a consistent estimator for the integral of the upper tail dependence function has not been established in prior literature. Again, the underlying upper tail dependence function is estimated using the empirical beta copula. The result is established in the following proposition.

Proposition 3. *Consider the estimator of λ as proposed in (12), then*

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt \xrightarrow[n \rightarrow \infty]{\text{Pr}} \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt.$$

Proof. The proof boils down to an application of the Dominated Convergence Theorem. First, the convergence of the integrand follows from Proposition 2. Second, as the empirical beta copula is a proper copula, it is easy to see that its upper tail dependence function satisfies the bounds

$$\lambda^{ik,\perp} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \leq \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \leq \lambda^{ik,+} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right)$$

and hence

$$0 \leq \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \leq \min \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right).$$

See for instance Proposition 2.1 in Kiriliouk et al. [34]. Third, it is shown in Maume-Deschamps et al. [42] (see the proof of Lemma 3.4) that

$$\begin{aligned} \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik,+} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt &= \frac{c_i}{c_k} \left(\frac{\beta_i}{\beta_k} \right)^{-\theta+1} \left(\left\{ \frac{\beta_k}{\beta_i} \left(\frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} - 1 \right\}_+ \frac{c_k}{c_i} \left(\frac{\beta_k}{\beta_i} \right)^{-\theta} \right. \\ &\quad \left. + \frac{1}{\theta-1} \left[1 + \left\{ \frac{\beta_i}{\beta_k} \left(\frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} - 1 \right\}_+ \right]^{-\theta+1} \right) < \infty. \end{aligned}$$

Thus, by dominated convergence, we get the result. \square

110 In Fig. 2, an illustration of the finite sample performance of the estimator is provided. In particular, for each sample size and subsequence combination, the calculation of $\int_{\frac{\beta_2}{\beta_3}}^{\infty} \lambda_{\text{Beta}}^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ was replicated 500 times. As can be seen for samples of size $n > 400$ and subsequences n^q for $q > 0.4$, the estimator performs well.

Once again, it is compared to an equivalent quantity, using the standard empirical copula. In fact, under the empirical copula, a closed form estimator for (13) exists, as outlined in Proposition 4 below.

Proposition 4. *Under the empirical copula*

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1} \left\{ \frac{R_{i,j}}{n} \leq u_j \right\},$$

it holds that

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}_n^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt = \frac{1}{\ell_\lambda} \sum_{j=1}^n \left[\left(\frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} - \frac{\beta_i}{\beta_k} \right]_+ \cdot \mathbb{1} \{ X_{kj} \geq X_{k:n-\ell_\lambda, n} \}.$$

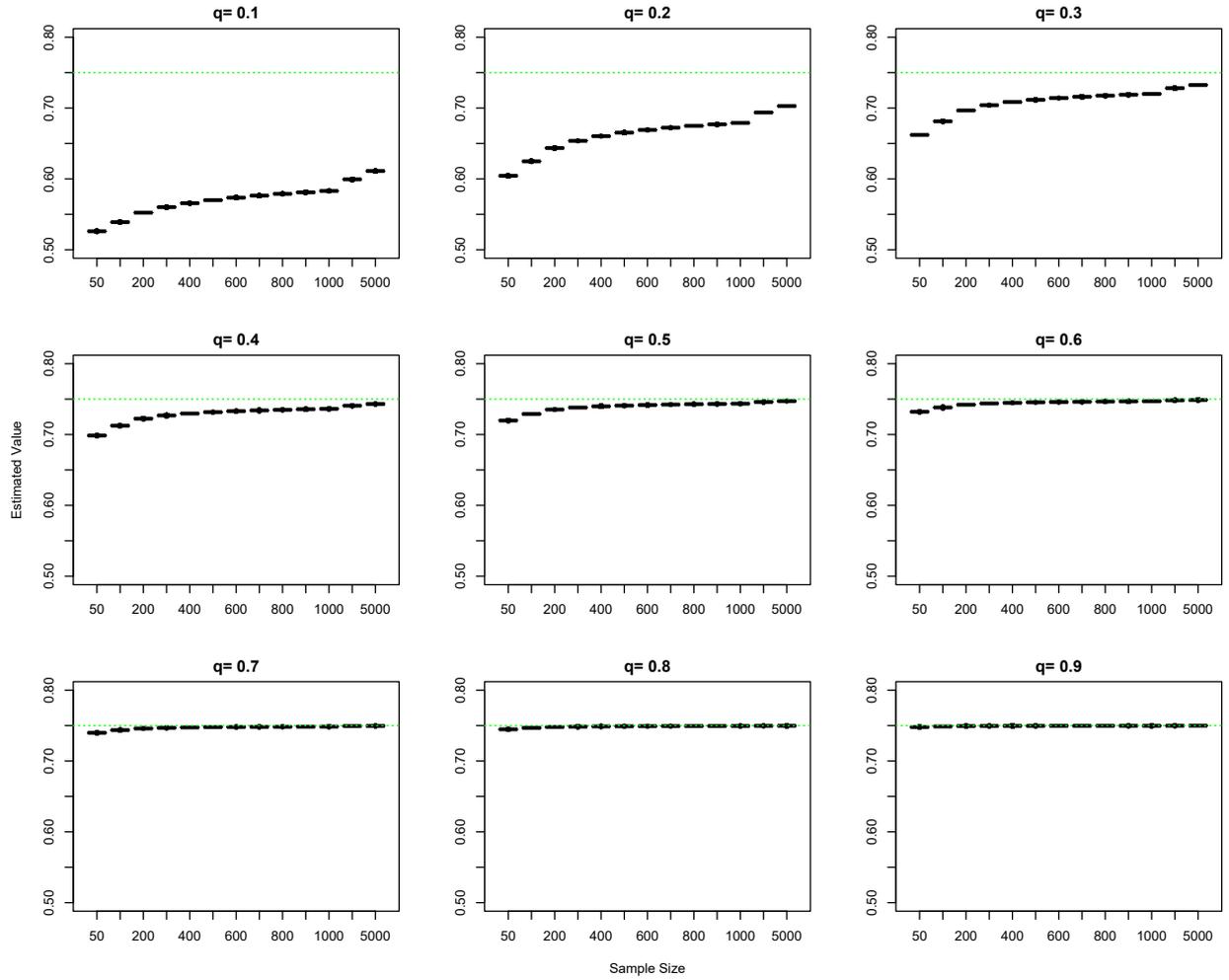


Fig. 2: Performance of $\int_{\beta_2}^{\infty} \lambda_{\text{Beta}}^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_\lambda = n^q$, $q \in \{0.1, 0.2, \dots, 0.9\}$ with sample size $n \in \{100, 200, \dots, 1000, 2500, 5000\}$. They are compared to the true value $\int_{\beta_2}^{\infty} \lambda_{\beta_3}^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ under Pareto type I margins with a comonotonic dependence structure (green horizontal line). Specifically, $X_i \sim P\{2, 1.25(1+i)\}$, $i \in \{1, 2, 3\}$.

Proof. One can write

$$\begin{aligned}
\int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}_n^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt &= \int_{\frac{\beta_i}{\beta_k}}^{\infty} \frac{1}{\ell_\lambda} \sum_{j=1}^n \mathbb{1} \left\{ 1 - F_{n,i}(X_{ij}) < \frac{\ell_\lambda \frac{c_i}{c_k} t^{-\theta}}{n}, 1 - F_{n,k}(X_{kj}) < \frac{\ell_\lambda}{n} \right\} dt \\
&= \int_{\frac{\beta_i}{\beta_k}}^{\infty} \frac{1}{\ell_\lambda} \sum_{j=1}^n \mathbb{1} \left\{ \frac{n - R_{i,j} + 1}{n} < \frac{\ell_\lambda \frac{c_i}{c_k} t^{-\theta}}{n} \right\} \mathbb{1} \{X_{kj} \geq X_{k:n-\ell_\lambda,n}\} dt \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \int_{\frac{\beta_i}{\beta_k}}^{\infty} \mathbb{1} \left\{ \left(\frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} > t \right\} \mathbb{1} \{X_{kj} \geq X_{k:n-\ell_\lambda,n}\} dt \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \int_{\frac{\beta_i}{\beta_k}}^{\left(\frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-\frac{1}{\theta}}} \mathbb{1} \left\{ \left(\frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} \geq \frac{\beta_i}{\beta_k} \right\} \mathbb{1} \{X_{kj} \geq X_{k:n-\ell_\lambda,n}\} dt \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \left[\left\{ \frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right\}^{-\frac{1}{\theta}} - \frac{\beta_i}{\beta_k} \right] \mathbb{1} \left\{ \left(\frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right)^{-\frac{1}{\theta}} \geq \frac{\beta_i}{\beta_k} \right\} \mathbb{1} \{X_{kj} \geq X_{k:n-\ell_\lambda,n}\} \\
&= \frac{1}{\ell_\lambda} \sum_{j=1}^n \left[\left\{ \frac{n - R_{i,j} + 1}{\ell_\lambda} \cdot \frac{c_k}{c_i} \right\}^{-\frac{1}{\theta}} - \frac{\beta_i}{\beta_k} \right]_+ \mathbb{1} \{X_{kj} \geq X_{k:n-\ell_\lambda,n}\}.
\end{aligned}$$

□

115 Fig. 3 illustrates the performance of $\int_{\beta_2/\beta_3}^{\infty} \hat{\lambda}_n \left\{ (c_2/c_3)t^{-\theta}, 1 \right\} dt$ as in Proposition 4. Comparing to Fig. 2, one sees that the empirical beta copula far surpasses the standard empirical copula in terms of accuracy.

While the convergence of each individual estimator has been established, they must now be combined to establish the convergence of our estimation procedure. Using the consistency of (10), (11), and (12) yields the following corollary

Corollary 2. Taking $\hat{\Lambda} = (\hat{\theta}, \hat{c}_2, \dots, \hat{c}_d, \hat{\lambda}_{\text{Beta}}^{ik})$ as defined in (10), (11), and (12), one has

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) dt \xrightarrow[n \rightarrow \infty]{\text{Pr}} \int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) dt.$$

Proof. For simplicity, we establish a similar result for $\hat{\lambda}_{\text{Beta}}^{ik}$, i.e.,

$$\hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right).$$

120 The proof makes use of the continuous mapping theorem (denoted CMT in the following); see, e.g., van der Vaart [49], Theorem 2.3. Beginning with the first statement, using several iterations of the triangle inequality yields

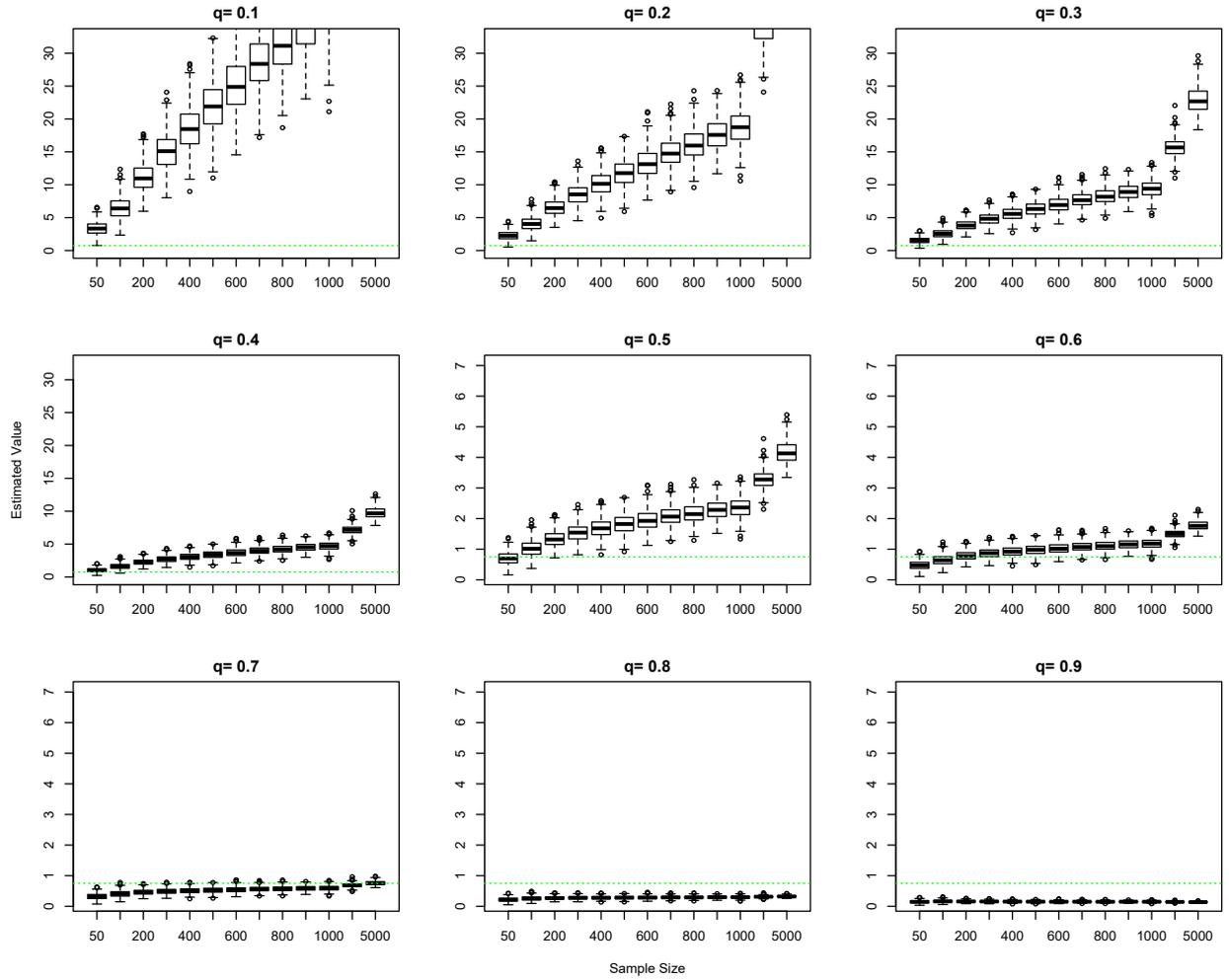


Fig. 3: Performance of $\int_{\beta_2}^{\infty} \hat{\lambda}_n^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ as in Proposition 4 for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, *i.e.*, $\ell_\lambda = n^q$, $q \in \{0.1, 0.2, \dots, 0.9\}$ with sample size $n \in \{100, 200, \dots, 1000, 2500, 5000\}$. They are compared to the true value $\int_{\beta_2}^{\infty} \lambda^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ under Pareto type I margins with a comonotonic dependence structure (green horizontal line). Specifically, $X_i \sim P(2, 1.25(1+i))$, $i \in \{1, 2, 3\}$.

$$\begin{aligned}
& \left| \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| = \left| \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| \\
& = \left| \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| \\
& = \left| \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) \right. \\
& \quad \left. + \lambda^{ik} \left(\frac{c_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) + \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right| \\
& \leq \underbrace{\left| \hat{\lambda}_{\text{Beta}}^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) \right|}_{(i)} + \underbrace{\left| \lambda^{ik} \left(\frac{\hat{c}_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) \right|}_{(ii)} \\
& \quad + \underbrace{\left| \lambda^{ik} \left(\frac{c_i}{\hat{c}_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) \right|}_{(ii)} + \underbrace{\left| \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\hat{\theta}}, 1 \right) - \lambda^{ik} \left(\frac{c_i}{c_k} t^{-\theta}, 1 \right) \right|}_{(iii)} \xrightarrow[n \rightarrow \infty]{\text{Pr}} 0,
\end{aligned}$$

where

(i) by the convergence of $\hat{\lambda}_{\text{Beta}}^{ik}$ (Proposition 2),

(ii) by the CMT on λ^{ik} and the consistency of \hat{c}_i , $i \in \{1, \dots, d\}$ (Maume-Deschamps et al. [42], Proposition 4.2), and

(iii) by the CMT on λ^{ik} and the consistency of $\hat{\theta}$ (de Haan and Ferreira [30], Theorem 3.2.2).

125 The proof of the main statement would follow in the same way, or as a result of the dominated convergence theorem, similar to Proposition 3. \square

For an illustration of Corollary 2, simulations are presented in Fig. 4. It is clear that adding extra uncertainty through calculation of $\hat{\theta}$ and \hat{c}_i , $i \in \{2, \dots, d\}$, decreases the accuracy of the full estimator. The variability of these quantities is well documented and this would naturally require an increased number of observations to estimate with confidence.

130 Moreover, the selection of the subsequences ℓ_θ , ℓ_λ and ℓ_i , $i \in \{2, \dots, d\}$ is crucial. For instance, the impact of the choice of ℓ_λ is illustrated in Figs. 2-4. In a real data study, the subsequences could be calibrated by some stability criterion. More precisely, in Section 7.2, we will use the selection procedure of intermediate sequences described, for instance, by Cai et al. [9] and Di Bernardino and Prieur [20]. We will consider a range of possibilities for intermediate sequences for which the associated estimators showed some stability. Then, to gain robustness, we average the estimates corresponding to the selected ranges.

3.4. Consistency

In Theorem 1 below, we first prove that the approximated loss function (approximated gradient function) in the optimization problem (9) converges to the loss function in (8) (to the gradient function). Furthermore, every iteration of the BFGS algorithm applied to the approximated optimization problem is consistent for the true optimization problem. Subsequently, we show that, under the assumption that BFGS captures the global minimum of the true optimization problem (8), the approximate optimization problem (9) can be used to solve for MEEs (see Algorithm 2 and Corollary 3 below).

Theorem 1. Let $\hat{\Lambda} = (\hat{\theta}, \hat{c}_2, \dots, \hat{c}_d, \hat{\lambda}_{\text{Beta}}^{ik})$ as in and (12) respectively. Then

$$L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\text{Pr}} L_{\Lambda}(\Theta) \quad \text{and} \quad \nabla L_{\hat{\Lambda}}(\Theta) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \nabla L_{\Lambda}(\Theta). \quad (14)$$

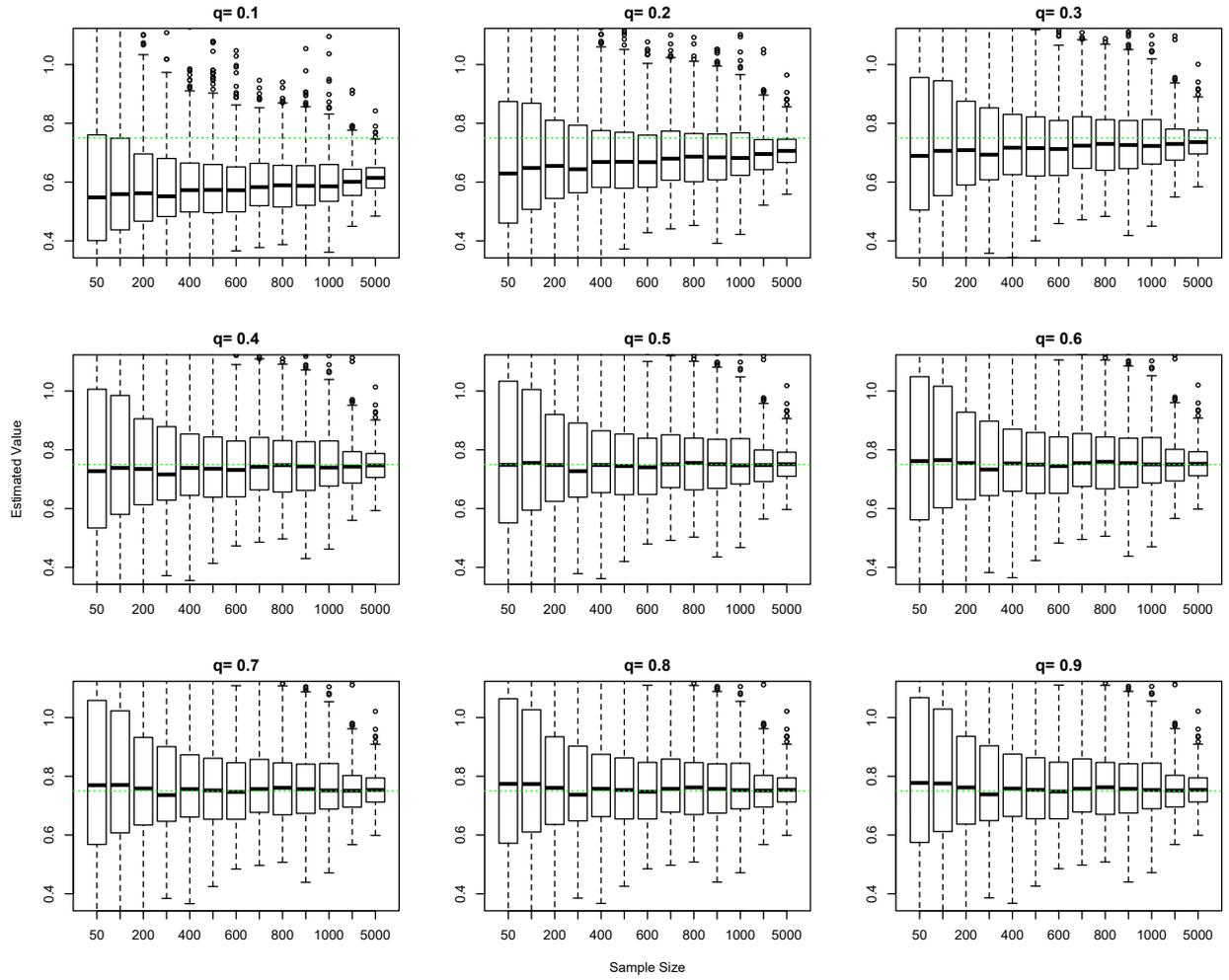


Fig. 4: Performance of $\int_{\beta_2}^{\infty} \lambda_{\text{Beta}}^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ for various sample sizes and subsequences. The variable q refers to the power used for our subsequence, i.e., $\ell_\lambda = n^q$, $q \in \{0.1, 0.2, \dots, 0.9\}$ with sample size $n \in \{100, 200, \dots, 1000, 2500, 5000\}$. They are compared to the true value $\int_{\beta_2}^{\infty} \lambda^{23} \left(\frac{c_2}{c_3} t^{-\theta}, 1 \right) dt$ under Pareto type I margins with a comonotonic dependence structure (green horizontal line). Specifically, $X_i \sim P\{2, 1.25(1+i)\}$, $i \in \{1, 2, 3\}$.

Moreover, given identical starting values Θ^0 , H_0 , $\sigma \in (0, 1/2)$, $\rho \in (\sigma, 1)$ and $\epsilon \geq 0$, for any step k , it holds that

$$\widehat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Theta^k. \quad (15)$$

Proof. Noting the structure of L_Λ and $\nabla L_\Lambda(\Theta)$ (detailed in 7.3), the consistency results in (14) are a direct result of the consistency of $\hat{\theta}$, \hat{c}_i , $i \in \{2, \dots, d\}$, Propositions 2, 3 and Slutsky's Theorem. For (15), we use induction. First, fix Θ^0 , H_0 , $\sigma \in (0, 1/2)$, $\rho \in (\sigma, 1)$ and $\epsilon \geq 0$. Recalling the iteration step of BFGS algorithm, we have

$$\hat{d}^0 = -H_0 \nabla L_{\hat{\Lambda}}(\Theta^0) \xrightarrow[n \rightarrow \infty]{\text{Pr}} -H_0 \nabla L_\Lambda(\Theta^0) = d^0.$$

Similarly, it is clear that $\hat{t}_0 \xrightarrow[n \rightarrow \infty]{\text{Pr}} t_0$. It follows that

$$\widehat{\Theta}^1 = \Theta^0 + \hat{t}_0 \hat{d}^0 \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Theta^0 + t_0 d^0 = \Theta^1,$$

as well as $\hat{y}^0 \xrightarrow[n \rightarrow \infty]{\text{Pr}} y^0$, $\hat{s}^0 \xrightarrow[n \rightarrow \infty]{\text{Pr}} s^0$ and $\widehat{H}_1 \xrightarrow[n \rightarrow \infty]{\text{Pr}} H_1$. Next, assume $\widehat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Theta^k$. Then,

$$\widehat{\Theta}^{k+1} = \widehat{\Theta}^k + \hat{t}_k \hat{d}_k = \widehat{\Theta}^k + \hat{t}_k \widehat{H}_k^{-1} \nabla L_{\hat{\Lambda}}(\widehat{\Theta}^k) \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Theta^k + t_k H_k^{-1} \nabla L_\Lambda(\Theta^k) = \Theta^{k+1}.$$

Thus, for any step $k \in \{1, 2, \dots\}$, $\widehat{\Theta}^k \xrightarrow[n \rightarrow \infty]{\text{Pr}} \Theta^k$ by induction. \square

We now proceed by using an iterated two-step procedure. We first provide an adequate estimate of the true loss function (and its gradient) and then proceed with the optimization procedure. This is formalized in Algorithm 2.

Algorithm 2 Two-steps estimation procedure for MEEs

(Step 1) Taking the limit $n \rightarrow \infty$.

Establish the consistency of $\hat{\Lambda}$ and subsequently $L_{\hat{\Lambda}}$ and $\nabla L_{\hat{\Lambda}}$ (see Theorem 1).

(Step 2) Taking the limit $k \rightarrow \infty$.

Optimize the consistently approximated problem from Step 1 using the BFGS algorithm.

145

The resulting asymptotic behavior of our estimation scheme using Algorithm 2 is established in Corollary 3. It is important to note the limit order in the non-exchangeable iterated limit (16).

Corollary 3. Under the assumptions of Corollaries 1, 2 and Theorem 1, and further assuming that the BFGS algorithm solves for the global minimum of (8), it holds that

$$\lim_{k \rightarrow \infty} (\text{plim}_{n \rightarrow \infty} \widehat{\Theta}^k) = \Theta^*. \quad (16)$$

Proof. Under the assumption that BFGS solves for the global minimum of (8), i.e., $\Theta^k \xrightarrow[k \rightarrow \infty]{} \Theta^*$, we have

$$\begin{aligned} \lim_{k \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \Pr(\|\widehat{\Theta}^k - \Theta^*\| \geq \epsilon) \right\} &= \lim_{k \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \Pr(\|\widehat{\Theta}^k - \Theta^k + \Theta^k - \Theta^*\| \geq \epsilon) \right\} \\ &\leq \lim_{k \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \Pr(\|\widehat{\Theta}^k - \Theta^k\| + \|\Theta^k - \Theta^*\| \geq \epsilon) \right\} \\ &\leq \lim_{k \rightarrow \infty} \left[\underbrace{\lim_{n \rightarrow \infty} \left\{ \Pr(\|\widehat{\Theta}^k - \Theta^k\| \geq \epsilon/2) \right\}}_{\text{(Step 1)}} + \underbrace{\Pr(\|\Theta^k - \Theta^*\| \geq \epsilon/2)}_{\text{(Step 2)}} \right] = 0, \end{aligned}$$

where firstly (Step 1) converges to zero as $n \rightarrow \infty$ by Theorem 1 and subsequently (Step 2), which does not depend on n , converges to 0 as $k \rightarrow \infty$ by assumption. \square

150 4. Numerical study

In this section, several numerical examples are provided demonstrating the effectiveness of our estimation procedure. In Section 4.1, we focus on the limiting cases of tail dependence: asymptotic independence and comonotonicity. As per Maume-Deschamps et al. [42], closed form solutions exist for these tail dependence structures so they will provide a good measuring stick for our methodology. In Section 4.2, we provide examples using non-trivial tail dependence structures, *i.e.*, neither asymptotic independence nor comonotonicity, where closed-form solutions for the system of equations in (4) exist. As such, we can further illustrate the performance of our estimates in this intermediate situation.

Results in each case are provided for both $\widehat{\Theta}$ and $\widehat{e}_\alpha(\mathbf{X})$. In the latter case, using Proposition 5.1 from Maume-Deschamps et al. [42], the final calculation of MEE estimates can be done using

$$\widehat{e}_\alpha(\mathbf{X}) = \widehat{\text{VaR}}_\alpha(X_1) \widehat{\eta}^{1/\widehat{\theta}}(1, \widehat{\beta}_2, \dots, \widehat{\beta}_d)^\top,$$

where

$$\widehat{\text{VaR}}_\alpha(X_1) = X_{1:n-k(n)+1,n} \left\{ \frac{k(n)}{(1-\alpha)n} \right\}^{1/\widehat{\theta}}, \quad (17)$$

called Weissman's estimator [51], for an intermediate integer sequence $k(n)$ is better suited than standard estimation techniques, using the empirical cdf, for large risk levels $\alpha \approx 1$. While both $\widehat{\Theta}$ and $\widehat{e}_\alpha(\mathbf{X})$ are provided, results for $\widehat{\Theta}$ will more effectively demonstrate the capabilities of our procedure as they do not rely on the performance of (17).

160 The simulations were conducted for sample sizes $n \in \{50, 100, 250, 500, 1000, 2500\}$ and repeated 500 times for each sample size. The subsequences for θ , c_i for $i \in \{2, \dots, d\}$ and λ were $\ell_\theta = n^{0.75}$, $\ell_i = n^{0.75}$ for $i \in \{2, \dots, d\}$ and $\ell_\lambda = n^{0.5}$ respectively. The risk level was set to $\alpha = 1 - 1/5000 = 0.9998$ to reflect the largest sample size considered. True system solutions Θ can be found in the last line of each table. First, we consider the limiting cases of tail dependence: asymptotic independence and comonotonicity.

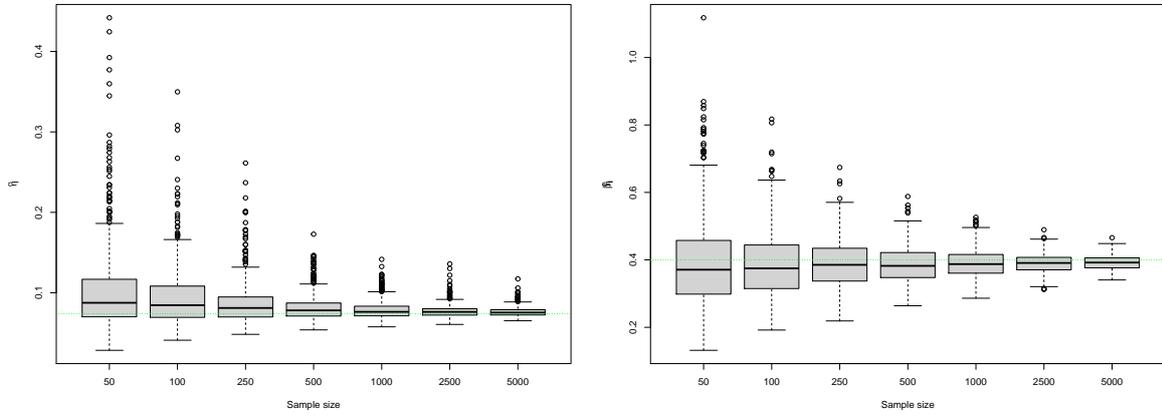
165 4.1. Asymptotic independence and comonotonicity

For the limiting cases of tail dependence, we consider three and five dimensional random vectors with Pareto type I marginals. Specifically, $X_i \sim P(b_i, a)$ where $a = \theta = 3.5$ and $b_i = 1.25(i + 1)$, $i \in \{1, \dots, d\}$, $d \in \{3, 5\}$. For Pareto margins with cdfs F_i , $i \in \{1, \dots, 5\}$ and these parameter settings, the tail ratios are

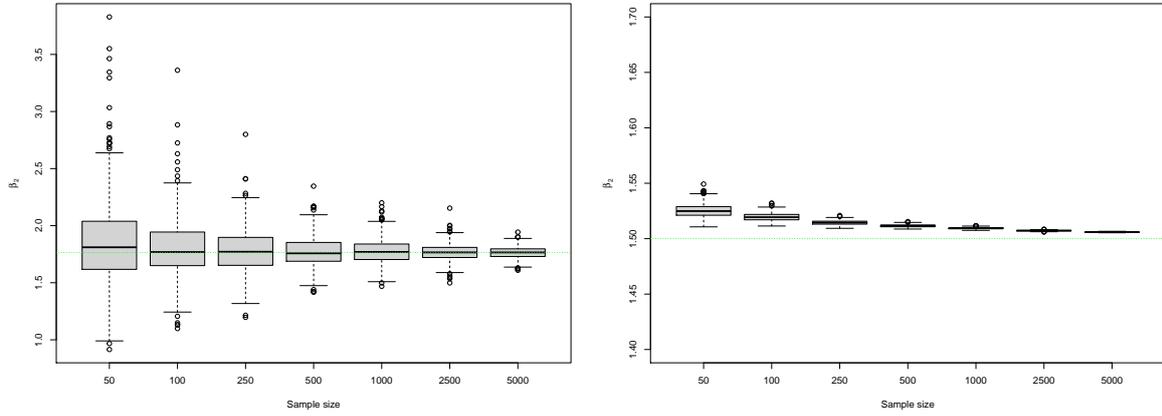
$$c_i = \lim_{x \rightarrow \infty} \frac{\overline{F}_i(x)}{\overline{F}_1(x)} = \left(\frac{b_i}{b_1} \right)^a = \begin{cases} 1.000, & i = 1, \\ 4.134, & i = 2, \\ 11.314, & i = 3, \\ 24.705, & i = 4, \\ 46.765, & i = 5. \end{cases}$$

Three-dimensional case

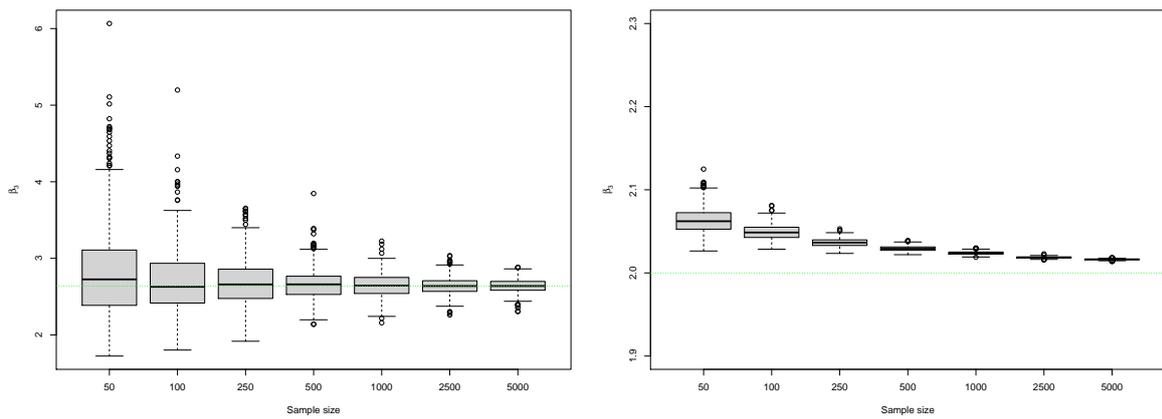
170 The summarized results of the simulations are presented in Table 1 and boxplots can be found in Fig. 5. In the three-dimensional case, it can be seen that the estimation procedure proposed in this paper performs quite well. At lower samples, the results are only slightly biased, but with a fair amount of standard error. As the sample size increases however, the algorithm is able to successfully estimate the true value of Θ and subsequently the true value of $e_\alpha(\mathbf{X})$ for $\alpha = 0.9998$ with minimal standard errors. It should be noted that for the comonotonic simulations the slight bias and small standard errors would mean that the true value would likely not be contained in any reasonable confidence interval, though the differences become smaller and smaller as the sample size increases.



(a) Results for $\hat{\eta}$



(b) Results for $\hat{\beta}_2$



(c) Results for $\hat{\beta}_3$

Fig. 5: Boxplots for the estimated solution vector for simulations using a three-dimensional random vector with Pareto type I margins for various sample sizes n . Specifically, $X_i \sim P(3.5, 1.25(1+i))$, $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$. Results are for independent margins (left) and comonotonic margins (right). True values Θ^- and Θ^+ (dashed green lines) provided for comparison.

Table 1: Medians of results from MEE simulation study for Pareto type I margins in three dimensions under independence (top) and comonotonicity (bottom) for various sample sizes n . Specifically, $n \in \{50, 100, 250, 500, 1000, 2500, 5000\}$ and $X_i \sim P\{3.5, 1.25(1+i)\}$, $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$. Standard deviations for the estimates are presented in parentheses.

(a) Independent Margins						
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.088 (0.056)	1.811 (0.382)	2.726 (0.588)	14.814 (9.370)	27.332 (20.79)	40.634 (36.25)
100	0.085 (0.039)	1.769 (0.264)	2.629 (0.426)	14.438 (5.674)	25.757 (11.88)	38.158 (19.61)
250	0.081 (0.028)	1.771 (0.189)	2.660 (0.304)	14.210 (3.700)	25.359 (7.075)	37.745 (11.78)
500	0.078 (0.017)	1.758 (0.132)	2.660 (0.211)	13.725 (2.488)	24.333 (4.833)	36.347 (7.912)
1000	0.076 (0.012)	1.770 (0.110)	2.644 (0.160)	13.786 (1.749)	24.466 (3.529)	36.468 (5.638)
2500	0.076 (0.008)	1.766 (0.073)	2.639 (0.115)	13.682 (1.105)	24.308 (2.169)	36.117 (3.536)
5000	0.075 (0.006)	1.765 (0.052)	2.639 (0.091)	13.636 (0.834)	24.060 (1.652)	36.006 (2.701)
Θ	0.074	1.764	2.639	13.545	23.894	35.744
(b) Comonotonic Margins						
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.371 (0.137)	1.525 (0.006)	2.062 (0.016)	22.523 (20.60)	34.338 (31.81)	46.446 (43.49)
100	0.375 (0.099)	1.519 (0.004)	2.049 (0.009)	21.737 (11.68)	33.028 (17.89)	44.534 (24.29)
250	0.385 (0.072)	1.514 (0.002)	2.036 (0.005)	22.109 (7.243)	33.485 (11.03)	45.029 (14.89)
500	0.382 (0.054)	1.512 (0.001)	2.029 (0.003)	21.580 (4.952)	32.618 (7.514)	43.789 (10.12)
1000	0.387 (0.041)	1.509 (0.001)	2.024 (0.002)	21.797 (3.555)	32.900 (5.383)	44.114 (7.238)
2500	0.391 (0.028)	1.507 (0.000)	2.019 (0.001)	21.834 (2.277)	32.908 (3.441)	44.069 (4.619)
5000	0.392 (0.022)	1.506 (0.000)	2.016 (0.001)	21.810 (1.761)	32.845 (2.658)	43.976 (3.563)
Θ	0.400	1.500	2.000	21.933	32.899	43.865

Five-dimensional case

175 The five-dimensional case is summarized in Table 2. It can be seen that the algorithm scales well to the increased dimension. This could be explained in part by the fact we are only ever dealing with pairwise dependencies through the pairwise upper tail dependence function λ^{ik} . The results presented focus on the independent five-dimensional case, though results for comonotonic random vectors (not included here) were very similar.

4.2. Pareto–Clayton model

In this section we present a simulation study using a non-trivial tail dependence structure. Specifically, we consider a random vector \mathbf{X} with Pareto type I margins, as defined previously, whose dependence structure is defined by a survival Clayton copula. The interested reader is referred to Cuberos et al. [14], Maume-Deschamps et al. [40], Yeh [54]. Recall that for a pair of standard uniform rv's (U, V) with a survival Clayton copula, we get

$$\bar{C}(u, v; \theta_C) = \left\{ \max(u^{-\theta_C} + v^{-\theta_C} - 1, 0) \right\}^{-1/\theta_C},$$

where the parameter $\theta_C \in [-1, \infty) \setminus \{0\}$ determines the strength of the dependence in the upper tail. In particular, the pairwise upper tail dependence function can now be defined as

$$\lambda(x, y; \theta_C) = (x^{-\theta_C} + y^{-\theta_C})^{-1/\theta_C}.$$

For a general development of MEEs for Archimedean copulas, see Maume-Deschamps et al. [42]. Notice that for this Pareto–Clayton model, it follows that

$$\lambda\left(\frac{c_i}{c_k} t^{-\theta}, 1; \theta_C\right) = \left\{ \left(\frac{c_i}{c_k}\right)^{-\theta_C} t^{-\theta\theta_C} + 1 \right\}^{-1/\theta_C}.$$

Table 2: Medians of results from MEE simulation study for Pareto type I margins in five dimensions linked with the independence copula for various sample sizes n . Specifically, $n \in \{50, 100, 250, 500, 1000, 2500, 5000\}$ and $X_i \sim P(3.5, 1.25(1+i))$, $i \in \{1, \dots, 5\}$ with $\ell_\theta = \ell_i = n^{0.75}$, $\ell_\lambda = n^{0.50}$. Standard deviations for the estimates are presented in parentheses.

(a) Parameter estimates.					
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
50	0.040 (0.058)	1.779 (0.773)	2.672 (1.159)	3.641 (1.405)	4.662 (1.887)
100	0.037 (0.033)	1.760 (0.404)	2.622 (0.602)	3.553 (0.851)	4.565 (1.136)
250	0.032 (0.017)	1.779 (0.288)	2.674 (0.410)	3.667 (0.547)	4.783 (0.757)
500	0.032 (0.012)	1.783 (0.205)	2.668 (0.300)	3.661 (0.410)	4.690 (0.545)
1000	0.031 (0.008)	1.782 (0.142)	2.653 (0.206)	3.661 (0.286)	4.723 (0.386)
2500	0.031 (0.005)	1.769 (0.109)	2.653 (0.148)	3.638 (0.210)	4.695 (0.282)
5000	0.030 (0.004)	1.763 (0.073)	2.645 (0.108)	3.618 (0.154)	4.667 (0.205)
Θ	0.029	1.764	2.639	3.607	4.655

(b) Expectiles estimates.					
n	$\hat{e}_\alpha^1(X)$	$\hat{e}_\alpha^2(X)$	$\hat{e}_\alpha^3(X)$	$\hat{e}_\alpha^4(X)$	$\hat{e}_\alpha^5(X)$
50	12.301 (7.676)	21.755 (22.48)	33.493 (31.78)	46.092 (43.99)	58.649 (53.44)
100	11.546 (4.261)	19.608 (8.535)	29.990 (12.93)	41.265 (19.67)	53.211 (26.50)
250	10.968 (2.510)	19.893 (5.057)	29.810 (7.971)	41.149 (11.05)	52.947 (15.67)
500	10.761 (1.870)	19.108 (3.689)	28.967 (5.726)	38.747 (8.207)	50.317 (11.57)
1000	10.653 (1.261)	19.185 (2.324)	28.377 (3.698)	38.982 (5.352)	50.274 (7.274)
2500	10.660 (0.784)	18.776 (1.591)	28.107 (2.491)	38.583 (3.622)	49.900 (4.956)
5000	10.581 (0.628)	18.653 (1.108)	27.999 (1.823)	38.329 (2.632)	49.508 (3.671)
$e_{0.9998}(X)$	10.390	18.330	27.420	37.475	48.372

For simplicity, and to facilitate closed form solutions to the system of equations in Proposition 1, we consider the case where $\theta_C = \theta^{-1}$. Then, the integral of λ^{ik} can be written as

$$\int_{\frac{\beta_i}{\beta_k}}^{\infty} \lambda \left(\frac{c_i}{c_k} t^{-\theta}, 1; \theta^{-1} \right) dt = \frac{1}{\theta-1} \frac{c_i}{c_k} \left\{ \frac{\beta_i}{\beta_k} + \left(\frac{c_i}{c_k} \right)^{1/\theta} \right\}^{-\theta+1},$$

giving a system of equations

$$\frac{1}{\theta-1} - \eta \frac{\beta_k^\theta}{c_k} = - \sum_{i \neq k} \left[\frac{1}{\theta-1} \frac{c_i}{c_k} \left\{ \frac{\beta_i}{\beta_k} + \left(\frac{c_i}{c_k} \right)^{1/\theta} \right\}^{-\theta+1} - \eta \frac{\beta_k^{\theta-1}}{c_k} \beta_i \right], \quad k \in \{1, \dots, d\}.$$

We now consider two cases: (i) $d = 2$, $\theta = 2$, $c_2 > 0$, and (ii) $d \geq 2$, $\theta > 1$, $c_i = 1$, $i \in \{1, \dots, d\}$. In case (i) and (ii), the closed-form MEEs system solutions are given by

$$\Theta = \left(\frac{1 + c_2(c_2^{3/4} + \sqrt{c_2})^{-1}}{c_2^{3/4} + 1}, c_2^{3/4} \right) \quad \text{and} \quad \Theta = \left(\frac{(d-1)2^{-\theta+1} + 1}{d(\theta-1)}, 1, \dots, 1 \right),$$

180 respectively. In the latter case, this solution makes heuristic sense. Having equivalent margins with a unit tail ratio would imply that they are identically distributed. As such, one would expect that the components of the MEEs are identical and solely based on the dependence parameter θ . In Table 3, we present the results of our simulation for each case. Here we see that performance is again satisfactory. In both cases our estimation procedure is able to accurately capture the true value of MEEs. An additional study of MEEs for the Pareto–Clayton model can be found in the
185 Supplementary material Section (see Section 7.1). In particular, we study their behavior for varying θ_C .

Table 3: Medians of results from MEE simulation study for Pareto type I margins linked with a survival Clayton copula for different dimensions d , dependence levels θ_C , and various sample sizes n . Specifically, $n \in \{50, 100, 250, 500, 1000, 2500, 5000\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$. Standard deviations for the estimates are presented in parentheses.

(a) Case (i), with $X_i \sim P(2, 1.25(1+i))$, $i \in \{1, 2\}$. This implies $d = 2$, $\theta_C = 1/2$ and $c_2 = 1.25$.

n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$
50	0.468 (0.802)	1.873 (9.347)	140.525 (1241.8)	304.536 (5744.1)
100	0.480 (0.645)	1.872 (1.175)	125.333 (678.19)	276.406 (826.40)
250	0.508 (0.399)	1.895 (1.249)	120.870 (189.65)	251.578 (332.57)
500	0.520 (0.313)	1.927 (0.776)	123.610 (99.534)	253.192 (179.82)
1000	0.580 (0.255)	1.848 (0.620)	126.894 (59.084)	247.724 (93.983)
2500	0.574 (0.213)	1.883 (0.517)	128.624 (42.478)	252.158 (64.426)
5000	0.575 (0.184)	1.857 (0.491)	129.376 (32.306)	250.832 (49.495)
Θ	0.590	1.837	135.798	249.477

(b) Case (ii), with $X_i \sim P(2, 2.5)$, $i \in \{1, 2, 3\}$. This implies $d = 3$, $\theta_C = 1/2$, and $c_i = 1$, $i \in \{1, 2, 3\}$

n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.574 (1.025)	0.989 (1.514)	1.024 (0.905)	151.343 (1689.1)	147.757 (1868.5)	144.849 (1057.6)
100	0.608 (0.703)	0.982 (0.760)	1.003 (0.809)	149.258 (562.22)	137.159 (386.92)	144.054 (411.26)
250	0.591 (0.444)	1.042 (0.502)	1.010 (0.555)	128.892 (152.95)	131.739 (125.49)	131.078 (134.63)
500	0.659 (0.386)	0.992 (0.464)	0.992 (0.416)	142.214 (115.13)	143.754 (104.74)	137.537 (102.22)
1000	0.677 (0.319)	0.990 (0.389)	0.967 (0.363)	141.696 (76.800)	138.610 (73.575)	136.954 (67.519)
2500	0.668 (0.252)	0.981 (0.323)	0.997 (0.323)	144.641 (46.530)	139.980 (44.727)	136.499 (43.140)
5000	0.660 (0.210)	1.002 (0.290)	0.992 (0.281)	140.850 (35.269)	142.948 (36.241)	141.022 (34.644)
Θ	0.667	1.000	1.000	144.338	144.338	144.338

4.3. Comparison with alternative methods

To further substantiate the performance of our estimation procedure, we compare the results in Table 1 to simulations using the fully parametric estimators

$$\hat{\eta}^\pm = \frac{1}{(\hat{\theta} - 1) \left(1 + \sum_{j=2}^d \hat{c}_j^{\frac{1}{\hat{\theta}-1}}\right)}, \quad \hat{\beta}_k^\pm = \hat{c}_k^{\frac{1}{\hat{\theta}-1}} \quad \text{and} \quad \hat{\eta}^\pm = \frac{1}{\hat{\theta} - 1}, \quad \hat{\beta}_k^\pm = \hat{c}_k^{\frac{1}{\hat{\theta}}} \quad (18)$$

for $k \in \{2, \dots, d\}$, based on (5) and (6), respectively, from Maume-Deschamps et al. [42]. Using these fully parametric estimators for the system solutions, estimators for MEEs can be written

$$\hat{e}_\alpha^\pm(\mathbf{X}) = \widehat{\text{VaR}}_\alpha(X_1) \left(\frac{1}{\hat{\theta} - 1}\right)^{\frac{1}{\hat{\theta}}} \left(\frac{1}{1 + \sum_{i=2}^d \hat{c}_i^{\frac{1}{\hat{\theta}-1}}}\right) \left(1, \hat{c}_2^{\frac{1}{\hat{\theta}-1}}, \dots, \hat{c}_d^{\frac{1}{\hat{\theta}-1}}\right)^\top \quad \text{and} \quad \hat{e}_\alpha^\pm(\mathbf{X}) = \widehat{\text{VaR}}_\alpha(X_1) \left(\frac{1}{\hat{\theta} - 1}\right)^{\frac{1}{\hat{\theta}}} \left(1, \hat{c}_2^{\frac{1}{\hat{\theta}}}, \dots, \hat{c}_d^{\frac{1}{\hat{\theta}}}\right)^\top$$

for independent and comonotonic random vectors, respectively, where $\widehat{\text{VaR}}_\alpha(X_1)$ is the Weissman's estimator [51] in (17). Analogously to Table 1, in Table 4 we consider median parameter estimates from the simulation study of independent and comonotonic Pareto margins in three dimensions. Standard deviations for the estimates are presented in parentheses. Here, as in Section 4.1, $X_i \sim P\{3.5, 1.25(1+i)\}$, $i \in \{1, 2, 3\}$.

Unsurprisingly, the fully parametric estimators perform better in particular for smaller sample sizes and have smaller standard errors. However, our semi-parametric method (see Table 1) does not require any pre-imposed dependence assumption and still provides satisfactory results. Note that the standard errors being zero for $\hat{\beta}_2$ and $\hat{\beta}_3$ for the comonotonic case are a result of the estimators structure and the distributional assumptions, *i.e.*,

$$\hat{\beta}_i^+ = \hat{c}_i^{\frac{1}{\hat{\theta}}} = \frac{X_{i:n-\ell_i+1,n}}{X_{i:n-\ell_i+1,n}} = \frac{F_i^{-1}(U_{n-\ell_i+1,n})}{F_1^{-1}(U_{n-\ell_i+1,n})} = \frac{b_i}{b_1}$$

for $i \in \{2, 3\}$, where $U_{n-\ell_i+1,n}$ is the $n - \ell_i + 1, n$ th order statistic from the standard uniform sample used to generate the data. For comonotonic random vectors, each realization of the copula is generated using the same uniform realization for each component, *i.e.*, $X_{ji} = F_j^{-1}(U_i)$, $j \in \{1, \dots, d\}$, $i \in \{1, \dots, n\}$.

In the full non-parametric framework, Maume-Deschamps et al. [41] use Robbins-Monro's stochastic optimization to solve the system of equations

$$\alpha \sum_{i=1}^d \mathbb{E}[(X_i - x_i)_+ \mathbb{1}\{X_k > x_k\}] = (1 - \alpha) \sum_{i=1}^d \mathbb{E}[(X_i - x_i)_- \mathbb{1}\{X_k < x_k\}], \quad k \in \{1, \dots, d\}.$$

195 For a comparison with our results under the same independent Pareto model, the interested reader is referred to Fig. 7 in Maume-Deschamps et al. [41] (page 40). It is easy to see that for elevated levels of α , the non-parametric approach solved through Robbins-Monro's stochastic optimization performs quite poorly.

Table 4: Medians of results from MEE simulation study for independent (top) and comonotonic (bottom) random vectors using the full parametric estimators for Θ and e_α in (18) with Pareto type I margins and various sample sizes n . Specifically, $n \in \{50, 100, 250, 500, 1000, 2500, 5000\}$ and $X_i \sim P\{3.5, 1.25(1 + i)\}$, $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$, $\ell_\lambda = n^{0.50}$. Standard deviations for the estimates are presented in parentheses.

(a) Independent Margins.						
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.075 (0.022)	1.794 (0.231)	2.672 (0.407)	13.917 (8.231)	25.373 (18.82)	38.092 (31.18)
100	0.074 (0.016)	1.770 (0.167)	2.645 (0.304)	13.569 (5.119)	24.443 (10.67)	36.305 (17.80)
250	0.075 (0.012)	1.770 (0.124)	2.656 (0.214)	13.736 (3.293)	24.610 (6.598)	36.895 (11.03)
500	0.074 (0.009)	1.765 (0.094)	2.639 (0.157)	13.540 (2.281)	23.933 (4.580)	35.459 (7.501)
1000	0.074 (0.007)	1.767 (0.075)	2.636 (0.126)	13.588 (1.626)	23.915 (3.340)	35.931 (5.452)
2500	0.074 (0.005)	1.767 (0.053)	2.637 (0.089)	13.560 (1.051)	23.957 (2.108)	35.685 (3.441)
5000	0.074 (0.004)	1.762 (0.042)	2.637 (0.074)	13.530 (0.804)	23.903 (1.602)	35.736 (2.667)
Θ	0.074	1.764	2.639	13.544	23.894	35.744
(b) Comonotonic Margins.						
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(\mathbf{X})$	$\hat{e}_\alpha^2(\mathbf{X})$	$\hat{e}_\alpha^3(\mathbf{X})$
50	0.411 (0.151)	1.500 (0)	2.000 (0)	23.228 (21.68)	34.843 (32.53)	46.457 (43.37)
100	0.404 (0.106)	1.500 (0)	2.000 (0)	22.222 (12.08)	33.333 (18.12)	44.444 (24.16)
250	0.407 (0.076)	1.500 (0)	2.000 (0)	22.460 (7.411)	33.690 (11.12)	44.921 (14.82)
500	0.399 (0.057)	1.500 (0)	2.000 (0)	21.842 (5.040)	32.763 (7.560)	43.684 (10.08)
1000	0.401 (0.043)	1.500 (0)	2.000 (0)	22.006 (3.605)	33.009 (5.407)	44.011 (7.210)
2500	0.401 (0.029)	1.500 (0)	2.000 (0)	21.980 (2.303)	32.970 (3.454)	43.960 (4.606)
5000	0.400 (0.023)	1.500 (0)	2.000 (0)	21.939 (1.776)	32.908 (2.664)	43.878 (3.552)
Θ	0.400	1.500	2.000	21.933	32.899	43.865

5. Discussion

As mentioned in Section 2, in the general case the loss function in (7) could be non-convex, and thus, might not satisfy the assumptions required for global convergence of the BFGS algorithm. Nevertheless, we chose this method for its relative simplicity and ease of implementation, as well as the satisfactory numerical results it has provided for our problem. To build a more rigorous estimation procedure, one could consider implementing methods which build upon the foundation of the original BFGS algorithm. For instance, Li and Fukushima [36, 37] discuss modifications to the BFGS algorithm which allow one to relax the assumption of convexity and still ensure global convergence. Specifically, they discuss the importance of Lipschitz continuity and using a damping method on the iterates of the approximate inverse Hessian matrix. Equivalently, Yuan et al. [55] discuss modifications to the Wolfe conditions in Algorithm 1 which ensures global convergence of BFGS for general functions. Such methods could further improve

our results. However, while BFGS is already included in many R packages, the above methods are not readily available. Moreover, manually programming them and doing so efficiently would require a specific work which is beyond the scope of the present work.

210 Additionally, while the BFGS algorithm works well for our problem, it was designed for unconstrained optimization problems. Clearly for extreme multivariate expectiles it is required that $\Theta > \mathbf{0}$ componentwise. A further improvement to this problem would be incorporating a box-constrained methodology for solving (8) and, more importantly, (9). One such choice would be box-constrained BFGS algorithm (or BFGS-B). This algorithm would allow us to incorporate properties of MEEs. For instance, as comonotonicity and independence represent the limiting cases of
 215 dependence in the upper tail, all expectiles will be bounded above and below by comonotonic and independent expectiles. Moreover, from Lemmas 2.4 and 2.5 Maume-Deschamps et al. [42] we know that for asymptotically independent and comonotonic random vectors the solutions to (8) are Θ^\perp and Θ^+ defined by (5) and (6), respectively.

As these situations represent the extreme cases of MEEs, this allows one to bound any solution Θ^* of (8) to the d -dimensional box $[\Theta_1^L, \Theta_1^U] \times \dots \times [\Theta_d^L, \Theta_d^U]$ where $\Theta_i^L = \min\{\Theta_i^\perp, \Theta_i^+\}$ and $\Theta_i^U = \max\{\Theta_i^\perp, \Theta_i^+\}$ for $i \in \{1, \dots, d\}$. However, because the estimation is done without prior knowledge of the underlying distribution, the box-constraints in the algorithm must also be estimated. To this end, the bounds of each box will be slightly relaxed. In other words, the algorithm would look for a solution in

$$[\widehat{\Theta}_1^L, \widehat{\Theta}_1^U] \times \dots \times [\widehat{\Theta}_d^L, \widehat{\Theta}_d^U],$$

where $\widehat{\Theta}_i^L = (1 - \gamma_L) \min\{\widehat{\Theta}_i^\perp, \widehat{\Theta}_i^+\}$ and $\widehat{\Theta}_i^U = (1 + \gamma_U) \max\{\widehat{\Theta}_i^\perp, \widehat{\Theta}_i^+\}$ for some $\gamma_L, \gamma_U > 0$.

Another change that could be made to the general BFGS algorithm is to incorporate limited memory storage of the
 220 inverse hessian H_k . This would prove most beneficial when the dimension of the problem is large and we would want to limit the number of operations on large inverse Hessian matrices. Combining these two modifications gives the limited-memory box-constrained BFGS algorithm (L-BFGS-B). Preliminary simulations using L-BFGS-B algorithm applied to (9) were conducted and the results were encouraging. Indeed, incorporating this methodology could provide further benefits in terms of both accuracy and computational efficiency. For more information on BFGS, L-BFGS or
 225 L-BFGS-B algorithms see Nocedal and Wright [44], Liu and Nocedal [38] and Byrd et al. [8]. Incorporating these changes along with the modifications described in Li and Fukushima [36, 37] and Yuan et al. [55] would also be interesting.

6. Conclusions

In this paper, we have presented an effective semi-parametric method for accurately estimating multivariate expectiles
 230 for extreme levels of risk $\alpha \approx 1$. Using the empirical beta copula, taking advantage of the fact that it is a genuine copula, we are able to establish the consistency of the approximate loss function in (9) when using empirical estimators for the tail index, tail ratio and upper tail dependence function. We have also proposed a new consistent estimator for the integral of the upper tail dependence function. Numerical simulations demonstrate the satisfactory finite-sample performance of our estimation scheme when compared to the true MEEs stemming from a variety of dependence
 235 structures, where closed form solutions exist. Furthermore, the set-up provided by the system of equations in Proposition 1 allows us to mitigate the curse of dimensionality as we are only ever required to consider pairwise dependencies. As such, the number of operations required for each iteration of the optimization algorithm will be $\binom{d}{2}$. In the future, it would be interesting to compare the performance of the algorithm for a variety of estimators. For instance, if one were to consider the checkerboard copula instead of the empirical beta copula. Finally, while finite-sample results are encouraging, the estimation procedure could be further streamlined by considering an algorithm which could take
 240 advantage of additional properties of our optimization problem. For example, L-BFGS-B algorithm could improve estimation by bounding solutions to (9) by the MEEs under the limiting cases of upper tail dependence.

Furthermore the present work could open interesting perspectives in functional statistics field. Indeed, the univariate well known conditional expectile/quantile model (see, *e.g.*, Daouia et al. [15], Usseglio-Carleve [48], Girard et al. [27]) can be adapted to our semi-parametric estimation scheme for new conditional multidimensional L^1 -expectile from heavy-tailed distributions when functional covariate information is available. The interested reader is also referred for instance to Gardes et al. [24], Gardes and Stupfler [25] where functional non-parametric estimators for conditional extreme quantiles are proposed.

More precisely, let (X_i, Z_i) , $i = 1, \dots, n$, be n independent copies of a random pair $(\mathbf{X}, \mathbf{Z}) \in \mathbb{R}^d \times \mathcal{E}$ where (\mathcal{E}, d) is a not necessarily finite-dimensional Polish space endowed with a semi-metric d . For instance, \mathcal{E} can be the standard p -dimensional space \mathbb{R}^p , a space of continuous functions over a compact metric space, or a Lebesgue space $L_p(\mathbb{R})$, to name a few. From Equation (2), a future development of this work would be to consider the functional conditional multidimensional L^1 -expectile extension given by

$$e_\alpha^1(\mathbf{X}, z) := \arg \min_{\mathbf{x} \in \mathbb{R}^d} \mathbb{E} \left\{ \alpha \left(\sum_{i=1}^d |X_i - x_{i+}| \right)^2 + (1 - \alpha) \left(\sum_{i=1}^d |X_i - x_{i-}| \right)^2 \middle| \mathbf{Z} = z \right\}.$$

Then the study in the present paper could be adapted to estimate extreme $e_\alpha^1(\mathbf{X}, z)$ by using the extrapolation technique when $\alpha \rightarrow 1$ together with a nonparametric estimation of the conditional tail copula (see, *e.g.*, Gardes and Girard [23]). The main objective of this semi-parametric approach for extreme functional $e_\alpha^1(\mathbf{X}, z)$ would be to balance the trade-off between the high sensitivity to dimension of the nonparametric models and the relative lack of flexibility of the completely parametric models (see, *e.g.*, Goia and Vieu [29]).

Acknowledgments: The authors express their gratitude to the Associate Editor and the three anonymous Referees for their valuable comments which have helped to greatly improve the quality of this manuscript. The authors would also like to thank Professor Christian Genest for providing several insightful comments during the earlier stages of this research. This work has been partially supported by the SIMONS Foundation, Society of Actuaries James C. Hickman Scholar Program and the Centre de recherche mathématiques de Montréal through the Simons-CRM Program. This work has been supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002.

References

- [1] C. Acerbi, D. Tasche, Expected shortfall: A natural coherent alternative to value at risk, *Economic Notes* 31 (2002) 379–388.
- [2] C. Acerbi, D. Tasche, On the coherence of expected shortfall, *Journal of Banking & Finance* 26 (2002) 1487–1503.
- [3] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, *Mathematical Finance* 9 (1999) 203–228.
- [4] N. Beck, M. Mailhot, A consistent estimator to the orthant-based tail value-at-risk, *ESAIM: Probability and Statistics* 22 (2018) 163–177.
- [5] F. Bellini, V. Bignozzi, On elicitable risk measures, *Quantitative Finance* 15 (2015) 725–733.
- [6] F. Bellini, E. Di Bernardino, Risk management with expectiles, *The European Journal of Finance* 23 (2017) 487–506.
- [7] F. Bellini, B. Klar, A. Müller, E. R. Gianin, Generalized quantiles as risk measures, *Insurance: Mathematics and Economics* 54 (2014) 41–48.
- [8] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM Journal on Scientific Computing* 16 (1995) 1190–1208.
- [9] J.-J. Cai, J. H. Einmahl, L. Haan, C. Zhou, Estimation of the marginal expected shortfall: the mean when a related variable is extreme, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77 (2015) 417–442.
- [10] M. Chaouch, C. Goga, Design-based estimation for geometric quantiles with application to outlier detection, *Computational Statistics & Data Analysis* 54 (2010) 2214–2229.
- [11] H. Cossette, M. Mailhot, É. Marceau, M. Mesfioui, Bivariate lower and upper orthant value-at-risk, *European Actuarial Journal* 3 (2013) 321–357.
- [12] H. Cossette, M. Mailhot, E. Marceau, M. Mesfioui, Vector-valued tail value-at-risk and capital allocation, *Methodology and Computing in Applied Probability* 18 (2016) 653–674.
- [13] A. Cousin, E. Di Bernardino, On multivariate extensions of value-at-risk, *Journal of Multivariate Analysis* 119 (2013) 32–46.
- [14] A. Cuberos, E. Masiello, V. Maume-Deschamps, Copulas checker-type approximations: Application to quantiles estimation of sums of dependent random variables, *Communications in Statistics - Theory and Methods* 49 (2020) 3044–3062.
- [15] A. Daouia, L. Gardes, S. Girard, On kernel smoothing for extremal quantile regression, *Bernoulli* 19 (2013) 2557–2589.

- [16] A. Daouia, S. Girard, G. Stupfler, Estimation of tail risk based on extreme expectiles, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80 (2018) 263–292.
- 285 [17] A. Daouia, S. Girard, G. Stupfler, et al., Extreme M-quantiles as risk measures: From L^1 to L^p optimization, *Bernoulli* 25 (2019) 264–309.
- [18] P. Deheuvels, E. Haeusler, D. M. Mason, Almost sure convergence of the Hill estimator, *Mathematical Proceedings of the Cambridge Philosophical Society* 104 (1988) 371–381.
- [19] E. Di Bernardino, T. Laloë, V. Maume-Deschamps, C. Prieur, Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory, *ESAIM: Probability and Statistics* 17 (2013) 236–256.
- 290 [20] E. Di Bernardino, C. Prieur, Estimation of the multivariate conditional tail expectation for extreme risk levels: Illustration on environmental data sets, *Environmetrics* 29 (2018) e2510.
- [21] I. Ekeland, A. Galichon, M. Henry, Comonotonic measures of multivariate risks, *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 22 (2012) 109–132.
- [22] P. Embrechts, G. Puccetti, Bounds for functions of multivariate risks, *Journal of Multivariate Analysis* 97 (2006) 526–547.
- 295 [23] L. Gardes, S. Girard, Nonparametric estimation of the conditional tail copula, *Journal of Multivariate Analysis* 137 (2015) 1–16.
- [24] L. Gardes, S. Girard, A. Lekina, Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis* 101 (2010) 419 – 433. *Statistical Methods and Problems in Infinite-dimensional Spaces*.
- [25] L. Gardes, G. Stupfler, An integrated functional Weissman estimator for conditional extreme quantiles, *REVSTAT - Statistical Journal* 17(1) (2019) 109–144.
- 300 [26] S. Girard, G. Stupfler, Extreme geometric quantiles in a multivariate regular variation framework, *Extremes* 18 (2015) 629–663.
- [27] S. Girard, G. Stupfler, A. Usseglio-Carleve, Nonparametric extreme conditional expectile estimation, 2019. Working paper or preprint.
- [28] T. Gneiting, Making and evaluating point forecasts, *Journal of the American Statistical Association* 106 (2011) 746–762.
- [29] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *Journal of Multivariate Analysis* 146 (2016) 1 – 6. Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.
- 305 [30] L. de Haan, A. Ferreira, *Extreme Value Theory: An Introduction.*, Springer Series in Operations Research and Financial Engineering. New York., 2006.
- [31] K. Herrmann, M. Hofert, M. Mailhot, Multivariate geometric expectiles, *Scandinavian Actuarial Journal* 2018 (2018) 629–659.
- [32] K. Herrmann, M. Hofert, M. Mailhot, Multivariate geometric tail-and range-value-at-risk, *ASTIN Bulletin: The Journal of the IAA* 50 (2020) 265–292.
- 310 [33] B. M. Hill, A simple general approach to inference about the tail of a distribution, *The Annals of Statistics* 3 (1975) 1163–1174.
- [34] A. Kiriliouk, J. Segers, L. Tafakori, An estimator of the stable tail dependence function based on the empirical beta copula, *Extremes* 21 (2018) 581–600.
- [35] H. Laniado, R. Lillo, J. Romo, Multivariate extremality measure, Preprint (2010) 10–19.
- [36] D.-H. Li, M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, *Journal of Computational and Applied Mathematics* 129 (2001) 15–35.
- 315 [37] D.-H. Li, M. Fukushima, On the global convergence of the BFGS method for nonconvex unconstrained optimization problems, *SIAM Journal on Optimization* 11 (2001) 1054–1064.
- [38] D. C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, *Mathematical Programming* 45 (1989) 503–528.
- [39] T. Mao, K. Ng, T. Hu, Asymptotic expansions of generalized quantiles and expectiles for extreme risks, *Probability in the Engineering and Informational Sciences* 29 (2015) 309–327.
- 320 [40] V. Maume-Deschamps, D. Rullière, K. Said, Impact of dependence on some multivariate risk indicators, *Methodology and Computing in Applied Probability* 19 (2017) 395–427.
- [41] V. Maume-Deschamps, D. Rullière, K. Said, Multivariate extensions of expectiles risk measures, *Dependence Modeling* 5 (2017) 20–44.
- [42] V. Maume-Deschamps, D. Rullière, K. Said, Extremes for multivariate expectiles, *Statistics & Risk Modeling* 35 (2018) 111–140.
- 325 [43] W. Newey, J. L. Powell, Asymmetric least squares estimation and testing, *Econometrica* 55 (1987) 819–47.
- [44] J. Nocedal, S. Wright, *Numerical Optimization*, Springer Science & Business Media, 2006.
- [45] F. Riedel, Dynamic coherent risk measures, *Stochastic processes and their applications* 112 (2004) 185–200.
- [46] R. Schmidt, U. Stadtmüller, Non-parametric estimation of tail dependence, *Scandinavian Journal of Statistics* 33 (2006) 307–335.
- [47] R. Torres, R. E. Lillo, H. Laniado, A directional multivariate value at risk, *Insurance: Mathematics and Economics* 65 (2015) 111–123.
- 330 [48] A. Usseglio-Carleve, Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors, *Electronic Journal of Statistics* 12 (2018) 4057–4093.
- [49] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 2000.
- [50] A. W. van der Vaart, J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer, 1996.
- 335 [51] I. Weissman, Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association* 73 (1978) 812–815.
- [52] Y. Yamai, T. Yoshida, On the validity of value-at-risk: Comparative analyses with expected shortfall, *Monetary and Economic Studies* 20 (2002) 57–85.
- [53] Y. Yamai, T. Yoshida, Value-at-risk versus expected shortfall: A practical perspective, *Journal of Banking & Finance* 29 (2005) 997–1015.
- 340 [54] H.-C. Yeh, The frailty and the Archimedean structure of the general multivariate Pareto distributions, *Bulletin Institute of Mathematics Academia Sinica* 2 (2007) 713–729.
- [55] G. Yuan, Z. Wei, X. Lu, Global convergence of BFGS and PRP methods under a modified weak Wolfe–Powell line search, *Applied Mathematical Modelling* 47 (2017) 811–825.
- [56] J. F. Ziegel, Coherence and elicibility, *Mathematical Finance* 26 (2016) 901–918.

7. Supplementary material

345 7.1. Pareto–Clayton model: additional studies

Here we present the results of several extra simulations demonstrating the impact of dependence parameter θ_C on our results. Note that in general $\theta_C \in [-1, +\infty) \setminus \{0\}$ where the cases $\theta_C = -1$, $\theta_C \rightarrow 0$ and $\theta_C \rightarrow \infty$ correspond to countermonotonic, independent and comonotonic dependence structure, respectively. As such, we test our methodology by considering very small and large values of θ_C so as to see if we can recapture these limiting results presented in Section 4.1. Furthermore, we consider a range of values in between the limiting cases to study the behavior of MEEs as the dependence becomes stronger.

Once more, we consider $X_i \sim P(b_i, a)$ where $a = \theta = 3.5$ and $b_i = 1.25(i + 1)$, $i \in \{1, 2, 3\}$. Recall that for these settings the system solutions for independent and comonotonic random variables are

$$\Theta^\ominus = (0.0740, 1.764, 2.639), \quad \Theta^\oplus = (0.400, 1.500, 2.000),$$

respectively (see Section 4.1). For all simulations in this section, we consider a survival Clayton copula with dependence parameter $\theta_C \in \{0.01, 0.1, 0.25, 0.5, 0.75, 1, 2.5, 5, 7.5, 10, 15, 20, 100\}$. Finally, $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$.

Table 5: Medians of results from MEE simulation study for Pareto type I margins using a Clayton survival copula with dependence parameter $\theta_C = 0.01$ (top) and $\theta_C = 100$ (bottom) for various sample sizes. Specifically, $n \in \{50, 100, 250, 500, 1000, 2500, 5000\}$ and $X_i \sim P\{3.5, 1.25(1 + i)\}$, $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$. Standard deviations for the estimates are presented in parentheses.

(a) Simulation results for Pareto–Clayton model with $\theta_C = 0.01$. The final row Θ represents the known solutions under independence.						
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(X)$	$\hat{e}_\alpha^2(X)$	$\hat{e}_\alpha^3(X)$
50	0.096 (0.081)	1.725 (0.392)	2.591 (0.620)	15.664 (13.81)	27.564 (23.86)	42.011 (39.55)
100	0.085 (0.042)	1.775 (0.280)	2.638 (0.468)	14.506 (5.847)	26.275 (12.51)	39.345 (21.03)
250	0.081 (0.026)	1.757 (0.199)	2.654 (0.331)	14.118 (3.618)	24.970 (7.340)	37.527 (12.78)
500	0.079 (0.021)	1.765 (0.149)	2.644 (0.248)	14.046 (2.637)	24.862 (5.236)	37.157 (8.546)
1000	0.077 (0.012)	1.765 (0.107)	2.642 (0.160)	13.761 (1.633)	24.379 (3.277)	36.605 (5.130)
2500	0.077 (0.009)	1.767 (0.077)	2.643 (0.117)	13.838 (1.139)	24.359 (2.303)	36.345 (3.620)
5000	0.076 (0.007)	1.763 (0.060)	2.637 (0.088)	13.715 (0.845)	24.189 (1.680)	36.242 (2.637)
Θ	0.074	1.764	2.639	13.545	23.894	35.744
(b) Simulation results for Pareto–Clayton model with $\theta_C = 100$. The final row Θ represents the known solutions under comonotonicity.						
n	$\hat{\eta}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{e}_\alpha^1(X)$	$\hat{e}_\alpha^2(X)$	$\hat{e}_\alpha^3(X)$
50	0.344 (0.121)	1.536 (0.013)	2.089 (0.025)	21.480 (19.66)	32.878 (30.82)	44.503 (42.91)
100	0.363 (0.100)	1.529 (0.009)	2.071 (0.017)	21.767 (12.42)	33.267 (19.20)	45.242 (26.28)
250	0.367 (0.062)	1.521 (0.007)	2.054 (0.010)	21.600 (6.097)	32.843 (9.355)	44.331 (12.70)
500	0.362 (0.046)	1.517 (0.005)	2.044 (0.007)	20.651 (4.001)	31.298 (6.117)	42.124 (8.273)
1000	0.366 (0.036)	1.515 (0.004)	2.037 (0.006)	20.494 (3.126)	31.068 (4.764)	41.747 (6.429)
2500	0.365 (0.023)	1.511 (0.003)	2.029 (0.004)	20.138 (1.826)	30.400 (2.789)	40.751 (3.746)
5000	0.375 (0.018)	1.509 (0.003)	2.024 (0.002)	20.705 (1.445)	31.210 (2.174)	41.904 (2.941)
Θ	0.400	1.500	2.000	21.933	32.899	43.865

In Table 5, it can be seen that our estimation procedure accurately estimates the true system solutions for Pareto–Clayton models which are arbitrarily close to independence and comonotonicity. Moreover, in Fig. 6, we plot the boxplots for $\hat{\beta}_2$ from our simulations. We also provide the true values under independence and comonotonicity for comparison (green and red dotted lines in Fig. 6, respectively). Here, we are able to extract more insight into the behavior of MEEs as the dependence parameter shifts. For the Pareto–Clayton model, we see that the parameter estimates seem to decrease monotonically from those representing independence to those representing comonotonicity.

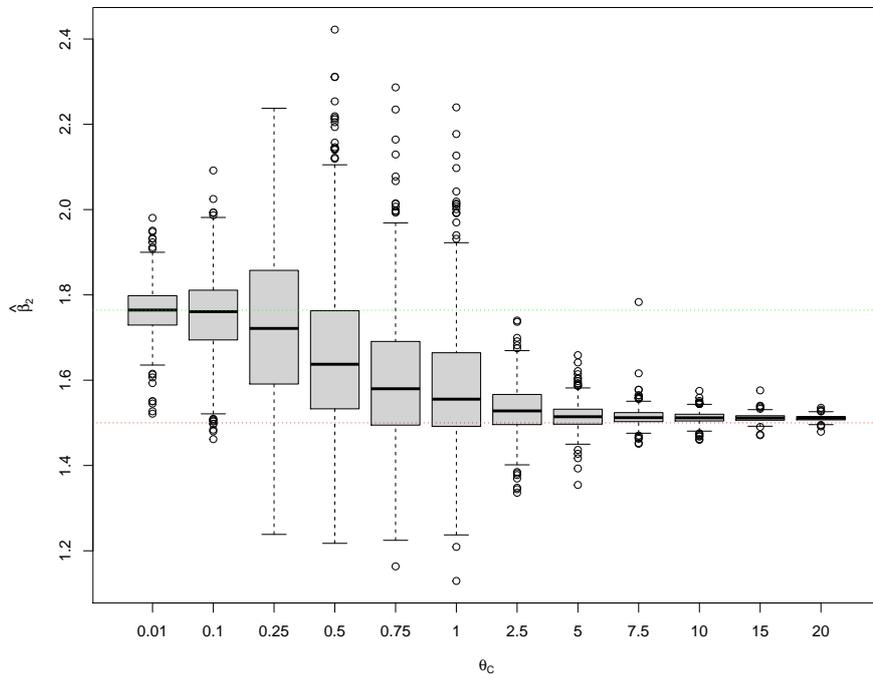


Fig. 6: Simulation results for the estimate of $\hat{\beta}_2$ under the Pareto–Clayton model with varying dependence parameter θ_C and Pareto type I margins. Specifically, $X_i \sim P\{3.5, 1.25(1 + i)\}$, $i \in \{1, 2, 3\}$ with $\ell_\theta = \ell_i = n^{0.75}$ and $\ell_\lambda = n^{0.50}$. Here the sample size is fixed at $n = 5000$ and $\theta_C \in \{0.01, 0.1, 0.25, 0.5, 0.75, 1, 2.5, 5, 7.5, 10, 15, 20\}$. Dotted lines provide true values for asymptotic independence (green) and comonotonicity (red) with $\beta_2^+ = 1.764$ and $\beta_2^* = 1.5$, respectively.

360 7.2. Data study

Here we present an application of MEEs to a real data set. In particular, we consider the Pima Indians Diabetes Database (<https://www.kaggle.com/uciml/pima-indians-diabetes-database>). This data set has been used in previous studies of multivariate risk measures, see, *e.g.*, Girard and Stupfler [26], Chaouch and Goga [10]. The data set consists of 9 variables (Pregnancies, Glucose, Diastolic Blood Pressure (DBP), Skin Thickness, Insulin Levels, 365 body mass index (BMI), Diabetes Pedigree Function, Age and Outcome) for $n = 768$ individuals from the Pima Indian population. The study was conducted in an attempt to understand the prevalence of diabetes in this population. We consider the two-dimensional case for pairs (X_{i1}, X_{i2}) where X_{i1} is the BMI and X_{i2} is the DBP for individual $i \in \{1, \dots, 768\}$.

The first step is to estimate $\hat{\theta}$ and \hat{c}_2 . Using the selection procedure described, for instance, by Cai et al. [9], we 370 consider a range of possibilities for intermediate sequences ℓ_θ and ℓ_2 for which the estimators showed some stability. Then, to gain robustness, we average the estimates corresponding to the selected ranges. The interested reader is also referred to Di Bernardino and Prieur [20]. For $\hat{\theta}$ this turned out to be values of $\ell_\theta \in [100, 150]$. The corresponding Hill plots are presented in Fig. 7. Considering these values the tail index was estimated to be $\hat{\theta} = 9.126$. The same procedure is followed for \hat{c}_2 and the resulting estimated value is $\hat{c}_2 = 1103.046$. In Fig. 8, the resulting expectiles are 375 plotted.

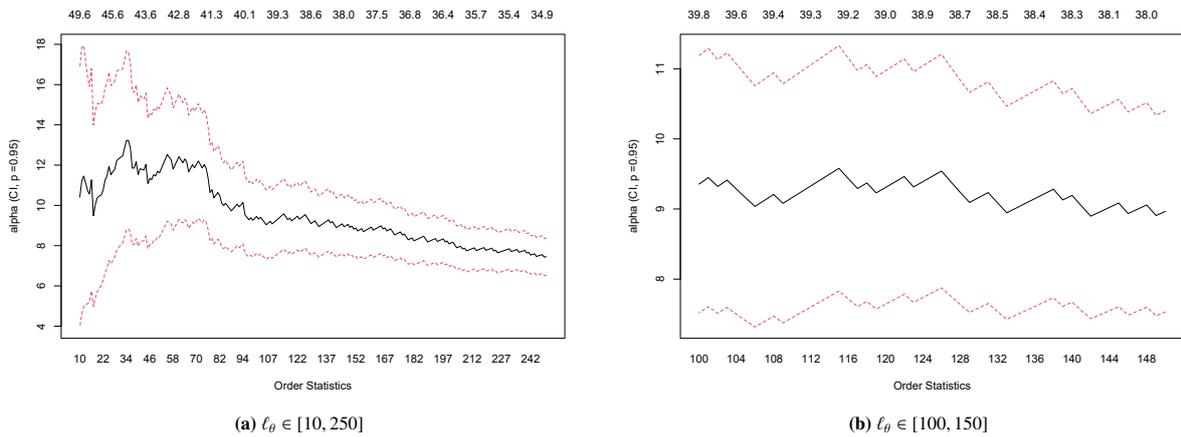


Fig. 7: Hill Plots for estimating $\hat{\theta}$ using BMI measurements. We consider a range of possibilities for intermediate sequences ℓ_θ for which the estimator $\hat{\theta}$ shows some stability. In left panel we consider a wide range where $\ell_\theta \in [10, 250]$ and the right panel shows the chosen estimation range where $\ell_\theta \in [100, 150]$.

Recalling the first order conditions of L^1 expectiles, $e_\alpha(\mathbf{X})$ is the value of \mathbf{x} such that

$$\frac{\sum_{i=1}^d \mathbb{E} [(X_i - x_i)_+ \mathbb{1}\{X_k > x_k\}]}{\sum_{i=1}^d \mathbb{E} [(X_i - x_i)_- \mathbb{1}\{X_k < x_k\}]} = \frac{1 - \alpha}{\alpha}, \quad k \in \{1, 2\}.$$

This could be interpreted as the value of \mathbf{x} which fixes the ratio of participation in positive scenarios to negative scenarios at $(1 - \alpha)/\alpha$ for all variables. Consider the MEE at level $\alpha = 1 - 1/768 = 0.9987$, we have $\hat{e}_{0.9987}(\mathbf{X}) = (45.433, 106.493)$. As was noted in similar studies of this data set, *e.g.*, Chaouch and Goga [10], an individual with a BMI above 31.7 kg/m² would qualify as severely overweight and a DBP value above 95 mm Hg would be at risk of hypertension. Both these factors are considered to impact the prevalence of diabetes. 380

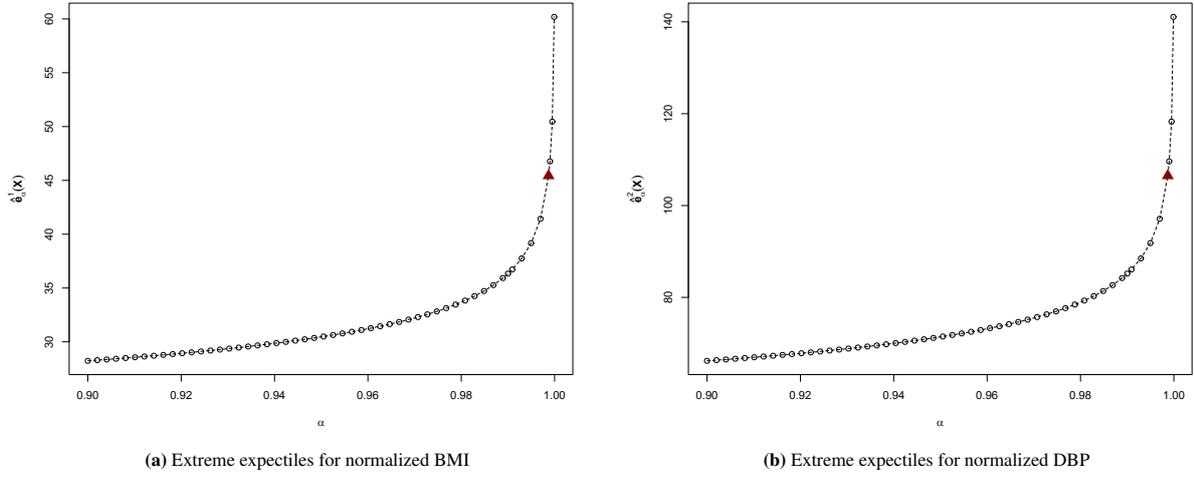


Fig. 8: Extreme expectiles ($\hat{\varrho}_\alpha^1(X)$, $\hat{\varrho}_\alpha^2(X)$) for pairs of BMI and DBP in the Pima Indians Diabetes Study for several risk values $\alpha \in [0.9, 1)$ (black dots). The extreme expectiles for $\alpha = 1 - 1/n \approx 0.9987$, (45.433, 106.493), are represented by red triangles.

7.3. Gradient $\nabla L_\Lambda(\cdot)$ in Algorithm 1

Recall that $\beta^* = (\beta_1, \dots, \beta_d)$ with $\beta_1 = 1$ and $\Theta = (\eta, \beta_2, \dots, \beta_d)$. Then, with our notation,

$$\Theta_j = \begin{cases} \eta, & j = 1 \\ \beta_j, & j \in \{2, \dots, d\} \end{cases}.$$

First, note that $\nabla L_\Lambda(\cdot)$ is given by

$$\begin{aligned} \nabla L_\Lambda(\cdot) &= \nabla \left[\frac{1}{2} \sum_{k=1}^d \{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\}^2 \right] \\ &= \begin{bmatrix} \sum_{k=1}^d [\{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\} \times \frac{\partial}{\partial \eta} \{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\}] \\ \sum_{k=1}^d [\{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\} \times \frac{\partial}{\partial \beta_2} \{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\}] \\ \vdots \\ \sum_{k=1}^d [\{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\} \times \frac{\partial}{\partial \beta_d} \{g_\Lambda^{(k)}(\cdot) + f_\Lambda^{(k)}(\cdot)\}] \end{bmatrix}. \end{aligned} \quad (19)$$

Then, the marginal derivatives in (19) can be written explicitly evaluated

$$\begin{aligned} \frac{\partial}{\partial \Theta_j} g_\Lambda^{(k)}(\Theta) &= \begin{cases} -\frac{\beta_k^\theta}{c_k}, & j = 1, \\ -\eta \theta \frac{\beta_k^{\theta-1}}{c_k}, & j = k \neq 1, \\ 0, & \text{otherwise,} \end{cases} \\ \frac{\partial}{\partial \Theta_j} f_\Lambda^{(k)}(\Theta) &= \begin{cases} -\frac{\beta_k^{\theta-1}}{c_k} \sum_{i \neq k} \beta_i, & j = 1, \\ \sum_{i \neq k} \left[\lambda^{ik} \left\{ \frac{c_i}{c_k} \left(\frac{\beta_i}{\beta_k} \right)^{-\theta}, 1 \right\} \frac{\beta_i}{\beta_k^2} - \eta \frac{(\theta-1)\beta_k^{\theta-2}}{c_k} \beta_i \right], & j = k \neq 1, \\ -\lambda^{jk} \left\{ \frac{c_j}{c_k} \left(\frac{\beta_j}{\beta_k} \right)^{-\theta}, 1 \right\} \frac{1}{\beta_k} - \eta \frac{\beta_k^{\theta-1}}{c_k}, & \text{otherwise,} \end{cases} \end{aligned}$$

$j, k \in \{1, \dots, d\}$.