

Clustered Archimax copulas

Simon Chatelain^{*1}, Samuel Perreault³,
Anne-Laure Fougères², and Johanna G. Nešlehová⁴

¹Geneva School of Economics and Management, Université de Genève,
e-mail: simon.chatelain.ev@gmail.com

²Université Claude Bernard Lyon 1, ICJ UMR5208, CNRS, École Centrale de Lyon, INSA
Lyon, Université Jean Monnet, 69622 Villeurbanne, France

³Department of Statistical Sciences, University of Toronto

⁴Department of Mathematics and Statistics, McGill University and Institute for Statistics
and Mathematics, Vienna University of Economics and Business

Abstract: When modeling multivariate phenomena, properly capturing the joint extremal behavior is often one of the many concerns. Archimax copulas appear as successful candidates in case of asymptotic dependence. In this paper, the class of Archimax copulas is extended via their stochastic representation to a clustered construction. These clustered Archimax copulas are characterized by a partition of the random variables into groups linked by a radial copula; each cluster is Archimax and therefore defined by its own Archimedean generator and stable tail dependence function. The proposed extension allows for both asymptotic dependence and independence between the clusters, a property which is sought, for example, in applications in environmental sciences and finance. The model also inherits from the ability of Archimax copulas to capture dependence between variables at pre-extreme levels. The asymptotic behavior of the model is established, leading to a rich class of stable tail dependence functions.

MSC2020 subject classifications: Primary 62H05, 60G70; secondary 62G10.

Keywords and phrases: Multivariate extremes, subasymptotic modeling, random scale constructions.

Received October 2022.

1. Introduction

The Archimax copula model introduced by Capéraà, Fougères and Genest (1997) has been advocated in Chatelain, Fougères and Nešlehová (2020) as a flexible way to model a group of variables whose asymptotic dependence is driven by a stable tail dependence function (stdf); or more precisely a random vector whose dependence structure follows an asymptotic extreme-value regime perturbed by the same distortion. However, as it is the case for rainfall over large territories for example, asymptotic independence between certain variables is likely to be present and this phenomenon cannot be handled by a single Archimax model without limiting the marginal dependence structure to be an (exchangeable)

arXiv: [2210.15622](https://arxiv.org/abs/2210.15622)

*Corresponding author.

Archimedean copula. In financial applications, a stock portfolio might contain stocks from the same industry, causing them to be dependent in the extreme regime, while stocks from different industries might be asymptotically independent. Likewise, assuming the same distortion for all variables may not be realistic when the number of variables is large.

We propose a dependence model based on the survival copula of a nonnegative, real random vector of the form

$$\mathbf{X} = (R_1 \mathbf{S}_1, \dots, R_K \mathbf{S}_K),$$

where for all $k \in \{1, \dots, K\}$, R_k are nonnegative random variables and \mathbf{S}_k are d_k -dimensional random vectors which will be characterized later. More specifically, each of the vectors $R_k \mathbf{S}_k = R_k \times (S_{k1}, \dots, S_{kd_k})$ will be defined so as to have an Archimax survival copula, while conditional independence of the vectors $\{\mathbf{S}_k\}_{k=1}^K$ given $\{R_k\}_{k=1}^K$ will introduce parsimony to the dependence model. This paper focuses on the survival copula of \mathbf{X} , which can subsequently be used in a copula model with arbitrary margins.

Stochastic representations involving a randomly scaled random vector are very common in the literature. For example, pseudo-polar representations exist for elliptical (see [Fang, Kotz and Ng \(1990\)](#)), generalized Pareto (see [Ferreira and de Haan \(2014\)](#)), Archimedean (see [McNeil and Nešlehová \(2009\)](#)) and Liouville (see [Belzile and Nešlehová \(2017\)](#)). Such constructions have garnered significant interest as models suited for extreme values; they can offer the flexibility needed to model across extremal dependence classes as shown in [Huser, Opitz and Thibaud \(2017\)](#), [Wadsworth et al. \(2017\)](#) and [Huser and Wadsworth \(2019\)](#). [Engelke, Opitz and Wadsworth \(2019\)](#) offer an extensive study of the extremal behavior of bivariate pseudo-polar vectors. As shown in [Charpentier et al. \(2014\)](#), Archimax copulas also allow a representation of the form $R\mathbf{S}$; the interpretation proposed by [Chatelain, Fougères and Nešlehová \(2020\)](#) is that of a radial variable R distorting the vector \mathbf{S} characterized by a stdf and therefore representing the extremal regime. The aim of this paper is to propose a dependence model in a way that its higher-dimensional margins are Archimax copulas but with possibly different distortions or stdfs. There is also a connection to be made with [Hofert, Huser and Prasad \(2018\)](#) which have extended the class of Archimax copulas with completely monotone Archimedean generators to hierarchical constructions. Therein, hierarchies can be introduced either via the frailties or stdfs, but the extremal behavior is not elicited.

In this paper, we define the new class of clustered Archimax copulas. Building a model from this family is straightforward as it only requires specifying the Archimax clusters and the dependence between the radial variables R_1, \dots, R_K . Under non-restrictive assumptions, the extremal behavior of the clustered Archimax copula is obtained and shown to be quite flexible. We find that different choices of popular Archimedean generators coupled with dependence models on the radial variables lead to various regimes falling under either asymptotic dependence or independence between clusters. Meanwhile, each Archimax cluster retains its flexibility in the extreme dependence regime. We also propose infer-

ence techniques to fit clustered Archimax copulas to data and illustrate with a data application on rainfall in France.

The paper is organized as follows. Section 2 contains preliminary notions; Section 3 defines the model; Section 4 presents results on its extremal behavior; Section 5 provides illustrative examples; Section 6 covers inference techniques and a simulation study; Section 7 contains the data application; finally, Section 8 concludes the paper with a discussion. Proofs are reported in Appendix A, while Appendix C formulates a conjecture extending Theorem 4.1.

2. Preliminaries

Copulas contain all the information pertaining to the dependence between the components of continuous random vectors. The decomposition of Sklar (1959) links the marginals and the joint distribution of a random vector via a copula, which is simply a distribution function on the unit hypercube with standard uniform margins. Consider $\mathbf{X} = (X_1, \dots, X_d) \sim F$ with margins F_1, \dots, F_d . Then there exists a copula C such that for any $(x_1, \dots, x_d) \in \mathbb{R}^d$,

$$F(x_1, \dots, x_d) = \Pr(X_1 \leq x_1, \dots, X_d \leq x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}.$$

Moreover, C is unique if the margins are all continuous. There is also a version of Sklar's decomposition for survival functions. That is, given the marginal survival functions $\bar{F}_1, \dots, \bar{F}_d$ and the joint marginal function \bar{F} , there exists a copula \bar{C} , called a survival copula of \mathbf{X} , such that for all $(x_1, \dots, x_d) \in \mathbb{R}^d$,

$$\bar{F}(x_1, \dots, x_d) = \Pr(X_1 > x_1, \dots, X_d > x_d) = \bar{C}\{\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)\}.$$

It can be shown that if $\mathbf{U} \sim C$, then $1 - \mathbf{U} \sim \bar{C}$ (see Nelsen (2006)). There is a large amount of literature on dependence modeling via copulas; we refer to the monographs of Nelsen (2006), Durante and Sempi (2010) and Joe (2015).

One particular family of copulas, called Archimax, was introduced in two dimensions by Capéraà, Fougères and Genest (2000), and extended to arbitrary dimensions by Mesiar and Jäger (2013) and Charpentier et al. (2014). Generalizing both Archimedean and extreme-value copula families, Archimax copulas take the form, for all $(u_1, \dots, u_d) \in (0, 1)$,

$$C_{\psi, \ell}(u_1, \dots, u_d) = \psi \circ \ell\{\phi(u_1), \dots, \phi(u_d)\}, \quad (2.1)$$

where ψ is an Archimedean generator with inverse ϕ and ℓ is a stdf. These two functional parameters are defined below.

Definition 2.1. A non-increasing and continuous function $\psi : [0, \infty) \rightarrow [0, 1]$ which satisfies $\psi(0) = 1$, $\lim_{x \rightarrow \infty} \psi(x) = 0$ and is strictly decreasing on $[0, x_\psi)$, where $x_\psi = \inf\{x : \psi(x) = 0\}$, is called an Archimedean generator.

A function $\ell : \mathbb{R}_+^d \rightarrow \mathbb{R}_+$ is called a d -variate stable tail dependence function (stdf) if there exists a finite measure μ on the d -dimensional unit simplex $\Delta_d =$

$\{\mathbf{w} \in [0, 1]^d : w_1 + \dots + w_d = 1\}$ satisfying $\int_{\Delta_d} s_j d\mu(\mathbf{s}) = 1$ for all $j \in \{1, \dots, d\}$, such that for all $\mathbf{x} \in \mathbb{R}_+^d$,

$$\ell(\mathbf{x}) = \int_{\Delta_d} \max(x_1 s_1, \dots, x_d s_d) d\mu(\mathbf{s}).$$

A d -dimensional copula C is called Archimax if it permits the representation (2.1) for some d -variate stdf ℓ and an Archimedean generator ψ with inverse $\phi : (0, 1] \rightarrow [0, \infty)$, where by convention $\psi(\infty) = 0$ and $\phi(0) = x_\psi$.

In the special case when $\psi(x) = \psi_\Pi(x) = \exp(-x)$, one has that $C_{\psi, \ell}$ reduces to the extreme value copula with stdf ℓ defined for all $(u_1, \dots, u_d) \in (0, 1)^d$ by

$$C_\ell(\mathbf{u}) = \exp\{-\ell(-\ln u_1, \dots, -\ln u_d)\}.$$

Stable tail dependence functions were introduced by Huang (1992) and are given a characterization by Ressel (2013) in terms of homogeneity, convexity and boundary properties. In this paper, the d -norm representation of stdfs is particularly useful. The following characterization, as discussed in Aulbach, Falk and Zott (2015), can be traced back to the work of Pickands (1975), de Haan and Resnick (1977), and Vatan (1985) on the representation of standard max-stable processes. Any d -dimensional stdf ℓ can be written, for $(x_1, \dots, x_d) \in \mathbb{R}_+^d$, as

$$\ell(x_1, \dots, x_d) = \mathbb{E}\left(\max_{1 \leq k \leq d} x_k W_k\right) \tag{2.2}$$

for some positive random variables W_1, \dots, W_d with unit mean. When $\ell(\mathbf{x}) = \ell_\Pi(\mathbf{x}) = x_1 + \dots + x_d$, i.e., the stdf corresponding to asymptotic independence, $C_{\psi, \ell}$ is simply an Archimedean copula given for all $(u_1, \dots, u_d) \in (0, 1)^d$ by

$$C_\psi(\mathbf{u}) = \psi\{\phi(u_1) + \dots + \phi(u_d)\}.$$

Conditions on ψ for C_ψ to be a copula were explored in McNeil and Nešlehová (2009), while conditions for ψ and ℓ for (2.1) to be a copula were explored in Charpentier et al. (2014).

Archimax copulas also admit a stochastic representation, which this paper builds upon. Theorem 3.3 of Charpentier et al. (2014) states under conditions on ψ and ℓ , that $C_{\psi, \ell}$ is the survival copula of a random vector

$$(X_1, \dots, X_d) = (RS_1, \dots, RS_d) = R\mathbf{S}, \tag{2.3}$$

where R is a positive random variable independent of $\mathbf{S} = (S_1, \dots, S_d)$. The distribution of R is linked to the Archimedean generator via the Williamson- d transform, i.e. $\psi = \mathfrak{W}_d(F_R)$ and $F_R = \mathfrak{W}_d^{-1}(\psi)$. We refer to McNeil and Nešlehová (2009) and Larsson and Nešlehová (2011) for more details. The survival function of \mathbf{S} is given, for any $\mathbf{s} \in \mathbb{R}_+^d$, by

$$\bar{G}_\ell(\mathbf{s}) = \Pr(S_1 > s_1, \dots, S_d > s_d) = [\max\{0, 1 - \ell(\mathbf{s})\}]^{d-1}. \tag{2.4}$$

Note in particular that the margins of \mathbf{S} are Beta. Specifically, $S_i \sim \text{Beta}(1, d-1)$ for all $i \in \{1, \dots, d\}$. We interpret (2.3) as a dependence structure defined by a distortion (or radial) random variable R applied to the extremal component \mathbf{S} .

Archimax copulas have a given extreme-value attractor, which motivates their use to model pre-extreme dependence. Recall that a function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is regularly varying with index $\alpha \in \mathbb{R}$ if and only if for all $x > 0$, $f(xt)/f(t) \rightarrow x^\alpha$ as $t \rightarrow \infty$, denoted $f \in \mathcal{R}_\alpha$. When $1 - \psi(1/\cdot) \in \mathcal{R}_{-\alpha}$ for $\alpha \in (0, 1]$, it is shown in Proposition 6.1 of Charpentier et al. (2014) that $C_{\psi, \ell}$ is in the maximum domain of attraction of the extreme-value copula C_{ℓ_α} , i.e., for any $\mathbf{u} \in [0, 1]^d$,

$$\lim_{n \rightarrow \infty} C_{\psi, \ell}^n(\mathbf{u}^{1/n}) = C_{\ell_\alpha}(\mathbf{u}), \quad (2.5)$$

where for any $\mathbf{x} \in \mathbb{R}_+^d$, $\ell_\alpha(\mathbf{x}) = \ell^\alpha(\mathbf{x}^{1/\alpha})$. It is apparent that the Archimax family is fully flexible in the asymptotic regime, meaning that any extreme-value copula C_ℓ corresponds to a subclass of Archimax copulas that will be attracted to it.

3. Model specification and notation

As a first step towards the specification of clustered Archimax copulas, we need to introduce the notion of clusters to the random vector \mathbf{X} . To that end, let $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_K\}$ be a partition of $\{1, \dots, d\}$ into K disjoint sets. Because the stochastic representation (2.3) only makes sense in dimensions two and higher, we shall require, throughout this paper, that $d_k = |\mathcal{G}_k| \geq 2$ for all $k \in \{1, \dots, K\}$. Note that singleton clusters could be included, but this would require more tedious notation. In this setup, $K \leq \lfloor d/2 \rfloor$ and of course also $d_1 + \dots + d_K = d$. For convenience, we treat the subsets \mathcal{G}_k as ordered sets. This allows us to refer to the subvector of \mathbf{X} associated with the k th cluster $\mathcal{G}_k = \{i_1, \dots, i_{d_k}\}$ as $\mathbf{X}_k = (X_{i_1}, \dots, X_{i_{d_k}})$ or, in shorter notation, $\mathbf{X}_k = (X_i)_{i \in \mathcal{G}_k}$.

As we shall see shortly, a clustered Archimax copula is defined through a partition \mathcal{G} as well as K stdfs and K distortion variables. To ease the reading, we will use the notation $\boldsymbol{\ell} = (\ell_1, \dots, \ell_K)$ and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_K)$, where for each $k \in \{1, \dots, K\}$, ℓ_k is a d_k -variate stdf and ψ_k is a d_k -monotone Archimedean generator, i.e., differentiable up to order $d_k - 2$ with derivatives satisfying $(-1)^m \psi_k^{(m)}(x) \geq 0$ for all $x \in (0, \infty)$ for $m \in \{1, \dots, d_k - 2\}$ and further $(-1)^{d_k - 2} \psi_k^{(d_k - 2)}$ is nonincreasing and convex on $(0, \infty)$.

Definition 3.1. A d -variate copula $C_{\mathcal{G}, \boldsymbol{\psi}, \boldsymbol{\ell}, Q}$ is called clustered Archimax copula with cluster partition $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_K\}$, stdfs $\boldsymbol{\ell}$, Archimedean generators $\boldsymbol{\psi} = \{\psi_1, \dots, \psi_K\}$ respectively d_1, \dots, d_K -monotone, and copula Q that we term the radial copula, if it is the survival copula of a random vector \mathbf{X} that satisfies the following:

- (i) For each $k \in \{1, \dots, K\}$, $\mathbf{X}_k = R_k \mathbf{S}_k$ for some d_k -dimensional random vector \mathbf{S}_k with survival function $\bar{G}_k = \bar{G}_{\ell_k}$ as in (2.4) and random variable R_k is distributed as the inverse Williamson d_k -transform of ψ_k .

- (ii) The random vectors $\mathbf{S}_1, \dots, \mathbf{S}_K$ are mutually independent.
- (iii) The random vector $\mathbf{R} = (R_1, \dots, R_K)$ is independent of $\mathbf{S}_1, \dots, \mathbf{S}_K$ and has copula Q .

Note that by Sklar’s Theorem, the joint distribution of \mathbf{R} in Definition 3.1 is fully determined by its copula Q and marginal distributions that are the inverse Williamson- d transforms of ψ .

In more explicit notation, Definition 3.1 states that, upon re-indexing so that \mathcal{G} is contiguous with ordered subsets, any clustered Archimax random vector \mathbf{X} admits the representation

$$\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_K) = (R_1 S_{11}, \dots, R_1 S_{1d_1}, \dots, R_K S_{K1}, \dots, R_K S_{Kd_K}). \quad (3.1)$$

As the name suggests, certain multivariate margins of a clustered Archimax copula $C_{\mathcal{G}, \psi, \ell, Q}$ are Archimax. Specifically, Theorem 3.3 of Charpentier et al. (2014) ensures that for each $k \in \{1, \dots, K\}$, the survival copula of \mathbf{X}_k is the d_k -dimensional Archimax copula C_{ψ_k, ℓ_k} . In particular, in the boundary case when $K = 1$, the entire copula is Archimax. It also follows from the proof of the latter theorem that the survival copula $C_{\mathcal{G}, \psi, \ell, Q}$ of \mathbf{X} in (3.1) is the distribution function of

$$\{\psi_1(R_1 S_{11}), \dots, \psi_1(R_1 S_{1d_1}), \dots, \psi_K(R_K S_{K1}), \dots, \psi_K(R_K S_{Kd_K})\}. \quad (3.2)$$

Throughout the paper, for each $k \in \{1, \dots, K\}$, we let $\mathbf{S}_k = (S_i)_{i \in \mathcal{G}_k}$ be the vector that contains the components of \mathbf{S} corresponding to cluster k . In particular, when \mathbf{X} is as in (3.1), then $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_K)$. Furthermore, analogously to \mathbf{X} and \mathbf{S} , for any vector $\mathbf{x} \in \mathbb{R}_+^d$ and all $k \in \{1, \dots, K\}$, we denote by $\mathbf{x}_k = (x_i)_{i \in \mathcal{G}_k} \in \mathbb{R}_+^{d_k}$ the subvector of \mathbf{x} associated with the k th cluster \mathcal{G}_k . Accordingly, for all $k \in \{1, \dots, K\}$ and $i \in \{1, \dots, d_k\}$, we use X_{ki} , S_{ki} and x_{ki} to denote the i th entry of \mathbf{X}_k , \mathbf{S}_k and \mathbf{x}_k , respectively. Finally, unless otherwise stated, all operations involving one- and multi-dimensional vectors (random or not) should be understood as componentwise, e.g., for $k \in \{1, \dots, K\}$, $R_k \mathbf{S}_k^2 = (R_k S_i^2)_{i \in \mathcal{G}_k}$.

4. Extremal properties

In this section, we investigate the extremal behavior of a clustered Archimax copula $C_{\mathcal{G}, \psi, \ell, Q}$. The main result, Theorem 4.1 below, delineates the conditions under which $C_{\mathcal{G}, \psi, \ell, Q}$ is in a copula domain of attraction of some extreme-value copula and identifies the latter. Since the survival copula $C_{\mathcal{G}, \psi, \ell, Q}$ of \mathbf{X} in (3.1) is also the copula of $1/\mathbf{X}$, we will study the extremal behavior of $1/\mathbf{X}$.

The distortion vector \mathbf{R} has an effect on both inter- and intra-cluster dependence at extreme levels. Its extreme behavior is important, so it is natural to make the following two assumptions. The first concerns the properties of the margins of $1/\mathbf{R}$. Recall that a univariate random variable X with distribution F is in the maximum domain of attraction of a non-degenerate distribution G , denoted $X \in \mathcal{M}(G)$ or $F \in \mathcal{M}(G)$ iff there exist sequences $a_n \in \mathbb{R}_+$, $b_n \in \mathbb{R}$

such that, for any $x \in \mathbb{R}$, $F^n(a_n x + b_n) \rightarrow G(x)$ as $n \rightarrow \infty$. Moreover, the Fisher-Tippett Theorem states that G , up to location and scale, is either Fréchet (Φ_ρ), Gumbel (Λ) or Weibull (Ψ_ρ) with $\rho > 0$.

Assumption 4.1. For a clustered Archimax copula as in Definition 3.1, assume that $\{1, \dots, K\}$ is the union of disjoint sets \mathcal{D}_1 and \mathcal{D}_2 , such that

- (i) $k \in \mathcal{D}_1$ if and only if $1/R_k \in \mathcal{M}(\Phi_{\rho_k})$ for some $\rho_k \in (0, 1)$.
- (ii) $k \in \mathcal{D}_2$ if and only if there exists an $\epsilon_k > 0$ such that $E\{1/R_k^{1+\epsilon_k}\} < \infty$.

While the two cases above cover most widely considered Archimedean generators, we do conjecture an extension in Appendix C.1 that includes the boundary case $1/R_k \in \mathcal{M}(\Phi_1)$ and $E(1/R_k) = \infty$. If Assumption 4.1 holds, $k \in \mathcal{D}_1$ means that $1/R_k$ is heavy-tailed and this occurs if and only if $1 - \psi(1/\cdot) \in \mathcal{R}_{-\rho_k}$, as shown in Theorem 2 in Larsson and Nešlehová (2011). In contrast, $k \in \mathcal{D}_2$ implies that $1 - \psi(1/\cdot) \in \mathcal{R}_{-1}$ by Proposition 2 in Belzile and Nešlehová (2017). By the same proposition, one then has that $1/X_{ki} \in \mathcal{M}(\Phi_{\rho_k})$ for $k \in \mathcal{D}_1$ and $i \in \{1, \dots, d_k\}$ and $1/X_{ki} \in \mathcal{M}(\Phi_1)$ for $k \in \mathcal{D}_2$ and $i \in \{1, \dots, d_k\}$. This means that under Assumption 4.1, the respective clustered Archimax copula is in the copula domain of attraction of an extreme-value copula C_0 if and only if $1/\mathbf{X}$ is in the maximum domain of attraction of an extreme-value distribution with copula C_0 . Such a domain of attraction result requires further assumptions on the extremal behavior of the entire vector $1/\mathbf{R}$.

Assumption 4.2. For a clustered Archimax copula as in Definition 3.1, assume that the reciprocal distortion vector $1/\mathbf{R}$ is in the maximum domain of attraction of a multivariate extreme-value distribution with stdf $\ell_{1/\mathbf{R}}$ given, for $\mathbf{x} \in \mathbb{R}_+^K$, by

$$\ell_{1/\mathbf{R}}(\mathbf{x}) = E\left(\max_{1 \leq k \leq K} x_k W_k\right)$$

for some positive random variables W_1, \dots, W_K with unit mean.

It will become apparent in the next result that the choice of the aforementioned d -norm representations for stdfs is convenient in this context. We are now in position to formulate the main result of this section.

Theorem 4.1. Let $C_{\mathcal{G}, \psi, \ell, Q}$ be a clustered Archimax copula such that Assumptions 4.1 and 4.2 hold with (W_1, \dots, W_K) independent of \mathbf{S} . For $k \in \mathcal{D}_1$, let $b_k = E\{1/Z_k^{\rho_k}\}$, $Z_k \sim \text{Beta}(1, d_k - 1)$. Then $1/\mathbf{X} \in \mathcal{M}(H)$ with $1/X_{ki} \in \mathcal{M}(H_{ki})$, where $H_{ki} = \Phi_{\rho_k}$ for $k \in \mathcal{D}_1$ and $i \in \{1, \dots, d_k\}$ and $H_{ki} = \Phi_1$ for $k \in \mathcal{D}_2$ and $i \in \{1, \dots, d_k\}$. The stdf of H is given for all $\mathbf{x} \in \mathbb{R}_+^d$ by

$$\ell_{\mathcal{G}, \psi, \ell, Q}(\mathbf{x}) = E\left(\max_{\substack{k \in \mathcal{D}_1 \\ 1 \leq i \leq d_k}} \frac{x_{ki} W_k}{b_k S_{ki}^{\rho_k}}\right) + \sum_{k \in \mathcal{D}_2} \ell_k(\mathbf{x}_k). \tag{4.1}$$

Inter-cluster asymptotic independence can also be achieved if the distortions are asymptotically independent, as shown in the following corollary.

Corollary 4.1. *If $\{1/R_k : k \in \mathcal{D}_1\}$ are asymptotically independent, then for $\mathbf{x} \in \mathbb{R}_+^d$, the limiting stdf in (4.1) simplifies to*

$$\ell_{\mathcal{G},\psi,\ell,Q}(\mathbf{x}) = \sum_{k \in \mathcal{D}_1} \ell_k^{\rho_k}(\mathbf{x}_k^{1/\rho_k}) + \sum_{k \in \mathcal{D}_2} \ell_k(\mathbf{x}_k).$$

Remark 4.1. Note that under the hypothesis of Theorem 4.1, the asymptotic behavior of $\{1/R_k : k \in \mathcal{D}_2\}$ has no influence on the form of $\ell_{\mathcal{G},\psi,\ell,Q}$.

The following corollary to Theorem 4.1 compares the inter-cluster stdf to that of the reciprocal distortions $(1/R_1, \dots, 1/R_K)$.

Corollary 4.2. *Under the hypothesis of Theorem 4.1, let $(i_1, \dots, i_K) \in \mathcal{G}_1 \times \dots \times \mathcal{G}_K$. Then, for all $\mathbf{x}' \in \mathbb{R}_+^K$ and $\mathbf{x} \in \mathbb{R}_+^d$ such that $(x_{i_1}, \dots, x_{i_K}) = \mathbf{x}'$ and $x_i = 0$ for all $i \in \{1, \dots, d\} \setminus \{i_1, \dots, i_K\}$,*

$$\ell_{1/\mathbf{R}}(\mathbf{x}') \leq \ell_{\mathcal{G},\psi,\ell,Q}(\mathbf{x}).$$

Remark 4.2. It is worth noting that (4.1) elicits a new method to combine different stdfs in a non-trivial way. Since the second component of (4.1) does not reveal any new combination of stdfs, suppose for now that $\mathcal{D}_2 = \emptyset$. For a given $k \in \{1, \dots, K\}$, we then automatically have that \mathcal{D}_1 . Setting $x_{li} = 0$ for all $l \neq k$ and all $i = 1, \dots, d_l$ recovers the marginal stdf of the cluster k . This marginal stdf is equal to

$$\mathbb{E} \left(\max_{1 \leq i \leq d_k} \frac{x_{ki}}{b_k S_{ki}^{\rho_k}} \right)$$

for $\mathbf{x}_k \in \mathbb{R}_+^{d_k}$, which itself is equal to $\ell_k^{\rho_k}(\mathbf{x}_k^{1/\rho_k})$ by Proposition (6.1) of Charpentier et al. (2014). In the bivariate case, the form above is a special case of (7) in Engelke, Opitz and Wadsworth (2019). The attractor of the bivariate Archimax copula is in particular obtained as a special case of their Proposition 1 and Equation (6), see Sections 2.1 and 4 therein. The complete stdf, defined for all $\mathbf{x} \in \mathbb{R}_+^d$, by

$$\mathbb{E} \left(\max_{\substack{k \in \mathcal{D}_1 \\ 1 \leq i \leq d_k}} \frac{x_{ki} W_k}{b_k S_{ki}^{\rho_k}} \right),$$

essentially mixes the marginal cluster stdfs $\ell_1^{\rho_1}(\mathbf{x}_1^{1/\rho_1}), \dots, \ell_K^{\rho_K}(\mathbf{x}_K^{1/\rho_K})$ with the limiting stdf of $(1/R_1, \dots, 1/R_K)$. Simply put, Corollary 4.2 shows that this mixing results in a weaker asymptotic dependence between clusters than that of the reciprocal distortions $(1/R_1, \dots, 1/R_K)$, characterized by $\ell_{1/\mathbf{R}}$.

5. Insights into modeling

In this section, we provide examples of parametric families that can be used to construct clustered Archimax copulas. Simulating from single Archimax copulas has recently garnered attention, as methods have been advanced by Mai (2022)

and Ng, Hasan and Tarokh (2022). Due to the popularity of Archimedean copulas, there is a wide array of parametric families for the distortions to choose from. When ψ is d -times differentiable, its inverse Williamson d -transform has the density, given, for $r > 0$, by

$$f_R(r) = (-1)^d \frac{r^{d-1} \psi^{(d)}(r)}{(d-1)!};$$

viz. Eq. (2) in McNeil and Nešlehová (2010).

Example 5.1 (Clayton Generator). Consider the Clayton generator defined in (4.2.1) in Nelsen (2006), for $x \in [0, \infty)$ and $\theta \in [-1, \infty) \setminus \{0\}$, by

$$\psi_\theta(x) = (1 + \theta x)^{-1/\theta}.$$

When $\theta \geq -1/(d-1)$, ψ_θ is d -monotone and one can obtain the distribution of R using the inverse Williamson d -transform. For example, when $\theta > 0$, McNeil and Nešlehová (2009) calculate that one has, for $r > 0$,

$$f_R(r) = \frac{(1 + \theta r)^{-d-1/\theta} r^{d-1}}{(d-1)!} \prod_{j=1}^{d-1} (1 + \theta j).$$

We can see that for any $d \geq 2$ and $\beta < d$,

$$E(1/R^\beta) = \frac{\theta^d \left\{ \prod_{j=0}^d (1/\theta + j) \right\}}{(d-1)!} \int_0^\infty \frac{r^{d-1-\beta}}{(1 + \theta r)^{1/\theta+d}} dr < \infty.$$

Thus if the k -th cluster has a Clayton distortion with $\theta > 0$, then its components are asymptotically independent from all other clusters since $k \in \mathcal{D}_2$ in Theorem 4.1.

Example 5.2 (Joe generator). Consider the Joe generator defined in (4.2.6) in Nelsen (2006), for $x \in [0, \infty)$ and $\theta \in [1, \infty)$, by

$$\psi_\theta(x) = 1 - \{1 - e^{-x}\}^{1/\theta}.$$

Since $1 - \psi_\theta(1/\cdot) \in \mathcal{R}_{-1/\theta}$, Theorem 2 from Larsson and Nešlehová (2011) implies that $1/R \in \mathcal{M}(\Phi_{1/\theta})$. Thus, if the k -th cluster has a Joe distortion, it is asymptotically dependent with all other clusters with $j \in \mathcal{D}_1$, whose distortions R_j are asymptotically dependent with R_k .

We now present synthetic examples of clustered Archimax copula based on the families presented above. By virtue of being constructed via the stochastic representation in Equation (3.2), random number generation from this model is straightforward. Recall that we do not require complete monotonicity of the Archimedean generators and therefore rely on the radial representation in Equation (2.3). Our simulation algorithm is a simple extension of Algorithm 4.2 in Charpentier et al. (2014). The R code to generate the samples is provided in the supplementary materials (Chatelain, Perreault, Fougères and Nešlehová, 2025).

Algorithm 5.1. Let $C_{\mathcal{G},\psi,\ell,Q}$ be as in Definition 3.1. To simulate an observation (U_1, \dots, U_d) from $C_{\mathcal{G},\psi,\ell,Q}$,

1. Simulate a vector (R_1, \dots, R_K) . This can be done by simulating a vector $(V_1, \dots, V_K) \sim \bar{Q}$ and applying the transformations $F_{R_k}^{-1}(1 - V_k)$ for each $k \in \{1, \dots, K\}$. Following *McNeil and Nešlehová (2009)*, for $r \in [0, \infty)$,

$$F_{R_k}(r) = 1 - \sum_{j=0}^{d_k-2} \frac{(-1)^j r^j \psi_k^{(j)}(r)}{j!} - \frac{(-1)^{d-1} r^{d-1} \psi_{k,+}^{(d_k-1)}(r)}{(d_k-1)!},$$

where $\psi_{k,+}^{(d-1)}$ is the right-hand derivative of $\psi_k^{(d-1)}$.

2. For each $k \in \{1, \dots, K\}$, generate an observation $\mathbf{S}_k = (S_{k1}, \dots, S_{kd_k})$ whose survival function is given, for any $\mathbf{s} \in \mathbb{R}_+^{d_k}$, by

$$\bar{G}_{\ell_k}(\mathbf{s}) = \Pr(S_{k1} > s_1, \dots, S_{kd_k} > s_{d_k}) = [\max\{0, 1 - \ell_k(\mathbf{s})\}]^{d_k-1}.$$

3. Construct \mathbf{U} by setting $U_1 = \psi_1(R_1 \mathbf{S}_1), \dots, U_K = \psi_K(R_K \mathbf{S}_K)$.

Remark 5.1. In fact, the dependence structure of the distortions does not need to be defined via a copula, as long as it can be simulated from.

In order to illustrate the findings of Section 4, we use Algorithm 5.1 to generate samples from two clustered Archimax copulas, Model A and Model B described in Table 5.1; they differ only in the choice of their radial copula Q . In both cases, three trivariate Archimax copulas representing three clusters are combined to form a 9-dimensional dependence structure. The first cluster is defined by a Clayton-Gumbel Archimax copula, while the other two are defined by Joe-Gumbel Archimax copulas. The partition of $\{1, \dots, 9\}$ is $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\} = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$. Assumption 4.1 holds with $1 \in \mathcal{D}_1$ and $2, 3 \in \mathcal{D}_2$. Figures D.1 and D.2 in Appendix D represent samples drawn from Model A and B. It is clearly visible when comparing the off-diagonal 3-by-3 blocks that both models have the same intra-cluster dependence while having different inter-cluster dependence.

For both samples, we produce chi plots to represent extremal intra and inter-cluster dependence. Out of the 36 possible variable pairs, we chose 6 to cover all pairings of clusters $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$. Figure 5.1 displays those of variables $\{1, 2\}$ (intra-cluster $\mathcal{G}_1 - \mathcal{G}_1$), $\{1, 7\}$ (inter-cluster $\mathcal{G}_1 - \mathcal{G}_3$) and $\{4, 7\}$ (inter-cluster $\mathcal{G}_2 - \mathcal{G}_3$). Figure D.3 in Appendix D pertains to the remaining three cluster pairings. Following *Coles, Heffernan and Tawn (1999)*, for the pair (i, j) , the quantity of interest is given, for $q \in (0, 1)$, by

$$\chi_{ij}(q) = 2 - \frac{\log \Pr(U_i < q, U_j < q)}{\log q}. \tag{5.1}$$

The well-known upper tail dependence coefficient of *Joe (2015)* is then simply expressed

$$\lambda_{ij} = \lim_{q \uparrow 1} \chi_{ij}(q), \tag{5.2}$$

TABLE 5.1
Simulated models.

	Model A	Model B
Radial (survival) copula \bar{Q}	Gaussian, $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$	Gumbel, $\vartheta_{\mathbf{R}} = 4$
Radial extremal behavior	Asymptotic independence $\lambda_{\mathbf{R}} = 0$	Asymptotic dependence $\lambda_{\mathbf{R}} = 0.81$
\mathbf{S} specification	ℓ_1 : Logistic, $\vartheta_1 = 1.25$ ℓ_2 : Logistic, $\vartheta_2 = 2$ ℓ_3 : Logistic, $\vartheta_3 = 1.5$	ℓ_1 : Logistic, $\vartheta_1 = 1.25$ ℓ_2 : Logistic, $\vartheta_2 = 2$ ℓ_3 : Logistic, $\vartheta_3 = 1.5$
\mathbf{R} specification	R_1 : Clayton, $\theta_1 = 1.5$ R_2 : Joe, $\theta_2 = 1.5$ R_3 : Joe, $\theta_3 = 2$	R_1 : Clayton, $\theta_1 = 1.5$ R_2 : Joe, $\theta_2 = 1.5$ R_3 : Joe, $\theta_3 = 2$
Intra-cluster extremal dependence	Cluster 1: Asymptotic dependence $\lambda^{(1)} \approx 0.26$ Cluster 2: Asymptotic dependence $\lambda^{(2)} \approx 0.74$ Cluster 3: Asymptotic dependence $\lambda^{(3)} \approx 0.74$	Cluster 1: Asymptotic dependence $\lambda^{(1)} \approx 0.26$ Cluster 2: Asymptotic dependence $\lambda^{(2)} \approx 0.74$ Cluster 3: Asymptotic dependence $\lambda^{(3)} \approx 0.74$
Inter-cluster extremal dependence	Clusters 1-2: Asymptotic independence $\lambda^{(12)} = 0$ Clusters 1-3: Asymptotic independence $\lambda^{(13)} = 0$ Clusters 2-3: Asymptotic independence $\lambda^{(23)} = 0$	Clusters 1-2: Asymptotic independence $\lambda^{(12)} = 0$ Clusters 1-3: Asymptotic independence $\lambda^{(13)} = 0$ Cluster 2-3: Asymptotic dependence $\lambda^{(23)} \approx 0.5$

provided the limit exists. For each $k \in \{1, 2, 3\}$, the values of λ_{ij} are in fact equal for all $i, j \in \mathcal{G}_k$ and $i \neq j$ due to the fact that the logistic stdf is symmetric with respect to permutation of its arguments. Therefore, for each $k \in \{1, 2, 3\}$, we can simplify the notation by having $\lambda^{(k)} = \lambda_{ij}$ with any $i, j \in \mathcal{G}_k$ such that $i \neq j$. Both samples exhibit the same intra-cluster extreme dependence, with $\lambda^{(1)} = 2 - 2^{1/\vartheta_1} \approx 0.26$ for the first cluster (see Figures 5.1 (a) and (b)), $\lambda^{(2)} = 2 - 2^{1/(\theta_2 \vartheta_2)} \approx 0.74$ (see Figures D.3 (c) and (d)) for the second cluster and $\lambda^{(3)} = 2 - 2^{1/(\theta_3 \vartheta_3)} \approx 0.74$ (see Figures D.3 (e) and (f)) for the third cluster. Note that the pairs (θ_2, ϑ_2) , (θ_3, ϑ_3) were chosen to be different while resulting in the same upper tail dependence coefficient. These values are the same within each cluster because $C_{\psi_{\theta_k}, \ell_{\vartheta_k}}$ was chosen to be exchangeable for each $k \in \{1, 2, 3\}$. The upper tail dependence coefficient of any bivariate copula C in the domain of attraction of an extreme-value copula C_ℓ can be shown to be $\lambda = 2 - \ell(1, 1)$. We therefore obtain the true values of $\lambda^{(k)}$ via (2.5), noting that the index of regular variation α is equal to 1 for any Clayton generator ψ_θ and equal to $1/\theta$ for any Joe generator ψ_θ .

Now, for $k, l \in \{1, 2, 3\}$ and $k \neq l$, let $\lambda^{(kl)} = \lambda_{ij}$ where $i \in \mathcal{G}_k$ and $j \in \mathcal{G}_l$. As for the intra-cluster extreme dependence, this simplified notation is possible due to the fact that given a pair \mathcal{G}_k and \mathcal{G}_l , all values of λ_{ij} such that $i \in \mathcal{G}_k$ and $j \in \mathcal{G}_l$ are equal. In Figures 5.1 (c) and (d) and D.3 (a) and (b), we

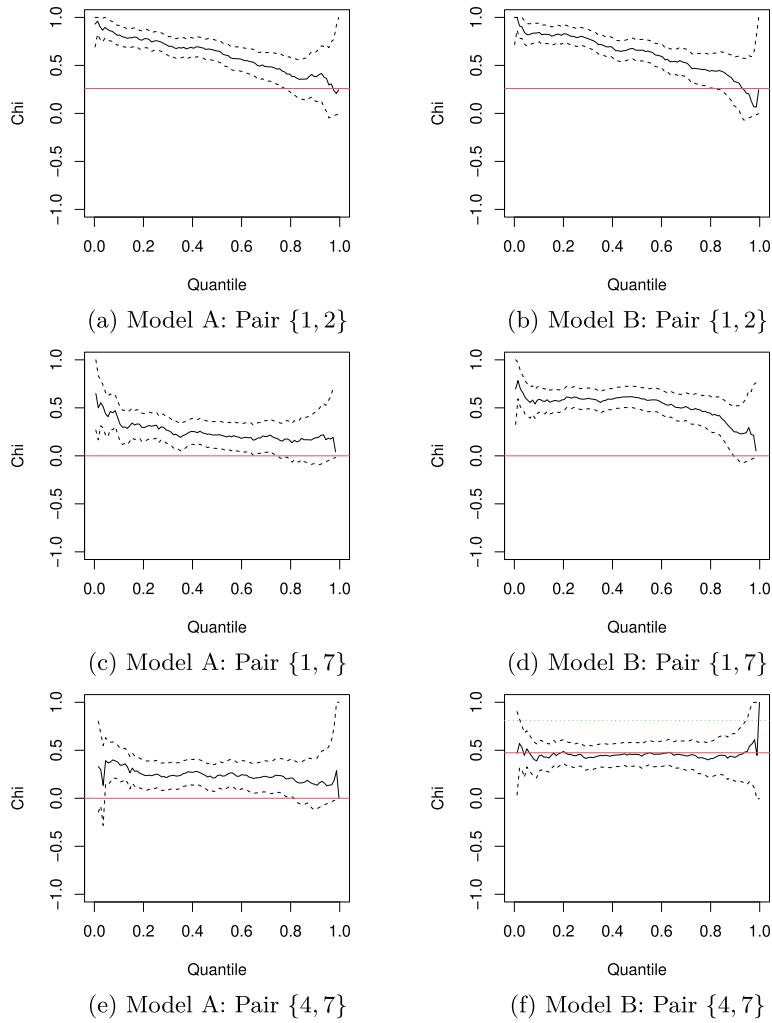


FIG 5.1. Pair chi plots for the sample from Model A (left) and Model B (right). The full black line is the empirical estimate of (5.1), the dotted lines are 95% confidence intervals and the red lines represent the true values of $\lim_{q \rightarrow 1} \chi_{ij}(q)$ for each pair (i, j) . The dotted green line in panel (f) represents the upper tail dependence coefficient of $(1/R_2, 1/R_3)$. The samples used for the empirical estimates are those of Figures D.1 and D.2.

have $\lambda^{(12)} = \lambda^{(13)} = 0$. As explained in Example 5.1, the choice of a Clayton generator forces cluster 1 to be asymptotically independent from clusters 2 and 3. In Figure 5.1 (e) we have $\lambda^{(23)} = 0$. This is due to the fact that the Gaussian copula used to model the dependence between distortions forces asymptotic independence between clusters; see Example 4.1. However, Figure 5.1 (f) shows asymptotic dependence between clusters 2 and 3, i.e. $\lambda^{(23)} > 0$. The value, approximately equal to 0.5, was obtained numerically by simulating from (4.1).

Corollary 4.2 is illustrated by the fact that $\lambda^{(23)}$ is lower than the upper tail dependence coefficient of $(1/R_2, 1/R_3)$.

6. Inference

From now on, let \mathbf{Y} denote the d -dimensional, continuous random vector of interest, whose underlying copula $C_{\mathcal{G},\psi,\ell,Q}$ is clustered Archimax. For $k \in \{1, \dots, K\}$ and $i \in \{1, \dots, d_k\}$ let F_{ki} denote the distribution function of Y_{ki} , recalling that $\mathbf{Y}_k = (Y_i)_{i \in \mathcal{G}_k}$ by definition, and let \mathbf{U} be the d -dimensional copula realization associated with \mathbf{Y} , i.e., $U_{ki} = F_{ki}(Y_{ki})$ and \mathbf{U} has distribution $C_{\mathcal{G},\psi,\ell,Q}$. Finally, suppose that we observe a sample $\{\mathbf{Y}^j\}_{j=1}^n$ of i.i.d. replicates of \mathbf{Y} with corresponding (unobserved) copula realizations $\{\mathbf{U}^j\}_{j=1}^n$.

Theorem 3.3 of Charpentier et al. (2014) implies that there exists a stochastic representation similar to (3.1) for the copula $C_{\mathcal{G},\psi,\ell,Q}$. We will lean on this representation to propose a method for inferring $C_{\mathcal{G},\psi,\ell,Q}$ from $\{\mathbf{Y}^j\}_{j=1}^n$. Given \mathcal{G} , we perform inference for ψ , ℓ and Q separately. To ease the reading, we denote by \mathcal{I} the set

$$\mathcal{I} = \{(i, j, k) : 1 \leq k \leq K, i, j \in \mathcal{G}_k, i < j\} \quad (6.1)$$

throughout the section, as well as in Appendix B.

In Section 6.1, we briefly review how each individual cluster can be modeled using existing inference techniques for Archimax copulas. In Section 6.2, we propose a method for estimating the dependence between clusters when the partition \mathcal{G} is known or hypothesized. In some cases, such as a portfolio containing stocks from distinct industries, this partition can indeed be inherent to the dataset. When \mathcal{G} is unknown, one can use the clustering techniques of, e.g., Bernard et al. (2013) or Saunders, Stephenson and Karoly (2021) to create a set of candidate partitions of distinct sizes. The number of clusters can then be settled heuristically with the help of standard clustering tools (e.g., a dendrogram), or by selecting the coarsest partition that yields a satisfactory fit. These proposed inference tools are investigated via simulations in Section 6.3.

6.1. Inference for individual clusters

This section pertains to the estimation of each marginal Archimax copulas of the clustered Archimax copula $C_{\mathcal{G},\psi,\ell,Q}$. As such, inference methods have already been developed in Chatelain, Fougères and Nešlehová (2020) and more recently in Ng, Hasan and Tarokh (2022).

6.1.1. Estimating ψ_1, \dots, ψ_K

We use a parametric approach to estimate the generators $\psi = (\psi_1, \dots, \psi_K)$, which requires selecting a (possibly different) parametric family for each generator. Certain properties observed in the data can guide the user to specific

choices of generators. For example, the presence of within-cluster lower tail dependence warrants the use of a Clayton generator. However, as explained in Example 5.1, this choice implies the asymptotic independence of \mathbf{Y}_k and \mathbf{Y}_ℓ for all $k, \ell \in \{1, \dots, K\}$ with $k \neq \ell$. This consequence is desirable when studying the precipitation data in Section 7, but would not be appropriate, for instance, when modeling temperatures over the same region; see Davison, Huser and Thibaud (2013) for more details.

Once the parametric family of each generator is chosen, the parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ of $\boldsymbol{\psi}$ remain to be estimated. For a given $k \in \{1, \dots, K\}$, our estimator of θ_k relies on the following remark.

Remark 6.1. Any margin of an Archimax copula is itself Archimax with the same generator. In particular, for all $(i, j, k) \in \mathcal{I}$, with \mathcal{I} as in (6.1), the distortions parameters θ_k and θ_{ij} associated with \mathbf{Y}_k and (Y_i, Y_j) , respectively, are identical, i.e., $\theta_k = \theta_{ij}$.

Remark 6.1 suggests that for each $(i, j, k) \in \mathcal{I}$, an estimator $\hat{\theta}_{ij}$ of θ_{ij} also estimates θ_k . In view of this, we estimate $\boldsymbol{\theta}$ by $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1, \dots, \bar{\theta}_K)$ given, for $k \in \{1, \dots, K\}$ and $\mathcal{G}_{kk} = \{(i, j) : i, j \in \mathcal{G}_k, i < j\}$, by

$$\bar{\theta}_k = \frac{1}{\binom{d_k}{2}} \sum_{(i,j) \in \mathcal{G}_{kk}} \hat{\theta}_{ij}, \tag{6.2}$$

where, for any $(i, j) \in \mathcal{G}_{kk}$, $\hat{\theta}_{ij}$ is the estimator defined in Section 7 of Chatelain, Fougères and Nešlehová (2020). These latter authors discuss the cases when the underlying generator is of the Clayton, Genest-Ghoudi, Frank or Joe families.

Remark 6.1 also suggests a simple way of assessing whether the partition \mathcal{G} is appropriate: if the underlying distribution is indeed a clustered Archimax with partition \mathcal{G} , then the hypothesis H_0 that $\theta_{ij} = \theta_k$ for all $(i, j, k) \in \mathcal{I}$ must necessarily hold. We propose to test H_0 using the statistic \mathbf{T} given by

$$T_{ijk} = \hat{\theta}_{ij} - \bar{\theta}_k. \tag{6.3}$$

Although its definition involves three indices, we treat \mathbf{T} as a vector of dimension $|\mathcal{I}|$; the specific ordering of its entries does not matter, as long as it is kept fixed. Because we expect that departures from the null will cause certain entries of \mathbf{T} to be large in absolute values, we test H_0 using either the supremum norm $\|\mathbf{T}\|_\infty = \max_{\iota \in \mathcal{I}} |T_\iota|$ or the (squared) Euclidean norm $\|\mathbf{T}\|_2^2 = \sum_{\iota \in \mathcal{I}} T_\iota^2$ of \mathbf{T} ; see Perreault, Nešlehová and Duchesne (2022) for similar hypothesis tests, albeit in a nonparametric context.

To derive the null distribution of $\|\mathbf{T}\|$, where $\|\cdot\|$ is either the supremum or Euclidean norm, we exploit the fact that for any $(i, j, k) \in \mathcal{I}$, $\hat{\theta}_{ij}$ is a function of two U-statistics with square integrable kernels (Chatelain, Fougères and Nešlehová, 2020, Section 7). Consequently, for any $\iota \in \mathcal{I}$, T_ι is a function of several such U-statistics. Standard results about U-statistics (Hoeffding, 1948) and the delta method then imply that, under the null, $\sqrt{n}\mathbf{T}$ is asymptotically Normal as $n \rightarrow \infty$, i.e., $\sqrt{n}\mathbf{T} \rightsquigarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ for some positive definite matrix $\boldsymbol{\Sigma}$ of

appropriate dimensions. In view of Slutsky's Lemma and the Continuous Mapping Theorem, a test of approximate level $\alpha \in (0, 1)$ then consists in rejecting H_0 whenever $\Pr(\|\mathbf{Z}\| > \|\mathbf{T}\|) < \alpha$, where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \hat{\Sigma})$ for some consistent estimator $\hat{\Sigma}$ of Σ . In the data application of Section 7, we use a jackknife estimator of Σ derived from Section 2(c) of Arvesen (1969); see Appendix B for the implementation details.

6.1.2. Estimating ℓ_1, \dots, ℓ_K

To estimate the stdfs $\boldsymbol{\ell} = (\ell_1, \dots, \ell_K)$, we exploit the fact that for any $k \in \{1, \dots, K\}$ and $\mathbf{x} \in \mathbb{R}^{d_k}$,

$$\ell_k(\mathbf{x}) = \|\mathbf{x}\| A_k(\mathbf{x}/\|\mathbf{x}\|), \quad (6.4)$$

where A_k is the Pickands dependence function associated to \mathbf{Y}_k . Indeed as characterized by Ressel (2013), stdfs are homogeneous functions of degree 1 and are thus determined by their restrictions on the unit simplex. These restrictions, $A_k(\mathbf{s}) = \ell_k(\mathbf{s})$ for $\mathbf{s} \in \Delta_{d_k}$, are none other than Pickands dependence functions as introduced in the bivariate case by Pickands (1981). Specifically, we replace A_k in (6.4) by its CFG-type estimator (Capéraà, Fougères and Genest, 1997), defined as follows. For $j \in \{1, \dots, n\}$, let \hat{U}^j be such that, for all $k \in \{1, \dots, K\}$ and $i \in \{1, \dots, d_k\}$,

$$\hat{U}_{ki}^j = r_{kij}/(n+1), \quad (6.5)$$

where r_{kij} is the rank of Y_{ki}^j among $Y_{ki}^1, \dots, Y_{ki}^n$. Now, for $\mathbf{w} \in \Delta_{d_k}$ with Δ_{d_k} as in Definition 2.1, let

$$\hat{\xi}_{k,j}(\mathbf{w}) = \min_{1 \leq i \leq d_k} \phi_{\hat{\theta}_k}(\hat{U}_{ki}^j)/w_i.$$

Then, for all $k \in \{1, \dots, K\}$ and $\mathbf{w} \in \Delta_{d_k}$, the (endpoint-corrected) CFG-type estimator of A_k is given by

$$\log \hat{A}_k^{\text{CFG}}(\mathbf{w}) = \frac{1}{n} \sum_{j=1}^n \left\{ \log \phi_{\hat{\theta}_k} \left(\frac{j}{n+1} \right) - \log \hat{\xi}_{k,j}(\mathbf{w}) \right\}. \quad (6.6)$$

We refer the reader to Section 3 of Chatelain, Fougères and Nešlehová (2020) for more details about the CFG-type estimator, as well as an alternative estimator based on that of Pickands (1981). The asymptotic behavior of the CFG and Pickands-type estimators is established under regularity conditions on both ψ and A ; see Section 4 in Chatelain, Fougères and Nešlehová (2020).

6.2. Inference for the dependence between distortions

To model the dependence between the components of \mathbf{R} , we suggest a parametric approach on the underlying copula Q . To this end, we make the assumption that Q belongs to a parametric family $\{C_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\}$ of K -dimensional copulas,

where the parameter space Θ is of arbitrary dimension. We further assume the existence of a multivariate density for \mathbf{R} in order to proceed with a likelihood-based method. While we cannot observe \mathbf{R} directly, we may still derive the corresponding likelihood based on pseudo-observations.

We follow (3.2) and define the d -dimensional random vector \mathbf{V} , for all $k \in \{1, \dots, K\}$ and $i \in \{1, \dots, d_k\}$, by $V_{ki} = \psi_k(R_k S_{ki})$, so that \mathbf{V} has distribution $C_{\mathcal{G}, \psi, \ell, Q}$. Now, fix $\boldsymbol{\iota} = (i_1, \dots, i_K) \in \mathcal{G}_\bullet = \mathcal{G}_1 \times \dots \times \mathcal{G}_K$ and note that since $X_{\iota_k} = R_k S_{\iota_k}$ for each $k \in \{1, \dots, K\}$, the density of the subvector $\mathbf{V}_\boldsymbol{\iota} = \{\psi_1(X_{\iota_1}), \dots, \psi_K(X_{\iota_K})\}$ of \mathbf{V} can be expressed, for $\mathbf{v} \in (0, 1)^K$, as

$$f_{\mathbf{V}_\boldsymbol{\iota}}(\mathbf{v}) = \int_{(0,1)^K} f_{\mathbf{R}}\{\boldsymbol{\phi}(\mathbf{v})/\mathbf{s}\} \left\{ \prod_{k=1}^K \frac{f_{S_k}(s_k)\phi'_k(v_k)}{s_k} \right\} d\mathbf{s} \tag{6.7}$$

where $\boldsymbol{\phi}(\mathbf{v})/\mathbf{s} = \{\phi_1(v_1)/s_1, \dots, \phi_K(v_K)/s_K\}$ and, for $k \in \{1, \dots, K\}$, f_{S_k} is a Beta(1, $d_k - 1$) density. Also note that, due to Sklar's Theorem, the joint density of \mathbf{R} can be written, for $\mathbf{r} \in \mathbb{R}_+^K$ and q the copula density of \mathbf{R} , as

$$f_{\mathbf{R}}(\mathbf{r}) = q\{F_{R_1}(r_1), \dots, F_{R_K}(r_k)\} \left\{ \prod_{k=1}^K f_{R_k}(r_k) \right\}. \tag{6.8}$$

Now, let $\mathcal{Y}_n = (\mathbf{Y}^j)_{j=1}^n$ and recall that $\mathbf{U}^j, j \in \{1, \dots, n\}$, and \mathbf{V} are equal in distribution. The pseudo-copula observations can be plugged into the (pseudo) composite marginal log-likelihood (Varin, Reid and Firth, 2011) associated with the density of $\mathbf{V}_\boldsymbol{\iota}$ and given by

$$\mathcal{L}_K(\boldsymbol{\vartheta}|\mathcal{Y}_n) = \sum_{j=1}^n \sum_{\boldsymbol{i} \in \mathcal{G}_\bullet} \log f_{\mathbf{V}_\boldsymbol{\iota}}(\hat{\mathbf{U}}_i^j), \tag{6.9}$$

where $\hat{\mathbf{U}}_i^j = (\hat{U}_{i_1}^j, \dots, \hat{U}_{i_K}^j)$ with $\hat{\mathbf{U}}^j$ as in (6.5), and $f_{\mathbf{V}_\boldsymbol{\iota}}$ as in (6.7).

Due to the multi-dimensional integral in the expression of $f_{\mathbf{V}_\boldsymbol{\iota}}$ and the possibly large cardinality of \mathcal{G}_\bullet , (6.9) will often be difficult to compute in practice. When the chosen parametric family for Q allows it, a natural approach for reducing the computational burden is to consider lower-dimensional margins to form the composite likelihood function. For example, when Q is a Gaussian copula, it seems reasonable to use the pairwise marginal likelihood

$$\mathcal{L}_2(\boldsymbol{\vartheta}|\mathcal{Y}_n) = \sum_{j=1}^n \sum_{1 \leq k < \ell \leq K} \sum_{\boldsymbol{i} \in \mathcal{G}_{k\ell}} \log f_{V_{\iota_k}, V_{\iota_\ell}}(\hat{\mathbf{U}}_i^j), \tag{6.10}$$

where $\mathcal{G}_{k\ell} = \mathcal{G}_k \times \mathcal{G}_\ell$, $\hat{\mathbf{U}}_i^j = (\hat{U}_{i_1}^j, \hat{U}_{i_2}^j)$, and $f_{V_{\iota_k}, V_{\iota_\ell}}$ is the corresponding bivariate analogue of $f_{\mathbf{V}_\boldsymbol{\iota}}$. For more details about composite likelihood estimation, we refer the reader to Cox and Reid (2004) and Varin, Reid and Firth (2011).

Remark 6.2. For some specific models, the similar yet simpler strategy underlying the estimation of $\boldsymbol{\vartheta}$ can also be used. For example, let Q be a Gaussian

copula with correlation matrix $(\vartheta_{k\ell})_{1 \leq k, \ell \leq K}$, so that ϑ is of length $\binom{K}{2}$. For each $k, \ell \in \{1, \dots, K\}$ with $k < \ell$, one can first independently consider the bivariate log-likelihoods associated with $\mathcal{G}_{k\ell}$ in (6.10) to compute a set of pairwise parameters $\{\hat{\rho}_i : i \in \mathcal{G}_{k\ell}\}$, which can then be averaged over to obtain a final estimate $\bar{\rho}_{k\ell}$ for $\vartheta_{k\ell}$. This estimator resembles the averaged maximum pseudo likelihood estimator (AMPLE) in Górecki and Hofert (2023). We conducted a preliminary assessment of this approach by contrasting the resulting model with the one based on (6.10) in the context of the application described in Section 7, which uses the Gaussian copula. We found that the two resulting models are almost indistinguishable in this particular case.

Remark 6.3. The composite log-likelihoods $\mathcal{L}_K(\vartheta|\mathcal{Y}_n)$ and $\mathcal{L}_2(\vartheta|\mathcal{Y}_n)$ in (6.9) and (6.10), respectively, depend on ϑ through q in (6.8) (or its bivariate analogues in the case of $\mathcal{L}_2(\vartheta|\mathcal{Y}_n)$). The functions f_{R_k} , F_{R_k} and ϕ_k , $k \in \{1, \dots, K\}$, are assumed to be known. In practice, one can plug in the parametric estimates proposed in Section 6.1.1. While outside of the scope of this work, one could consider the feasibility of estimating the distortion parameters together with ϑ via pseudo-maximum composite likelihood.

6.3. Simulation study

We assess our proposed inference technique in a simulation study on a range of different clustered Archimax copula models. Inference on individual clusters has been extensively studied in Chatelain, Fougères and Nešlehová (2020) and Ng, Hasan and Tarokh (2022), here the onus is on the evaluation of our method for inference of the underlying copula Q of the distortion variables.

To remain in realistic scenarii, we perform the moment-based estimation of the distortion parameters as described in Chatelain, Fougères and Nešlehová (2020) before applying the pairwise-likelihood estimation method of the previous section, viz. (6.10). This allows us to take into account the propagation of error that would also occur in practice. Throughout, we assumed that the families of Archimedean generators for each clusters were correctly specified, as were the radial copula Q and the clusters.

It can be the case in certain applications that the clusters do not arise naturally. For each simulated dataset, we therefore also employed several clustering algorithms with the correct number of clusters specified and computed Rand index measures. This allowed us to quantify how difficult it is to obtain an accurate data-driven clustering for each studied clustered Archimax copula configuration.

The class of clustered Archimax models proposed here is very rich due to the vast number of combinations of cluster configurations, families, parameters and dimensions. In the vein of the illustrations in Section 5, we restrict ourselves to two types of radial (survival) copulas \bar{Q} (Gaussian and Gumbel copulas) and two types of radial families (Joe and Clayton generators). For comparison's sake, we parameterize according to Kendall's rank correlation for \bar{Q} ($\tau_Q \in \{1/3, 2/3\}$) and the radial components ($\tau_\psi \in \{1/4, 1/2\}$). We explore the effects of the number of clusters ($K \in \{2, 3\}$) as well as their sizes ($d_k \in \{2, 3\}$ for

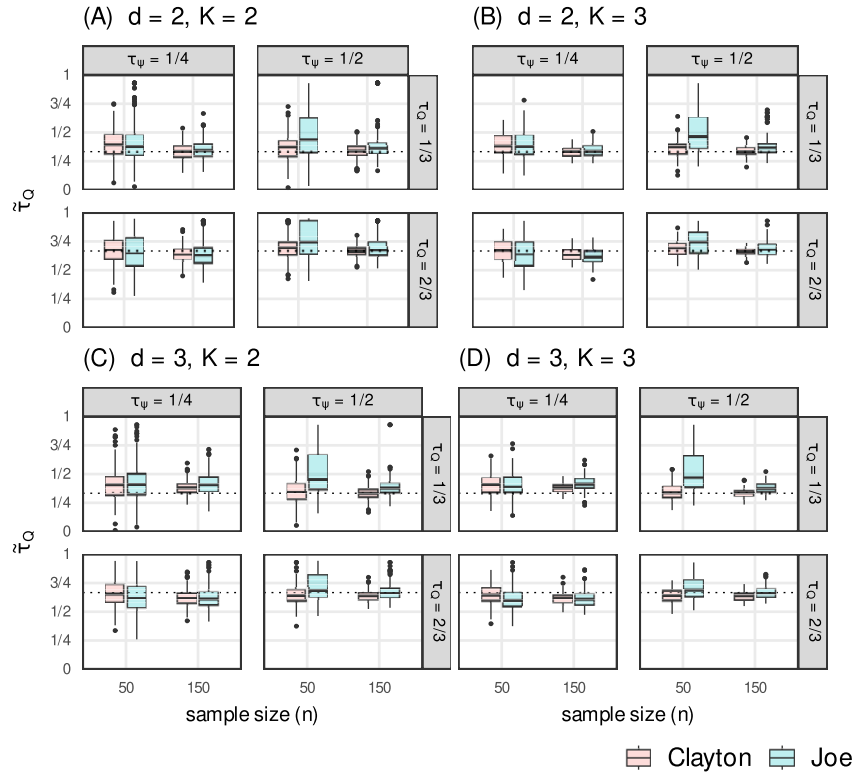


FIG 6.1. Boxplots of the fitted values of Kendall's Tau for the radial copula \tilde{Q} in the Gaussian case. True values τ_Q are plotted in dotted lines. The simulation is based on 100 Monte-Carlo replicates.

$k \in \{1, \dots, K\}$) in small samples ($n \in \{50, 150\}$) motivated by environmental applications with data-scarcity. In our proposed inference method, the intra-cluster stdfs $\{\ell_k\}_{k=1}^K$ can be estimated, e.g., via the semi-parametric approach of [Chatelain, Fougères and Nešlehová \(2020\)](#) after having performed inference on the radial components and their copula. Since the estimation of intra-cluster stdfs is not the focus of this simulation study, we limit ourselves to the logistic stdf with parameter corresponding to the upper-tail dependence coefficient $\lambda^{(k)} = 1/3, k \in \{1, \dots, K\}$.

The main results are reported in Figures 6.1 and E.1. As expected, we observe better performance with a larger sample size, as well as higher dimensional clusters. This latter behavior is because of the exchangeability of the choices for \tilde{Q} and the subsequent averaging of parameter estimates. The number of clusters only affects computational time. For both families, there isn't a clear effect of the value of τ_Q on the estimation performance, although particularly small or large degrees of dependence were not considered in this simulation study. Across the board, there seems to be a slight reduction of variance when τ_ψ increases

from $1/4$ to $1/2$. The most notable difference pertains to the choice of the radial families; the performance is better for the Clayton case than for the Joe case. Note that the presence of outliers in the plot is due to numerical instabilities that mostly happen for small sample sizes and Joe radial components.

To investigate the difference between these two Archimedean families, we provide the boxplots of the obtained moment-based estimates in Figure E.2 that are fed into the inference for \bar{Q} . These results make it clear that the difference in performance at the radial copula level is due to the propagation of the error from the estimation of the individual radial components, i.e. of the parameters θ_k in Table 5.1. Indeed, we observe superior performance for the Clayton family. Figure E.3 shows that, unsurprisingly, performance when the model includes a mix of Clayton and Joe clusters falls between the cases for all clusters being Clayton (best) and for all clusters being Joe (worst).

Figure E.4 reports the average Rand index values obtained for all configurations of cluster numbers and sizes. As expected, performance increases with sample size and cluster size while it decreases with the number of clusters. There is a clear indication that Kendall's tau provides a better dissimilarity matrix for determining the correct Archimax clusters than the upper tail dependence coefficient. Hierarchical clustering algorithms seem to be more suited than the partitioning around medoids approach; in particular complete linkage performs best.

7. Data illustration

In this section, we illustrate the proposed methodology through an application to flood monitoring. The data, provided by Météo France, consists of daily precipitation amounts measured from 1976 to 2015, inclusively, at $d = 23$ meteorological stations in France. As shown in Figure 7.1, the stations are agglomerated into three clusters centered around the cities of Paris (\mathcal{G}_1 , 9 stations), Lyon (\mathcal{G}_2 , 6 stations) and Montpellier (\mathcal{G}_3 , 8 stations). We thus let $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\}$ be the partition underlying our model.

Note that our application is purposely similar to that of Chatelain, Fougères and Nešlehová (2020), who fitted an Archimax copula to precipitation amounts (monthly maxima) measured at three nearby stations. The greater flexibility of clustered Archimax copulas allows us to increase the number of variables considered.

7.1. Data preprocessing

A preliminary analysis of the data reveals the presence of seasonality and temporal dependence within the univariate series. To mitigate the effect of seasonality, we consider only the observations from the months of September to December, inclusively, which encompass most of the extreme precipitation events. Since our primary focus is on extreme precipitations, we then take monthly maxima of the series, yielding a total of $n = 160$ observations per station. The resulting

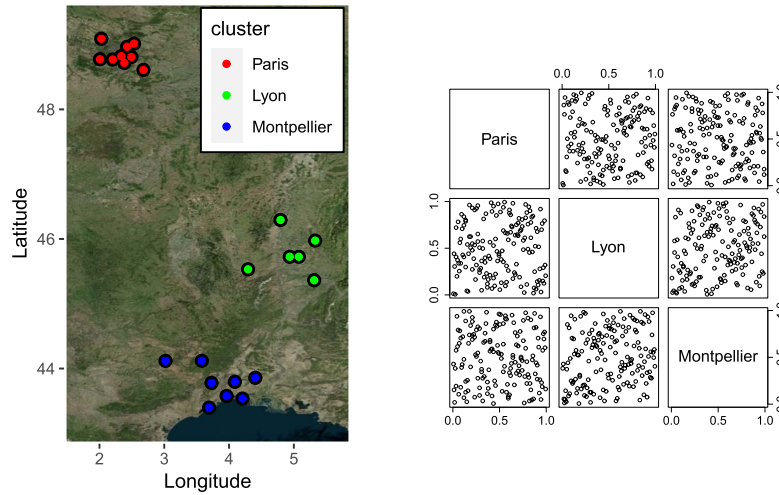


FIG 7.1. *Left: Geographical position of the stations used in the application of Section 7, along with their respective cluster label. Right: Pair plots of the scaled componentwise ranks of monthly maxima of precipitations for the stations around which the clusters were formed.*

series show no obvious sign of temporal dependence according to the Ljung-Box tests (Ljung and Box, 1978); note however that the asymptotic results of Chatelain, Fougères and Nešlehová (2020) upon which the present work relies hold for alpha-mixing sequences, meaning that temporal dependence vanishing with increasing lag is indeed allowed.

The pair plots of the scaled componentwise ranks of monthly maxima of precipitations involving only the most central station of each cluster are displayed in the right panel of Figure 7.1. They suggest a very weak dependence between precipitation maxima in Lyon and Montpellier, and an even weaker one, if at all, between any of these and the stations in Paris. In contrast, similar plots for all pairs of stations within the same cluster (Figures F.1–F.3) indicate much stronger dependencies. In these latter plots, we also note the presence of asymmetry; this is particularly pronounced in Figure F.1 (Paris).

As a final preliminary step, we used the procedure of Kojadinovic, Segers and Yan (2011) to test the hypothesis that the copula underlying each cluster of variables is an extreme-value copula. In all three cases, the test clearly rejects the hypothesis. This may be explained by the presence of masses of points near the bottom-left corner of many pair plots, combined with the fact that extreme-value distributions cannot allow lower-tail dependence. In contrast with extreme-value distributions, clustered Archimax copulas may indeed allow for both lower-tail and extremal dependence; in particular, letting ψ_θ be the Clayton generator leads to (pairwise) lower-tail dependence coefficients equal to $2A(1/2, 1/2)^{-1/\theta}$, where A is the Pickands dependence function characterizing the Archimax, as explained in Chatelain, Fougères and Nešlehová (2020).

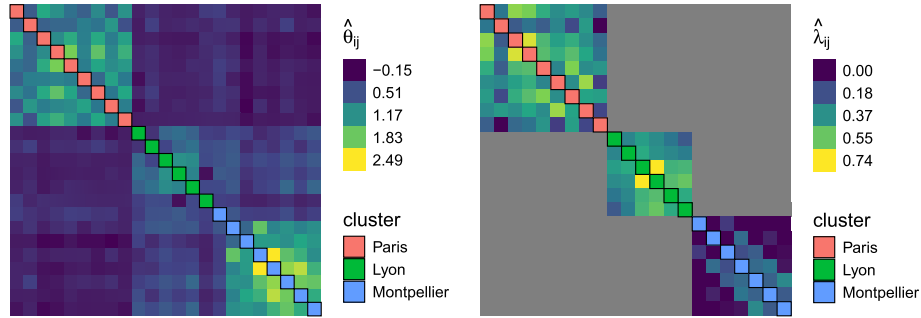


FIG 7.2. *Estimated quantities from Section 7.2. Left: Matrix of intra-cluster pairwise upper tail dependence coefficients ($d = 22$ stations). Right: Matrix of pairwise distortion parameters ($d = 23$ stations).*

7.2. Inference for ψ and ℓ

The apparent lack of inter-cluster dependence in the right panel of Figure 7.1 and the presence of lower-tail dependence in the data suggest that the Clayton generators could be good candidates for modeling the distortions, as these would produce asymptotically independent clusters. We thus begin by formally testing whether \mathcal{G} defines three asymptotically independent clusters. To do so, we apply the test of independence for random vectors proposed by [Kojadinovic and Holmes \(2009\)](#). Because we are interested in asymptotic independence, we apply the test procedure not to our dataset of monthly maxima, but to the corresponding dataset of yearly maxima, which includes 40 observations per station; the corresponding matrix of empirical Kendall's τ correlations is depicted in the left panel of Figure F.4 in the Appendix. The test yields a p-value of 0.16 for the global hypothesis of independence between the three clusters and a p-value above 0.35 for each of the three hypotheses of pairwise independence. Although the fact that only 40 observations are available might arguably yield a test with limited power, we move on with our analysis assuming that the clusters are asymptotically independent and that Clayton generators are reasonable choices.

The next step is to estimate the parameters θ_1 , θ_2 and θ_3 of the three Clayton generators. This involves computing, for each $k \in \{1, 2, 3\}$, the $\binom{d_k}{2}$ pairwise estimates $\{\hat{\theta}_{ij} : i, j \in \mathcal{G}_k, i < j\}$; these are gathered in a $d_k \times d_k$ matrix illustrated in the left panel of Figure 7.2. The resulting distortion estimates, defined in (6.2), are $\bar{\theta}_1 \approx 1.08$ (Paris), $\bar{\theta}_2 \approx 0.62$ (Lyon) and $\bar{\theta}_3 \approx 1.20$ (Montpellier), suggesting that the hypothesis of different distortions affecting the three clusters is reasonable.

At this point, one can already suspect a problem with the Montpellier cluster, as there seems to be strong discrepancies among the entries of its corresponding matrix in the left panel of Figure 7.2, violating the statement in Remark 6.1. To check this more formally, we perform the test described in Section 6.1.1 for

the hypothesis H_0 that $\theta_{ij} = \theta_k$ for all $k \in \{1, 2, 3\}$ and $i, j \in \mathcal{G}_k$ such that $i < j$. While the version of the test based on the supremum norm yields an acceptable p-value (approximately 0.18), the version based on the Euclidean norm yields an approximate p-value of 0.03, thus rejecting H_0 at nominal level 0.05. Three similar cluster-specific tests, each involving only the pairwise distortion estimates from a single cluster, indeed reveal an anomaly with Montpellier; its corresponding approximate p-values are 0.06 (supremum norm) and < 0.01 (Euclidean norm). A last series of entry-specific tests, each involving a single pairwise distortion estimate from the Montpellier cluster, strongly suggests that one particular station, the west-most station of the cluster, is at the root of the rejection. Concretely, the $d_3 - 1 = 7$ p-values associated with this latter station are below the nominal level 0.05, with most of them being smaller than 10^{-3} . From the left panel of Figure 7.1, one can see that it effectively appears isolated from the other stations of the Montpellier cluster, as it is located on the other side of the Cévennes mountain range. Removing its corresponding column from the data yields a new distortion parameter $\hat{\theta}_3 \approx 1.48$ for the Montpellier cluster, as well as satisfactory p-values for the tests of H_0 and the preceding test of asymptotic independence. We thus remove the problematic station from the data and redefine \mathcal{G} to be the ensuing partition of $\{1, \dots, 22\}$, assuming that the variables were re-indexed. Similarly, we now use θ_3 to refer to the estimate of θ_3 computed without the problematic station.

Given the new partition \mathcal{G} , we then construct, for each $k \in \{1, 2, 3\}$, the semi-parametric estimate of the stdf ℓ_k based on \hat{A}_k^{CFG} defined in (6.6). Note that for the estimated values of $\theta_1, \theta_2, \theta_3$, Conditions 4.1 and 4.2 in Chatelain, Fougères and Nešlehová (2020) used for convergence of \hat{A}_k^{CFG} are met. To better visualize the strength of the dependence between stations of the same cluster, we also computed, for each $k \in \{1, 2, 3\}$ and all $i, j \in \mathcal{G}_k$ such that $i \neq j$, pairwise estimates $\hat{\lambda}_{ij}$ of the upper tail dependence coefficients λ_{ij} defined in (5.2). To this end, we note that when Clayton generators are used, $\alpha = 1$ in (2.5) and the estimator in (6.6) can be slightly modified to obtain an estimator of λ_{ij} as follows. Recall the definition of \mathcal{I} in (6.1). For all $(i, j, k) \in \mathcal{I}$, let $\ell_{ij}^*(1, 1) = \ell_k(\mathbf{e}^{(ij)})$, where $\mathbf{e}^{(ij)} \in \mathbb{R}^{d_k}$ has only two non-zero entries, those associated with the i th and j th stations, equal to one. Then, for any $(i, j, k) \in \mathcal{I}$, one can estimate λ_{ij} combining the identities $\lambda_{ij} = 2 - \ell_{ij}^*(1, 1)$ and $\ell_k(\mathbf{x}) = \|\mathbf{x}\|A_k(\mathbf{x}/\|\mathbf{x}\|)$. Because the bivariate margins of an Archimax copula are also Archimax (Remark 6.1), we get that $\lambda_{ij} = 2 - 2A_{ij}(1/2, 1/2)$ with $A_{ij}(1/2, 1/2) = A_k(\mathbf{e}^{(ij)}/2)$. This leads to an estimator of λ_{ij} given by $\hat{\lambda}_{ij} = 2 - 2\hat{A}_{ij}^{\text{CFG}}(1/2, 1/2)$. The resulting estimates are shown in the right panel of Figure 7.2. Although this is not required, we expect many of these estimates to be large; this is clearly the case for the Paris and Lyon clusters, and still true, perhaps to a lesser extent, for the Montpellier cluster.

7.3. Inference for Q

To capture the dependence between the components of $\mathbf{R} = (R_1, R_2, R_3)$, we suppose that their copula Q belongs to the family of Gaussian copulas with correlation parameters $\boldsymbol{\vartheta} = (\vartheta_{12}, \vartheta_{13}, \vartheta_{23})$, which we estimate by numerically maximizing (6.10). The resulting estimates are given by $\bar{\rho}_{12} \approx 0.14$ for the Paris-Lyon pair, $\bar{\rho}_{13} \approx -0.02$ for the Paris-Montpellier pair and $\bar{\rho}_{23} \approx 0.45$ for Lyon-Montpellier pair. The results suggest in particular that the clusters centered around Paris and Montpellier are nearly independent, which is coherent with the fact they are the most distant pair of clusters among the three.

When reconducting the analysis using the simpler estimation method described in Remark 6.2, we obtained virtually identical parameters. This suggests that in some cases the simpler method may be used as a viable alternative to maximizing (6.10). The pairwise Pearson correlations computed during the process, and whose inter-cluster averages are used as final estimates, are shown in the right panel of Figure F.4 in the Appendix.

This application to rainfall illustrates some of the strengths of clustered Archimax copulas. Namely, we are able to model the dependence of multi-dimensional data that consists of groups of variables which are asymptotically dependent. Within each cluster, dependence is modeled in a fully flexible way, allowing for asymmetry. Inter-cluster dependence is modeled more parsimoniously while maintaining the possibility for both asymptotic dependence and independence. Finally, the proposed copula family allows to fit data at a pre-asymptotic level while boasting flexibility at the extreme level.

8. Discussion

The clustered Archimax model studied in this paper is related to several other articles in the literature. Hierarchical constructions based on Archimax copulas were proposed by Hofert, Huser and Prasad (2018). Specifically, their construction is based on the frailty representation of Archimax copulas, which only holds for completely monotone generators. Hierarchies can be induced via the frailties, the stdf, or both. Introducing hierarchies through the stdf alone leads to an Archimax copula again, but introducing hierarchies through the frailties leads to a hierarchical model which would be interesting to compare to the clustered Archimax construction proposed here. Notably, it would be interesting to establish the attractor of the model of Hofert, Huser and Prasad (2018) and compare it to that of the clustered Archimax copula advocated here. Unfortunately, the Hofert, Huser and Prasad (2018) hierarchical model with hierarchies induced through frailties is not, to the best of our knowledge, supplemented with inference techniques; the recent methodology of Górecki and Hofert (2023) only considers hierarchies induced through the stdf. Therefore, performance comparisons with our proposed clustered Archimax copula model are not yet possible.

The extremal dependence structure of Liouville copulas is established in Belzile and Nešlehová (2017). The stochastic representation of Liouville copulas

is similar to that of Archimax copulas, as they are survival copulas of vectors of the form $R\mathbf{D}$, with R a nonnegative random variable and \mathbf{D} a Dirichlet random vector. The work presented in this paper differs from this by replacing the Dirichlet component by a vector \mathbf{S} characterized by an stdf and by allowing for multiple distorting random variables R_1, \dots, R_K , thus inducing a hierarchy (or clustering). Finally, Engelke, Opitz and Wadsworth (2019) establish the extremal dependence of bivariate vectors of the form $R \times (W_1, W_2)$ for an extensive combination of asymptotic behaviors of both R and (W_1, W_2) . The attractor of the bivariate Archimax copula is in particular obtained as a special case of their Proposition 1 and Equation (6), see Sections 2.1 and 4 therein.

A. Proofs of Section 4

This section contains the proofs of the results from Section 4. We begin with auxiliary results in Section A.1; Theorem 4.1 and its Corollaries are proved in Sections A.2 and A.3, respectively.

A.1. Auxiliary results

The following proposition is used to prove Theorem 4.1 but is also of independent interest.

Proposition A.1. *Let $\mathbf{S} = (S_1, \dots, S_d)$ be a random vector with joint survival function \bar{G}_d as in (2.4) for some stdf ℓ . Then $1/\mathbf{S}$ belongs to the maximum domain of attraction of a multivariate extreme-value distribution with unit Fréchet margins and stdf ℓ .*

Proof of Proposition A.1. For the margins, recall that for each $i \in \{1, \dots, d\}$, $S_i \sim \text{Beta}(1, d - 1)$. The survival function of $1/S_i$ is thus given by $\bar{F}_{1/S_i}(s) = 1 - (1 - 1/s)^{d-1}$; it is easily seen that $\bar{F}_{1/S_i} \in \mathcal{R}_{-1}$. Now set $c_n = \{1 - (1 - 1/n)^{1/(d-1)}\}^{-1}$. From Equation 3.13 in Embrechts, Klüppelberg and Mikosch (1997), for all $s_i \in \mathbb{R}$, it then holds that $\Pr(1/S_i \leq c_n s_i) \rightarrow \Phi_1(s_i)$ as $n \rightarrow \infty$. Thus $1/\mathbf{S}$ is in the domain of attraction of a multivariate extreme-value distribution with unit Fréchet margins and stdf ℓ if and only if for all $\mathbf{s} \in \mathbb{R}_+^d$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \{1 - \Pr(1/S_1 \leq c_n s_1, \dots, 1/S_d \leq c_n s_d)\} \\ = \lim_{n \rightarrow \infty} n [1 - \bar{G}_d\{1/(c_n s_1), \dots, 1/(c_n s_d)\}] = \ell(1/s_1, \dots, 1/s_d). \end{aligned}$$

To show this, fix an arbitrary $\mathbf{s} \in \mathbb{R}_+^d$ and observe that because $c_n \rightarrow \infty$ as $n \rightarrow \infty$,

$$\bar{G}_d\{1/(c_n s_1), \dots, 1/(c_n s_d)\} = \{1 - (1/c_n)\ell(1/s_1, \dots, 1/s_d)\}^{d-1}$$

for all n sufficiently large. Now note that as $n \rightarrow \infty$, n/c_n^k converges to 0 for all $k \in \{2, \dots, d - 1\}$ and to $1/(d - 1)$ for $k = 1$. Consequently,

$$\begin{aligned} & \lim_{n \rightarrow \infty} n \left[1 - \left\{ 1 - (1/c_n)\ell(1/s_1, \dots, 1/s_d) \right\}^{d-1} \right] \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^{d-1} \binom{d-1}{k} (-1)^{k+1} \frac{n}{c_n^k} \ell^k(1/s_1, \dots, 1/s_d) = \ell(1/s_1, \dots, 1/s_d) \end{aligned}$$

as claimed. □

The following lemma determines the normalizing sequences needed for the proof of Theorem 4.1.

Lemma A.1. *Let $C_{G,\psi,\ell,Q}$ be a clustered Archimax copula such that Assumptions 4.1 and 4.2 are satisfied. Then the following hold:*

- (i) *For each $k \in \mathcal{D}_1$ and $i \in \{1, \dots, d_k\}$, $1/(R_k S_{ki}) \in \mathcal{M}(\Phi_{\rho_k})$. Recall that for $k \in \mathcal{D}_1$, $b_k = E\{1/Z_k^{\rho_k}\}$ where $Z_k \sim \text{Beta}(1, d_k - 1)$. Moreover, there exists a sequence of positive constants $\{a_{nk}\}$ such that for all $x > 0$, $n \Pr(1/R_k > a_{nk}x) \rightarrow x^{-\rho_k}$ as $n \rightarrow \infty$ and $n \Pr(1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x) \rightarrow x^{-\rho_k}$ as $n \rightarrow \infty$.*
- (ii) *For each $k \in \mathcal{D}_2$ and $i \in \{1, \dots, d_k\}$, $1/(R_k S_{ki}) \in \mathcal{M}(\Phi_1)$. Moreover, there exists a sequence of positive constants $\{a_{nk}\}$ such that for all $x > 0$, $n \Pr(1/S_{ki} > a_{nk}x) \rightarrow x^{-1}$ as $n \rightarrow \infty$ and $n \Pr(1/(R_k S_{ki}) > a_{nk} b_k x) \rightarrow x^{-1}$ as $n \rightarrow \infty$, where $b_k = E(1/R_k)$.*

Proof of Lemma A.1. (i) Let $k \in \mathcal{D}_1$ and $i \in \{1, \dots, d_k\}$. We then have $(1/R_k) \in \mathcal{M}(\Phi_{\rho_k})$ by assumption and $1/S_{ki} \in \mathcal{M}(\Phi_1)$ owing to the fact that $S_{ki} \sim \text{Beta}(1, d - 1)$. By Proposition 3.1.1 in Embrechts, Klüppelberg and Mikosch (1997), there exists a sequence of positive constants $\{a_{nk}\}$ such that for all $x > 0$, $n \Pr(1/R_k > a_{nk}x) \rightarrow x^{-\rho_k}$ as $n \rightarrow \infty$. Because $\rho_k < 1$, $E(1/S_{ki})^{\rho_k + \varepsilon} < \infty$ for some ε sufficiently small. Using the lemma of Breiman (1965) and the fact that $b_k = E\{1/S_{ki}^{\rho_k}\}$, we then have, for all $x > 0$ and with $\zeta_{kn} = a_{nk} b_k^{1/\rho_k}$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \Pr\left(\frac{1}{R_k S_{ki}} > \zeta_{kn} x\right) &= \lim_{n \rightarrow \infty} n \Pr\left(\frac{1}{R_k} > \zeta_{kn} x\right) \frac{\Pr\left(\frac{1}{R_k S_{ki}} > \zeta_{kn} x\right)}{\Pr\left(\frac{1}{R_k} > \zeta_{kn} x\right)} \\ &= (x b_k^{1/\rho_k})^{-\rho_k} b_k = x^{-\rho_k}. \end{aligned}$$

Indeed, $n \Pr(1/R_k > a_{nk} b_k^{1/\rho_k} x) \rightarrow (x b_k^{1/\rho_k})^{-\rho_k}$ as $n \rightarrow \infty$ by the choice of normalizing constants $\{a_{nk}\}$. The convergence of the fraction in the above display is due to Breiman’s lemma. The Fisher-Tippett-Gnedenko Theorem (Fisher and Tippett, 1928; Gnedenko, 1943) implies that since $1/R_k \in \mathcal{M}(\Phi_{\rho_k})$ and $\rho_k \in (0, 1)$, $\bar{F}_{1/R_k} \in \mathcal{R}_{-\rho_k}$. We also have that $1/S_{ki}$ and $1/R_k$ are independent, positive, and $E(1/S_{ki}^\gamma) < \infty$ for $\gamma \in (\rho_k, 1)$. By Breiman’s lemma, $1/(R_k S_{ki}) \in \mathcal{M}(\Phi_{\rho_k})$ and

$$\frac{\Pr\left(\frac{1}{R_k S_{ki}} > a_{nk} b_k^{1/\rho_k} x\right)}{\Pr\left(\frac{1}{R_k} > a_{nk} b_k^{1/\rho_k} x\right)} \rightarrow E(1/S_{ki}^{\rho_k}) = b_k$$

as $n \rightarrow \infty$.

(ii) Let $k \in \mathcal{D}_2$ and $i \in \{1, \dots, d_k\}$. The proof of the result relies again on Breiman's lemma; see also Proposition 2(b) of [Belzile and Nešlehová \(2017\)](#). Since $1/S_{ki} \in \mathcal{M}(\Phi_1)$, Proposition 3.1.1 in [Embrechts, Klüppelberg and Mikosch \(1997\)](#) implies that there exist sequences of positive constants $\{a_{nk}\}$ such that for all $x > 0$, $n \Pr(1/S_{ki} > a_{nk}x) \rightarrow x^{-1}$ as $n \rightarrow \infty$, and this for all $i = 1, \dots, d_k$. Recall that $b_k = E(1/R_k)$. Similarly to the proof of part (i), Breiman's lemma then implies that for all $x > 0$ and with $\zeta_{kn} = a_{nk}b_k$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \Pr\left(\frac{1}{R_k S_{ki}} > \zeta_{kn} x\right) &= \lim_{n \rightarrow \infty} n \Pr\left(\frac{1}{S_{ki}} > \zeta_{kn} x\right) \frac{\Pr\left(\frac{1}{R_k S_{ki}} > \zeta_{kn} x\right)}{\Pr\left(\frac{1}{S_{ki}} > \zeta_{kn} x\right)} \\ &= (x b_k)^{-1} b_k = x^{-1}. \end{aligned}$$

The convergence of the first part of the above is due to the choice of the normalizing constants $\{a_{nk}\}$. For the convergence of the second term, note that $\bar{F}_{1/S_{ki}} \in \mathcal{R}_{-1}$ and by assumption, $E\{1/R_k^{1+\epsilon_k}\}$ for some $\epsilon_k > 0$. Finally, since $1/S_{ki}$ and $1/R_k$ are independent and positive, Breiman's lemma implies that $1/(R_k S_{ki}) \in \mathcal{M}(\Phi_1)$ and that

$$\frac{\Pr\left(\frac{1}{R_k S_{ki}} > a_{nk} b_k x\right)}{\Pr\left(\frac{1}{S_{ki}} > a_{nk} b_k x\right)} \rightarrow E(1/R_k) = b_k$$

as $n \rightarrow \infty$. This completes the proof. \square

The lemma below establishes asymptotic independence between clusters in \mathcal{D}_1 and clusters in \mathcal{D}_2 .

Lemma A.2. *Suppose that $k \in \mathcal{D}_1$, $l \in \mathcal{D}_2$, $i \in \{1, \dots, d_k\}$ and $j \in \{1, \dots, d_l\}$. Let $\{a_{nk}\}$ and $\{a_{nj}\}$ be normalizing sequences as in Lemma A.1. As in Lemma A.1 (ii), let $b_l = E(1/R_l)$. Then for all $x, y > 0$,*

$$\lim_{n \rightarrow \infty} n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x, 1/(R_l S_{lj}) > a_{nl} b_l y\} = 0.$$

Proof. Fix $x, y > 0$ and recall that $\rho_k \in (0, 1)$. The probability of interest can be written as follows, for $\zeta_{nk} = a_{nk} b_k^{1/\rho_k}$ and $\eta_{nl} = a_{nl} b_l$,

$$\begin{aligned} &n \Pr\{1/(R_k S_{ki}) > \zeta_{nk} x, 1/(R_l S_{lj}) > \eta_{nl} y\} \\ &= \int_{\mathbb{R}_+^2} n \Pr(1/S_{ki} > \zeta_{nk} x r_k, 1/S_{lj} > \eta_{nl} y r_l) dF_{R_k, R_l}(r_k, r_l) \\ &= \int_{\mathbb{R}_+^2} n \Pr\{1/S_{ki} > \zeta_{nk} x r_k\} \Pr\{1/S_{lj} > \eta_{nl} y r_l\} dF_{R_k, R_l}(r_k, r_l), \end{aligned}$$

where the first equality is due to the independence between (R_k, R_l) and (S_{ki}, S_{lj}) and the last equality is due to the independence of S_{ki} and S_{lj} . Next, consider

the integrand as a sequence of functions $\{f_n\}$ defined on \mathbb{R}_+^2 . Observe that for each $r_k, r_l > 0$,

$$f_n(r_k, r_l) \leq g_n(r_k, r_l),$$

where $\{g_n\}$ is itself a sequence of functions on \mathbb{R}_+^2 defined by

$$g_n(r_k, r_l) = g_n(r_l) = n \Pr\{1/S_{lj} > a_{nl} b_l y r_l\}.$$

From the choice of $\{a_{nl}\}_{n \in \mathbb{N}}$, for all $r_k, r_l > 0$, $\lim_{n \rightarrow \infty} g_n(r_k, r_l) = g(r_k, r_l)$, where $g(r_k, r_l) = 1/(b_l y r_l)$. Moreover,

$$\int_{\mathbb{R}_+^2} g(r_k, r_l) dF_{R_k, R_l}(r_k, r_l) = \int_{\mathbb{R}_+^2} \frac{1}{b_l y r_l} dF_{R_k, R_l}(r_k, r_l) = \frac{1}{y},$$

and

$$\int_{\mathbb{R}_+^2} g_n(r_k, r_l) dF_{R_k, R_l}(r_k, r_l) = n \Pr\{1/(R_l S_{lj}) > a_{nl} b_l y\} \rightarrow \frac{1}{y}$$

as $n \rightarrow \infty$. We therefore have a sequence of nonnegative functions $\{g_n\}$ bounding $\{f_n\}$ from above such that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}_+^2} g_n(r_k, r_l) dF_{R_k, R_l}(r_k, r_l) = \int_{\mathbb{R}_+^2} \lim_{n \rightarrow \infty} g_n(r_k, r_l) dF_{R_k, R_l}(r_k, r_l).$$

Finally, note that

$$f_n(r_k, r_l) = n \Pr\{1/S_{ki} > a_{nk} b_k^{1/\rho_k} x r_k\} \Pr\{1/S_{lj} > a_{nl} b_l y r_l\} \rightarrow 0$$

as $n \rightarrow \infty$ since

$$\Pr\{1/S_{ki} > a_{nk} b_k^{1/\rho_k} x r_k\} \rightarrow 0 \quad \text{and} \quad n \Pr\{1/S_{lj} > a_{nl} b_l y r_l\} \rightarrow 1/(b_l y r_l)$$

as $n \rightarrow \infty$. The desired result then follows by the generalized Lebesgue dominated convergence theorem (see Theorem 1.21 in [Kallenberg \(2002\)](#), for example). \square

We are now ready to prove Theorem [4.1](#).

A.2. Proof of Theorem [4.1](#)

A random vector (Y_1, \dots, Y_d) is in the maximum domain of attraction of the extreme-value distribution H with Fréchet margins if and only if there exist sequences of positive constants $(a_{ni}) \in (0, \infty)$, $i \in \{1, \dots, d\}$, so that, for all $(y_1, \dots, y_d) \in \mathbb{R}_+^d$,

$$\lim_{n \rightarrow \infty} n \{1 - \Pr(Y_1 \leq a_{n1} y_1, \dots, Y_d \leq a_{nd} y_d)\} = -\ln H(y_1, \dots, y_d).$$

This is a multivariate extension of Proposition 3.1.1 in [Embrechts, Klüppelberg and Mikosch \(1997\)](#), as used in [Belzile and Nešlehová \(2017\)](#). For each

$k \in \{1 \dots, K\}$, set the sequences $\{a_{nk}\}$ as done in Lemma A.1. Then the fact that the marginals of H are Fréchet follows from the said Lemma. With the normalizing constants now set, the limit of interest is, for any fixed $\mathbf{x} \in \mathbb{R}_+^d$ and, for convenience, let $\mathcal{J} = \{(k, i) : k = 1, \dots, K, i = 1 \dots, d_k\}$,

$$\lim_{n \rightarrow \infty} n \left[1 - \Pr \left(\bigcap_{(k,i) \in \mathcal{J}} \left\{ \frac{1}{R_k S_{ki}} \leq a_{nk} b_k^{1/\rho_k} x_{ki} \right\} \right) \right], \tag{A.1}$$

where for $k \in \mathcal{D}_2$, $b_k = E(1/R_k)$ as in Lemma A.1 (ii) and for ease of notation, $\rho_k = 1$. Letting $\mathcal{P}(\mathcal{J})$ denote the power set of \mathcal{J} , (A.1) can be rewritten as

$$\lim_{n \rightarrow \infty} n \sum_{p \in \mathcal{P}(\mathcal{J})} (-1)^{|p|+1} \Pr \left(\bigcap_{(k,i) \in p} \left\{ \frac{1}{R_k S_{ki}} > a_{nk} b_k^{1/\rho_k} x_{ki} \right\} \right). \tag{A.2}$$

We now consider the subset $\mathcal{P}(\mathcal{J})|_{\mathcal{D}_1, \mathcal{D}_2}$ of $\mathcal{P}(\mathcal{J})$ such that $p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1, \mathcal{D}_2}$ if and only if there exists $(k, i), (l, j) \in p$ so that $k \in \mathcal{D}_1$ and $l \in \mathcal{D}_2$. As it turns out, the summands in (A.2) for which $p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1, \mathcal{D}_2}$ are asymptotically negligible. To see this, fix an arbitrary $p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1, \mathcal{D}_2}$ and pick $(k, i), (l, j) \in p$ so that $k \in \mathcal{D}_1$ and $l \in \mathcal{D}_2$. Then for all $\{x_{ca} : (c, a) \in p\} \in \mathbb{R}_+^{|p|}$,

$$\begin{aligned} n \Pr \left(\bigcap_{(c,a) \in p} \left\{ \frac{1}{R_c S_{ca}} > a_{nc} b_c^{1/\rho_c} x_{ca} \right\} \right) \\ \leq n \Pr \{ 1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x_{ki}, 1/(R_l S_{lj}) > a_{nl} b_l^{1/\rho_l} x_{lj} \} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ by Lemma A.2.

Now, let $\mathcal{P}(\mathcal{J})|_{\mathcal{D}_1}$ be the subset of $\mathcal{P}(\mathcal{J})$ such that $p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1}$ if and only if $(c, a) \in p$ implies that $c \in \mathcal{D}_1$. In other words, $\mathcal{P}(\mathcal{J})|_{\mathcal{D}_1}$ contains only sets of indices (c, a) with $c \in \mathcal{D}_1$. Let $N_1 = \sum_{k \in \mathcal{D}_1} d_k$, define $\mathbf{S}^{(1)} = (\mathbf{S}_k : k \in \mathcal{D}_1)$, and rewrite the summands in (A.2) with $p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1}$ as follows:

$$\begin{aligned} n \sum_{p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1}} (-1)^{|p|+1} \Pr \left(\bigcap_{(c,a) \in p} \left\{ \frac{1}{R_c S_{ca}} > a_{nc} b_c^{1/\rho_c} x_{ca} \right\} \right) \\ = n \left[1 - \Pr \left(\bigcap_{k \in \mathcal{D}_1} \bigcap_{1 \leq i \leq d_k} \left\{ \frac{1}{R_k S_{ki}} \leq a_{nk} b_k^{1/\rho_k} x_{ki} \right\} \right) \right] \\ = \int_{[0,1]^{N_1}} n \left[1 - \Pr \left(\bigcap_{k \in \mathcal{D}_1} \bigcap_{1 \leq i \leq d_k} \left\{ \frac{1}{R_k} \leq a_{nk} b_k^{1/\rho_k} x_{ki} s_{ki} \right\} \right) \right] dF_{\mathbf{S}^{(1)}}(\mathbf{s}^{(1)}) \\ = \int_{[0,1]^{N_1}} n \left[1 - \Pr \left(\bigcap_{k \in \mathcal{D}_1} \left\{ \frac{1}{R_k} \leq a_{nk} b_k^{1/\rho_k} \min_{1 \leq i \leq d_k} (x_{ki} s_{ki}) \right\} \right) \right] dF_{\mathbf{S}^{(1)}}(\mathbf{s}^{(1)}), \end{aligned}$$

where $\mathbf{s}^{(1)} \in [0, 1]^{N_1}$ is indexed so that $\mathbf{s}^{(1)} = (s_{ki} : k \in \mathcal{D}_1, 1 \leq i \leq d_k)$. Now consider the integrand as a sequence of functions $\{f_n\}$ defined on $[0, 1]^{N_1}$

and observe that for each $n \in \mathbb{N}$, $0 \leq f_n \leq g_n$, where g_n is given, for each $\mathbf{s}^{(1)} \in [0, 1]^{N_1}$ as above, by

$$g_n(\mathbf{s}^{(1)}) = n \sum_{k \in \mathcal{D}_1} \sum_{i=1}^{d_k} \Pr \left(1/R_k > a_{nk} b_k^{1/\rho_k} x_{ki} s_{ki} \right)$$

Clearly, $g_n(\mathbf{s}^{(1)}) \rightarrow g(\mathbf{s}^{(1)})$ as $n \rightarrow \infty$, where

$$g(\mathbf{s}^{(1)}) = \sum_{k \in \mathcal{D}_1} \sum_{i=1}^{d_k} \frac{1}{b_k \{x_{ki} s_{ki}\}^{\rho_k}}$$

with

$$\int_{[0,1]^{N_1}} g(\mathbf{s}^{(1)}) dF_{\mathcal{S}^{(1)}}(\mathbf{s}^{(1)}) = \sum_{k \in \mathcal{D}_1} \sum_{i=1}^{d_k} x_{ki}^{-\rho_k}.$$

Moreover,

$$\begin{aligned} \int_{[0,1]^{N_1}} g_n(\mathbf{s}^{(1)}) dF_{\mathcal{S}^{(1)}}(\mathbf{s}^{(1)}) &= n \left\{ \sum_{k \in \mathcal{D}_1} \sum_{i=1}^{d_k} \Pr \left(1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x_{ki} \right) \right\} \\ &\rightarrow \sum_{k \in \mathcal{D}_1} \sum_{i=1}^{d_k} x_{ki}^{-\rho_k} \end{aligned}$$

as $n \rightarrow \infty$. Therefore, we have a sequence of majorants $\{g_n\}$ for which the identity $\lim_{n \rightarrow \infty} \int g_n = \int \lim_{n \rightarrow \infty} g_n$ is satisfied. Now recall that the vector of reciprocal distortions $1/\mathbf{R}$ has a limiting stdf $\ell_{1/\mathbf{R}}$ defined in terms of the positive, unit-mean variables W_1, \dots, W_K in Assumption 4.2. Therefore, $f_n \rightarrow f$ pointwise, where for all $\mathbf{s}^{(1)} \in [0, 1]^{N_1}$,

$$f(\mathbf{s}^{(1)}) = \mathbb{E} \left\{ \max_{k \in \mathcal{D}_1} \frac{W_k}{b_k \min_{1 \leq i \leq d_k} (x_{ki} s_{ki})^{\rho_k}} \right\}.$$

Now, integrating over the $\mathbf{s}^{(1)}$ yields

$$\int_{[0,1]^{N_1}} f(\mathbf{s}^{(1)}) dF_{\mathcal{S}^{(1)}}(\mathbf{s}^{(1)}) = \mathbb{E} \left\{ \max_{k \in \mathcal{D}_1} \frac{W_k}{b_k (S_{ki} x_{ki})^{\rho_k}} \right\}.$$

Using the generalized Lebesgue dominated convergence theorem, we can thus conclude that for all $\mathbf{x}^{(1)} = (\mathbf{x}_k : k \in \mathcal{D}_1) \in \mathbb{R}_+^{N_1}$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \sum_{p \in \mathcal{P}(\mathcal{J})|_{\mathcal{D}_1}} (-1)^{|p|+1} \Pr \left(\bigcap_{(c,a) \in p} \{1/(R_c S_{ca}) > a_n b_c^{\rho_c} x_{ca}\} \right) \\ = \mathbb{E} \left\{ \max_{k \in \mathcal{D}_1} \frac{W_k}{b_k (S_{ki} x_{ki})^{\rho_k}} \right\}. \end{aligned}$$

Analogously to $\mathcal{P}(\mathcal{I})|_{\mathcal{D}_1}$, let $\mathcal{P}(\mathcal{I})|_{\mathcal{D}_2}$ contain only sets of indices (c, a) with $c \in \mathcal{D}_2$. Let $K_2 = |\mathcal{D}_2|$ and $N_2 = \sum_{k \in \mathcal{D}_2} d_k$, define $\mathbf{R}^{(2)} = (R_k : k \in \mathcal{D}_2)$, and recall that $\rho_k = 1$ for all $k \in \mathcal{D}_2$. Next, for $\mathbf{x}^{(2)} = (\mathbf{x}_k : k \in \mathcal{D}_2) \in \mathbb{R}_+^{N_2}$, rewrite the summands of (A.2) with $p \in \mathcal{P}(\mathcal{I})|_{\mathcal{D}_2}$ as follows:

$$\begin{aligned} & n \sum_{p \in \mathcal{P}(\mathcal{I})|_{\mathcal{D}_2}} (-1)^{|p|+1} \Pr\left(\bigcap_{(c,a) \in p} \{1/(R_c S_{ca}) > a_{nc} b_c^{1/\rho_c} x_{ca}\}\right) \\ &= n \left[1 - \Pr\left(\bigcap_{k \in \mathcal{D}_2} \bigcap_{1 \leq i \leq d_k} \{1/(R_k S_{ki}) \leq a_{nk} b_k x_{ki}\}\right)\right] \\ &= \int_{\mathbb{R}_+^{K_2}} n \left[1 - \Pr\left(\bigcap_{k \in \mathcal{D}_2} \bigcap_{1 \leq i \leq d_k} \{1/S_{ki} \leq a_{nk} b_k x_{ki} r_k\}\right)\right] dF_{\mathbf{R}^{(2)}}(\mathbf{r}), \end{aligned}$$

where $\mathbf{r}^{(2)} = (r_k : k \in \mathcal{D}_2)$ is indexed as $\mathbf{R}^{(2)}$. Now consider the integrand as a sequence of functions $\{f_n\}$ defined on $\mathbb{R}_+^{K_2}$ and observe that for each $n \in \mathbb{N}$, $0 \leq f_n \leq g_n$, where g_n is given, for all $\mathbf{r}^{(2)} \in \mathbb{R}_+^{K_2}$, by

$$g_n(\mathbf{r}^{(2)}) = n \left\{ \sum_{k \in \mathcal{D}_2} \sum_{i=1}^{d_k} \Pr(1/S_{ki} > a_{nk} b_k x_{ki} r_k) \right\}$$

Clearly, for all $\mathbf{r}^{(2)} \in \mathbb{R}_+^{K_2}$ and as $n \rightarrow \infty$,

$$g_n(\mathbf{r}^{(2)}) \rightarrow g(\mathbf{r}^{(2)}) = \sum_{k \in \mathcal{D}_2} \sum_{i=1}^{d_k} \frac{1}{b_k x_{ki} r_k}.$$

Furthermore,

$$\int_{\mathbb{R}_+^{K_2}} g(\mathbf{r}^{(2)}) dF_{\mathbf{R}^{(2)}}(\mathbf{r}^{(2)}) = \sum_{k \in \mathcal{D}_2} \sum_{i=1}^{d_k} \frac{1}{x_{ki}}$$

and, as $n \rightarrow \infty$,

$$\begin{aligned} \int_{\mathbb{R}_+^{K_2}} g_n(\mathbf{r}^{(2)}) dF_{\mathbf{R}^{(2)}}(\mathbf{r}^{(2)}) &= n \left\{ \sum_{k \in \mathcal{D}_2} \sum_{i=1}^{d_k} \Pr(1/(R_k S_{ki}) > a_{nk} b_k x_{ki}) \right\} \\ &\rightarrow \sum_{k \in \mathcal{D}_1} \sum_{i=1}^{d_k} \frac{1}{x_{ki}}. \end{aligned}$$

Analogously to the treatment of $\mathcal{P}(\mathcal{I})|_{\mathcal{D}_1}$, we have a sequence of majorants $\{g_n\}$ such that $\lim_{n \rightarrow \infty} \int g_n = \int \lim_{n \rightarrow \infty} g_n$. It remains to determine the limit of the sequence of functions $\{f_n\}$ defined for all $\mathbf{r}^{(2)} \in \mathbb{R}_+^{K_2}$ by

$$n \left[1 - \Pr\left(\bigcap_{k \in \mathcal{D}_2} \bigcap_{1 \leq i \leq d_k} \{1/S_{ki} > a_{nk} b_k x_{ki} r_k\}\right)\right].$$

By assumption, \mathbf{S}_k and \mathbf{S}_l are independent if $k \neq l$ and are therefore asymptotically independent as well. Using Proposition A.1 and the fact that $1/S_{ki} \in \mathcal{M}(\Phi_1)$ for all $k \in \{1, \dots, K\}$ and $i \in \{1, \dots, d_k\}$, one has that $f_n \rightarrow f$ pointwise, where for all $\mathbf{r}^{(2)} \in \mathbb{R}_+^{K_2}$,

$$\begin{aligned} f(\mathbf{r}^{(2)}) &= \sum_{k \in \mathcal{D}_2} \ell_k \{(b_k x_{k1} r_k)^{-1}, \dots, (b_k x_{kd_k} r_k)^{-1}\} \\ &= \sum_{k \in \mathcal{D}_2} (b_k r_k)^{-1} \ell_k(x_{k1}^{-1}, \dots, x_{kd_k}^{-1}) . \end{aligned}$$

Integrating the limit f yields

$$\int_{\mathbb{R}_+^{K_2}} f(\mathbf{r}^{(2)}) dF_{\mathbf{R}^{(2)}}(\mathbf{r}^{(2)}) = \sum_{k \in \mathcal{D}_2} \ell_k(x_{k1}^{-1}, \dots, x_{kd_k}^{-1}) .$$

Thus for all $\mathbf{x} \in \mathbb{R}_+^d$, the limit (A.1) is equal to

$$\mathbb{E} \left\{ \max_{\substack{k \in \mathcal{D}_1 \\ 1 \leq i \leq d_k}} \frac{W_k}{b_k (S_{ki} x_{ki})^{\rho_k}} \right\} + \sum_{k \in \mathcal{D}_2} \ell_k(x_{k1}^{-1}, \dots, x_{kd_k}^{-1}) .$$

Recalling that $1/(R_k S_{ki}) \in \mathcal{M}(\Phi_{\rho_k})$, one obtains (4.1) by plugging in the appropriate Fréchet margins. \square

A.3. Proofs of Corollaries 4.1 and 4.2

Proof of Corollary 4.1. Let $K_1 = |\mathcal{D}_1|$ and recall that $\mathbb{E}[\max_{k \in \mathcal{D}_1} y_k W_k]$ is the limiting stdf of $\{1/R_k : k \in \mathcal{D}_1\}$, defined for all $(y_1, \dots, y_{K_1}) \in \mathbb{R}_+^{K_1}$. Letting $\{W_k : k \in \mathcal{D}_1\}$ be a (uniformly) random permutation of $(K_1, 0, \dots, 0)$ yields the independence stdf $\mathbb{E}[\max_{k \in \mathcal{D}_1} y_k W_k] = y_1 + \dots + y_{K_1}$. Due to the fact that, for all $k, l \in \{1, \dots, d\}$ and $i \in \{1, \dots, d_l\}$, W_k is independent of S_{li} , plugging this into (4.1) yields, for all $\mathbf{x} \in \mathbb{R}_+^d$,

$$\ell_{\mathcal{G}, \psi, \ell, Q}(\mathbf{x}) = \sum_{k \in \mathcal{D}_1} \mathbb{E} \left(\max_{i=1, \dots, d_k} \frac{x_{ki}}{b_k S_{ki}^{\rho_k}} \right) + \sum_{k \in \mathcal{D}_2} \ell_k(\mathbf{x}_k) .$$

By Proposition 6.1 in Charpentier et al. (2014), as mentioned in the Section 1, for each $k \in \mathcal{D}_1$,

$$\mathbb{E} \left(\max_{1 \leq i \leq d_k} \frac{x_{ki}}{b_k S_{ki}^{\rho_k}} \right) = \ell_k^{\rho_k}(\mathbf{x}_k^{1/\rho_k}) .$$

This completes the proof. \square

Proof of Corollary 4.2. Observe first that by Assumption 4.2 and the fact that the variables W_k have unit mean,

$$\ell_{1/\mathbf{R}}(\mathbf{x}') = \mathbb{E} \left\{ \max_{1 \leq k \leq K} (x'_k W_k) \right\}$$

$$\begin{aligned}
 &\leq \mathbb{E}\left\{\max_{k \in \mathcal{D}_1}(x'_k W_k) + \sum_{k \in \mathcal{D}_2} x'_k W_k\right\} \\
 &= \mathbb{E}\left\{\max_{k \in \mathcal{D}_1}(x'_k W_k)\right\} + \sum_{k \in \mathcal{D}_2} x'_k. \tag{A.3}
 \end{aligned}$$

Next, note that $\ell_{\mathcal{G},\psi,\ell,Q}(\mathbf{x}) = \mathcal{A}(\mathbf{x}') + \mathcal{B}(\mathbf{x}')$, where

$$\mathcal{A}(\mathbf{x}') = \sum_{k \in \mathcal{D}_2} \ell_k(\mathbf{x}_k) = \sum_{k \in \mathcal{D}_2} x_{i_k} = \sum_{k \in \mathcal{D}_2} x'_k$$

and

$$\mathcal{B}(\mathbf{x}) = \mathbb{E}\left(\max_{\substack{k \in \mathcal{D}_1 \\ 1 \leq j \leq d_k}} \frac{\mathbf{x}_{kj} W_k}{b_k S_{kj}^{\rho_k}}\right) = \mathbb{E}\left(\max_{k \in \mathcal{D}_1} \frac{x'_k W_k}{b_k S_{kj}^{\rho_k}}\right).$$

Because for each $k \in \mathcal{D}_1$, $b_k = \mathbb{E}(1/S_{kj}^{\rho_k})$, we have that for any $\mathbf{w} \in \mathbb{R}_+^K$ and $k \in \mathcal{D}_1$,

$$\mathbb{E}\left(\max_{k \in \mathcal{D}_1} \frac{x'_k w_k}{b_k S_{kj}^{\rho_k}}\right) \geq \mathbb{E}\left(\frac{x'_k w_k}{b_k S_{kj}^{\rho_k}}\right) = x'_k w_k,$$

so that

$$\mathbb{E}\left(\max_{k \in \mathcal{D}_1} \frac{x'_k w_k}{b_k S_{kj}^{\rho_k}}\right) \geq \max_{k \in \mathcal{D}_1} (x'_k w_k).$$

This implies that

$$\mathcal{B}(\mathbf{x}') \geq \mathbb{E}\left\{\max_{k \in \mathcal{D}_1}(x'_k W_k)\right\},$$

which together with (A.3) yields the desired result. \square

B. Jackknife estimation of the covariance matrix of \mathbf{T}

Here, we provide a consistent estimator of the asymptotic variance Σ of $\sqrt{n}\mathbf{T}$, where \mathbf{T} is as in (6.3). Recall the definition of \mathcal{I} in (6.1). For each $\nu \in \{1, \dots, n\}$ and $(i, j, k) \in \mathcal{I}$, let $\hat{\theta}_{ij\nu}$ be a version of $\hat{\theta}_{ij}$ based on all but the ν th observation; $\bar{\theta}_{k\nu}$ be the average of all $\hat{\theta}_{ij\nu}$ such that, given k , $(i, j, k) \in \mathcal{I}$; $T_{ijk\nu} = \hat{\theta}_{ij\nu} - \bar{\theta}_{k\nu}$; and $\mathbf{T}_\nu = (T_{\nu\nu})_{\nu \in \mathcal{I}}$. In particular, \mathbf{T}_ν is a version of \mathbf{T} based on all but the ν th observation. It then follows from Theorem 9 of Arvesen (1969) that, for $\mathbf{T}_\nu^* = n\mathbf{T} - (n-1)\mathbf{T}_\nu$ and $\mathbf{T}_\bullet^* = (1/n) \sum_{\nu=1}^n \mathbf{T}_\nu^*$,

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{\nu=1}^n (\mathbf{T}_\nu^* - \mathbf{T}_\bullet^*)(\mathbf{T}_\nu^* - \mathbf{T}_\bullet^*)^\top \tag{B.1}$$

is a consistent estimator of Σ .

C. Conjectured extension of Theorem 4.1

As it is stated, Theorem 4.1 does not account for the boundary case when $1/R_k \in \mathcal{M}(\Phi_1)$, which can occur. It would thus be desirable to replace Assumption 4.1 of Theorem 4.1 by the following requirement.

Assumption C.1. For a clustered Archimax copula as in Definition 3.1, assume that $\{1, \dots, K\}$ is the union of disjoint sets \mathcal{D}_1 , \mathcal{D}_2 , and \mathcal{D}_3 such that

- (i) $k \in \mathcal{D}_1$ if and only if $1/R_k \in \mathcal{M}(\Phi_{\rho_k})$ for some $\rho_k \in (0, 1)$.
- (ii) $k \in \mathcal{D}_2$ if and only if there exists an $\epsilon_k > 0$ such that $E\{1/R_k^{1+\epsilon_k}\} < \infty$.
- (iii) $k \in \mathcal{D}_3$ if and only if $1/R_k \in \mathcal{M}(\Phi_1)$ and $E(1/R_k) = \infty$.

We conjecture that the variables whose distortions are in \mathcal{D}_3 have the same asymptotic behavior as those whose distortions are in \mathcal{D}_2 . More precisely, we surmise that the following statement holds.

Conjecture C.1. Let $C_{\mathcal{G}, \psi, \ell, \mathcal{Q}}$ be a clustered Archimax copula such that Assumptions C.1 and 4.2 hold. For $k \in \mathcal{D}_1$, let $b_k = E(1/Z_k^{\rho_k})$, $Z_k \sim \text{Beta}(1, d_k - 1)$. Then $1/\mathbf{X} \in \mathcal{M}(H)$ with $1/X_{ki} \in \mathcal{M}(H_{ki})$, where $H_{ki} = \Phi_{\rho_k}$ for $k \in \mathcal{D}_1$ and $i \in \{1, \dots, d_k\}$ and $H_{ki} = \Phi_1$ for $k \in \mathcal{D}_2 \cup \mathcal{D}_3$ and $i \in \{1, \dots, d_k\}$. The stdf of H is given for all $\mathbf{x} \in \mathbb{R}_+^d$ by

$$\ell_{\mathcal{G}, \psi, \ell, \mathcal{Q}}(\mathbf{x}) = E \left(\max_{\substack{k \in \mathcal{D}_1 \\ 1 \leq i \leq d_k}} \frac{x_{ki} W_k}{b_k S_{ki}^{\rho_k}} \right) + \sum_{k \in \mathcal{D}_2 \cup \mathcal{D}_3} \ell_k(\mathbf{x}_k). \quad (\text{C.1})$$

One part of Conjecture C.1 is clear, namely that $H_{ki} = \Phi_1$ for $k \in \mathcal{D}_3$. Indeed, for any such k , the Corollary to Theorem 3 in Embrechts and Goldie (1980) implies that $1/(R_k S_{ki}) \in \mathcal{M}(\Phi_1)$. So one can again find a sequence $\{a_{nk}\}$ of positive constants ensuring that for all $x \in \mathbb{R}_+$, $n \Pr\{1/(R_k S_{ki}) > a_{nk}x\} \rightarrow 1/x$ as $n \rightarrow \infty$. The main difficulty in establishing the validity of Conjecture C.1 is the fact that, for $k \in \mathcal{D}_3$ and $i \in \{1, \dots, d_k\}$, the relation between the above normalizing sequence $\{a_{nk}\}$ and the normalizing sequences for $1/R_k$ and $1/S_{ki}$ is unclear. In order to prove the conjectured result, it suffices to prove three sister lemmas which are reported below. The first two, Lemmas C.1 and C.2, are analogous to Lemma A.2 and are proved therein. The third, Conjecture C.2, that states asymptotic independence between different clusters in \mathcal{D}_3 , is the missing result that if established would prove Conjecture C.1.

Lemma C.1. Under the hypothesis of Conjecture C.1, suppose that $k \in \mathcal{D}_1$, $l \in \mathcal{D}_3$, $i \in \{1, \dots, d_k\}$ and $j \in \{1, \dots, d_l\}$. Let $\{a_{nk}\}$ be a sequence of positive constants such that for all $x > 0$, $n \Pr(1/R_k > a_{nk}x) \rightarrow x^{-\rho_k}$ as $n \rightarrow \infty$ and $n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x\} \rightarrow x^{-\rho_k}$ as $n \rightarrow \infty$. Furthermore, let $\{a_{nl}\}$ be a sequence of positive constants so that for all $x > 0$, $n \Pr(1/(R_l S_{lj}) > a_{nl}x) \rightarrow 1/x$ as $n \rightarrow \infty$. Then for all $x, y \in \mathbb{R}_+$,

$$\lim_{n \rightarrow \infty} n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x, 1/(R_l S_{lj}) > a_{nl}y\} = 0.$$

Proof of Lemma C.1. The proof is quite similar to the one of Lemma A.2. Observe first that the assumed sequences $\{a_{nk}\}$ and $\{a_{nl}\}$ indeed exist, by Lemma A.1 and the discussion in the paragraph following Conjecture C.1. Fix some arbitrary $x, y > 0$ and recall that $\rho_k \in (0, 1)$. The probability of interest can be written as follows

$$\begin{aligned} & n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x, 1/(R_l S_{lj}) > a_{nl} b_l y\} \\ &= \int_{(0,1)^2} n \Pr\{1/R_k > a_{nk} b_k^{1/\rho_k} x s_{ki}, 1/R_l > a_{nl} y s_{lj}\} dF_{S_{ki}, S_{lj}}(s_{ki}, s_{lj}). \end{aligned}$$

Consider the integrand as a function f_n defined on $(0, 1)^2$ and note that for all $n \in \mathbb{N}$, $0 \leq f_n \leq g_n$, where g_n is given, for all $(s_{ki}, s_{lj}) \in (0, 1)^2$ by

$$g_n(s_{ki}, s_{lj}) = g_n(s_{ki}) = n \Pr\{1/R_k > a_{nk} b_k^{1/\rho_k} x s_{ki}\}.$$

As in the proof of Lemma A.2, for all $(s_{ki}, s_{lj}) \in (0, 1)^2$,

$$\lim_{n \rightarrow \infty} g_n(s_{ki}, s_{lj}) = g(s_{ki}, s_{lj}) = 1/\{b_k(x s_{ki})^{\rho_k}\}.$$

Moreover,

$$\int_{(0,1)^2} g(s_{ki}, s_{lj}) dF_{S_{ki}, S_{lj}}(s_{ki}, s_{lj}) = \frac{1}{x^{\rho_k}}$$

and

$$\int_{(0,1)^2} g_n(s_{ki}, s_{lj}) dF_{S_{ki}, S_{lj}}(s_{ki}, s_{lj}) = n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x\} \rightarrow \frac{1}{x^{\rho_k}}$$

as $n \rightarrow \infty$. We therefore have a sequence of functions $\{g_n\}$ bounding $\{f_n\}$ from above such that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{(0,1)^2} g_n(s_{ki}, s_{lj}) dF_{S_{ki}, S_{lj}}(s_{ki}, s_{lj}) \\ &= \int_{(0,1)^2} \lim_{n \rightarrow \infty} g_n(s_{ki}, s_{lj}) dF_{S_{ki}, S_{lj}}(s_{ki}, s_{lj}). \end{aligned}$$

Finally, note that

$$f_n(s_{ki}, s_{lj}) = n \Pr\{1/R_k > a_{nk} b_k^{1/\rho_k} x s_{ki}, 1/R_l > a_{nl} b_l^{1/\rho_l} y s_{lj}\} \rightarrow 0$$

as $n \rightarrow \infty$ since $n \Pr\{1/R_k > a_{nk} b_k^{1/\rho_k} x s_{ki}\} \rightarrow \{b_k^{1/\rho_k} x s_{ki}\}^{-\rho_k}$ and $\Pr\{1/R_l > a_{nl} b_l^{1/\rho_l} y s_{lj}\} \rightarrow 0$ as $n \rightarrow \infty$. The desired result then follows from the generalized Lebesgue dominated convergence theorem. \square

Lemma C.2. *Under the hypothesis of Conjecture C.1, suppose that $k \in \mathcal{D}_2$, $l \in \mathcal{D}_3$, $i \in \{1, \dots, d_k\}$ and $j \in \{1, \dots, d_l\}$. Let $\{a_{nk}\}$ such that for all $x > 0$, $n \Pr(1/S_{ki} > a_{nk} x) \rightarrow x^{-1}$ as $n \rightarrow \infty$ and $n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k x\} \rightarrow x^{-1}$ as $n \rightarrow \infty$. Furthermore, let $\{a_{nl}\}$ be a sequence of positive constants so that for all $x > 0$, $n \Pr\{1/(R_l S_{lj}) > a_{nl} x\} \rightarrow x^{-1}$ as $n \rightarrow \infty$. Then for all $x, y \in \mathbb{R}_+$,*

$$\lim_{n \rightarrow \infty} n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x, 1/(R_l S_{lj}) > a_{nl} y\} = 0.$$

Proof of Lemma C.2. This proof is almost exactly the same as the proof of Lemma A.2. Again, the existence of the norming constants $\{a_{nk}\}$ and $\{a_{nl}\}$ follows from Lemma A.1 and the discussion in the paragraph following Conjecture C.1. Fix some arbitrary $x, y > 0$. We are interested in the limit as $n \rightarrow \infty$ of

$$\begin{aligned} & n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k x, 1/(R_l S_{lj}) > a_{nl} y\} \\ &= \int_{\mathbb{R}_+^2} n \Pr\{1/S_{ki} > a_{nk} b_k x r_k\} \Pr\{1/S_{lj} > a_{nl} y r_l\} dF_{R_k, R_l}(r_k, r_l). \end{aligned}$$

Consider the integrand as a function f_n defined on \mathbb{R}_+^2 . Observe that for each $n \in \mathbb{N}$, $0 \leq f_n \leq g_n$ where for all $(r_k, r_l) \in \mathbb{R}_+^2$,

$$g_n(r_k, r_l) = g_n(r_k) = n \Pr\{1/S_{ki} > a_{nk} b_k x r_k\}.$$

From the choice of $\{a_{nk}\}$, for all $(r_k, r_l) \in \mathbb{R}_+^2$,

$$\lim_{n \rightarrow \infty} g_n(r_k, r_l) = g(r_k, r_l) = 1/(b_k x r_k).$$

Moreover, since $b_k = E(1/R_k)$,

$$\int_{\mathbb{R}_+^2} g(r_k, r_l) dF_{R_k, R_l}(r_k, r_l) = \int_{\mathbb{R}_+^2} \frac{1}{b_k x r_k} dF_{R_k, R_l}(r_k, r_l) = \frac{1}{x}.$$

and

$$\int_{\mathbb{R}_+^2} g_n(r_k, r_l) dF_{R_k, R_l}(r_k, r_l) = n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k x\} \rightarrow \frac{1}{x}$$

as $n \rightarrow \infty$. We therefore have a sequence of functions $\{g_n\}$ bounding $\{f_n\}$ from above such that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}_+^2} g_n(r_k, r_l) dF_{R_k, R_l}(r_k, r_l) = \int_{\mathbb{R}_+^2} \lim_{n \rightarrow \infty} g_n(r_k, r_l) dF_{R_k, R_l}(r_k, r_l).$$

Finally, note that

$$f_n(r_k, r_l) = n \Pr\{1/S_{ki} > a_{nk} b_k x r_k\} \Pr\{1/S_{lj} > a_{nl} y r_l\} \rightarrow 0$$

as $n \rightarrow \infty$ since

$$n \Pr\{1/S_{ki} > a_{nk} b_k x r_k\} \rightarrow 1/\{b_k x r_k\} \quad \text{and} \quad \Pr\{1/S_{lj} > a_{nl} y r_l\} \rightarrow 0$$

as $n \rightarrow \infty$. Using the generalized Lebesgue dominated convergence theorem concludes the proof. \square

Conjecture C.2. *Under the hypothesis of Conjecture C.1, suppose that $k, l \in \mathcal{D}_3$, $i \in \{1, \dots, d_k\}$ and $j \in \{1, \dots, d_l\}$. Let $\{a_{nk}\}$ and $\{a_{nl}\}$ be sequences of positive constants such that for all $x > 0$, $n \Pr(1/(R_k S_{ki}) > a_{nk} x) \rightarrow x^{-1}$ and $n \Pr(1/(R_l S_{lj}) > a_{nl} x) \rightarrow x^{-1}$ as $n \rightarrow \infty$. Then for all $x, y \in \mathbb{R}_+$,*

$$\lim_{n \rightarrow \infty} n \Pr\{1/(R_k S_{ki}) > a_{nk} b_k^{1/\rho_k} x, 1/(R_l S_{lj}) > a_{nl} y\} = 0.$$

D. Additional figures for Section 5

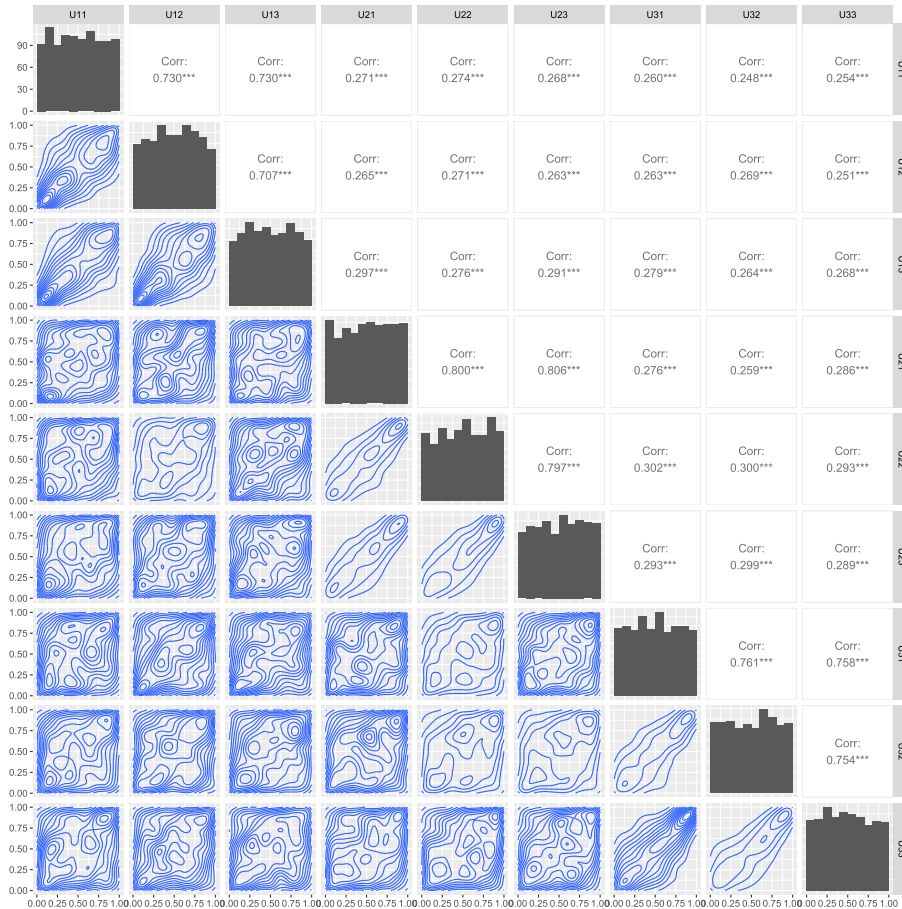


FIG D.1. (Model A) Pairwise plots of a sample of size $n = 1000$ from the copula $C_{\mathcal{G}, \psi, \ell, Q}$ where $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\} = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$. $\psi = \{\psi_{\theta_1}, \psi_{\theta_2}, \psi_{\theta_3}\}$ with $(\theta_1, \theta_2, \theta_3) = (1.5, 1.5, 2)$ where ψ_{θ_1} is Clayton while $\psi_{\theta_2}, \psi_{\theta_3}$ are Joe. $\ell = \{\ell_{\vartheta_1}, \ell_{\vartheta_2}, \ell_{\vartheta_3}\}$ with $(\vartheta_1, \vartheta_2, \vartheta_3) = (1.25, 2, 1.5)$ where $\ell_{\vartheta_1}, \ell_{\vartheta_2}, \ell_{\vartheta_3}$ are Gumbel-Hougaard. The radial survival copula D is trivariate Gaussian with correlations all equal to 0.5. Upper: linear correlation, Lower: contour density, Diagonal: univariate histogram.

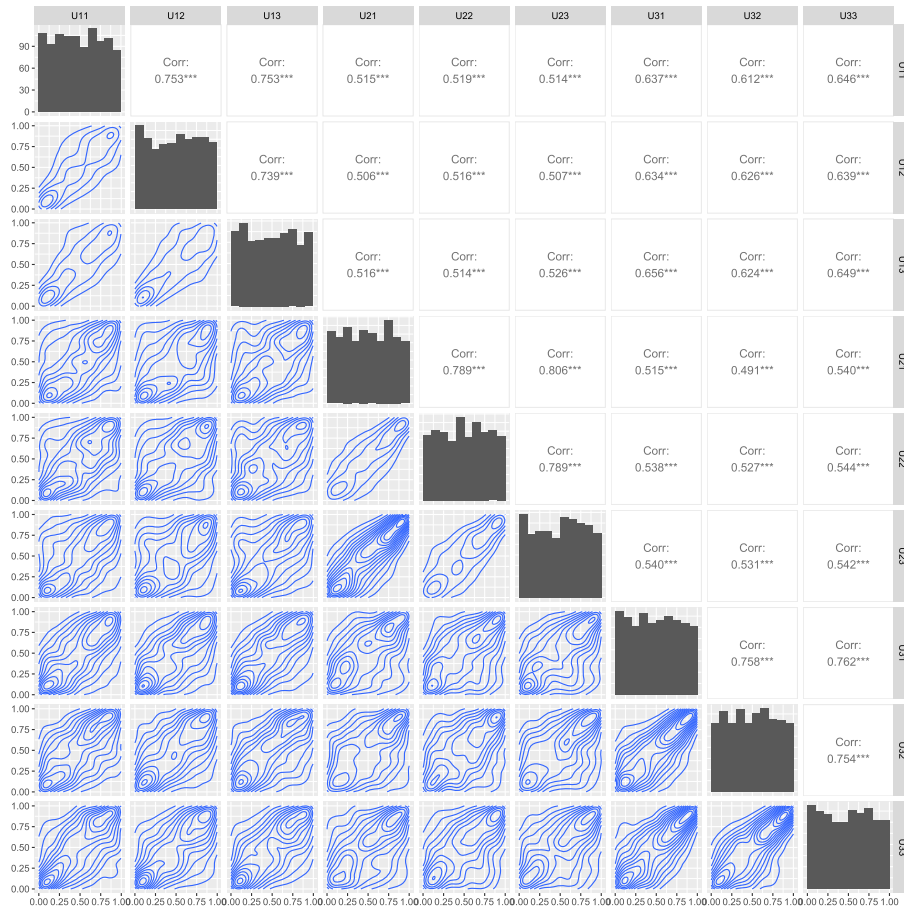


FIG D.2. (Model B) Pair plots of a sample of size $n = 1000$ from the copula $C_{\mathcal{G}, \psi, \ell, \mathcal{Q}}$ where $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\} = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$. $\psi = \{\psi_{\theta_1}, \psi_{\theta_2}, \psi_{\theta_3}\}$ with $(\theta_1, \theta_2, \theta_3) = (1.5, 1.5, 2)$ where ψ_{θ_1} is Clayton while $\psi_{\theta_2}, \psi_{\theta_3}$ are Joe. $\ell = \{\ell_{\vartheta_1}, \ell_{\vartheta_2}, \ell_{\vartheta_3}\}$ with $(\vartheta_1, \vartheta_2, \vartheta_3) = (1.25, 2, 1.5)$ where $\ell_{\vartheta_1}, \ell_{\vartheta_2}, \ell_{\vartheta_3}$ are Gumbel-Hougaard. The radial survival copula \bar{D} is trivariate Gumbel-Hougaard with parameter $\vartheta_R = 4$. Upper: linear correlation, Lower: contour density, Diagonal: univariate histogram.

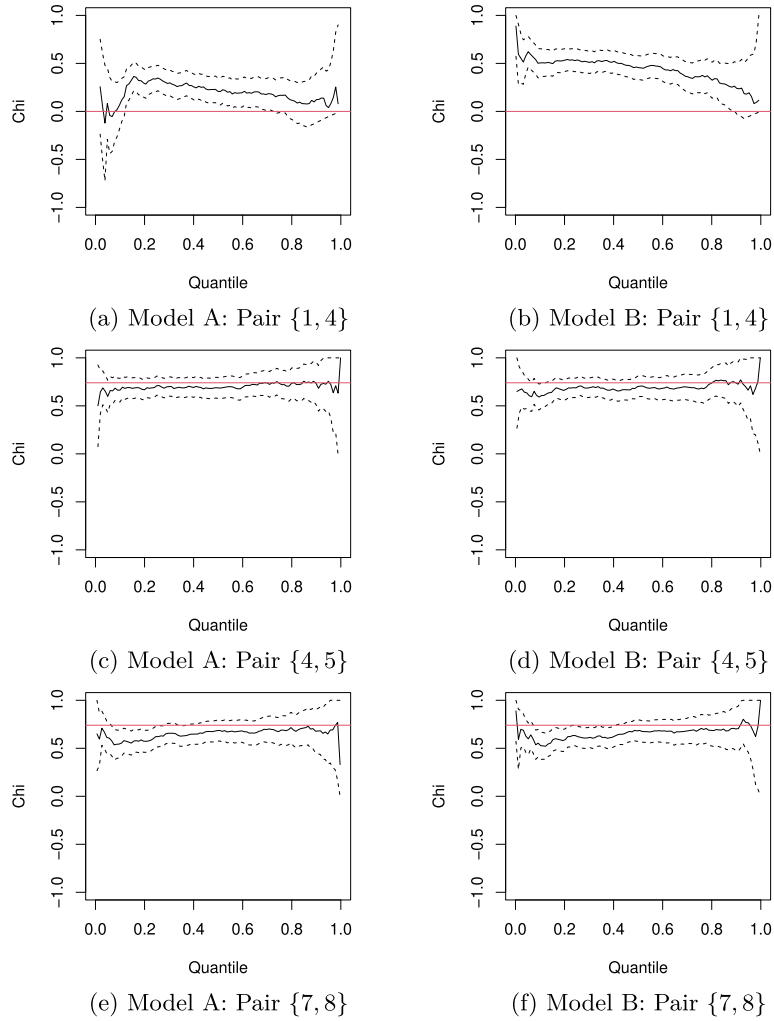


FIG D.3. Pair chi plots for the sample from Model A (left) and Model B (right). The full black line is the empirical estimate of (5.1), the dotted lines are 95% confidence intervals and the red lines represent the true values of $\lim_{q \rightarrow 1} \chi_{ij}(q)$ for each pair (i, j) . The samples used for the empirical estimates are those of Figures D.1 and D.2.

E. Additional figures for the simulation study of Section 6

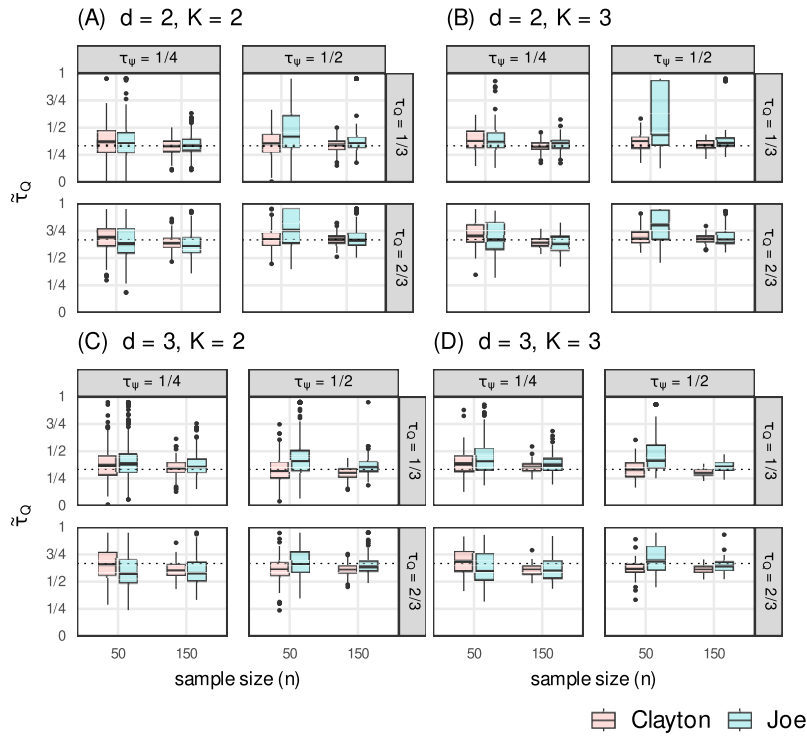


FIG E.1. Boxplots of the fitted values of Kendall's Tau for the radial copula \bar{Q} in the Gumbel case. True values τ_Q are plotted in dotted lines. The simulation is based on 100 Monte-Carlo replicates.

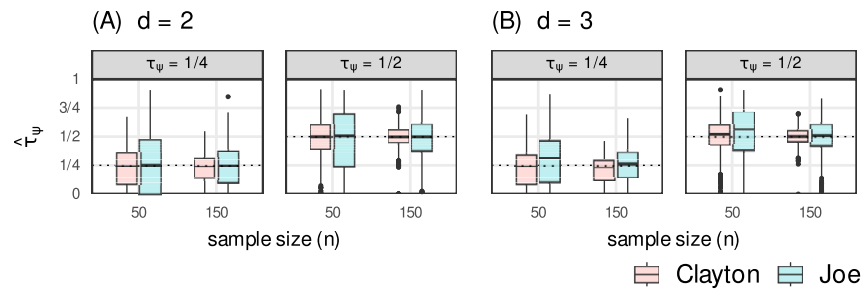


FIG E.2. Boxplots of the fitted values of Kendall's Tau for the radial components. True values τ_ψ are plotted in dotted lines.

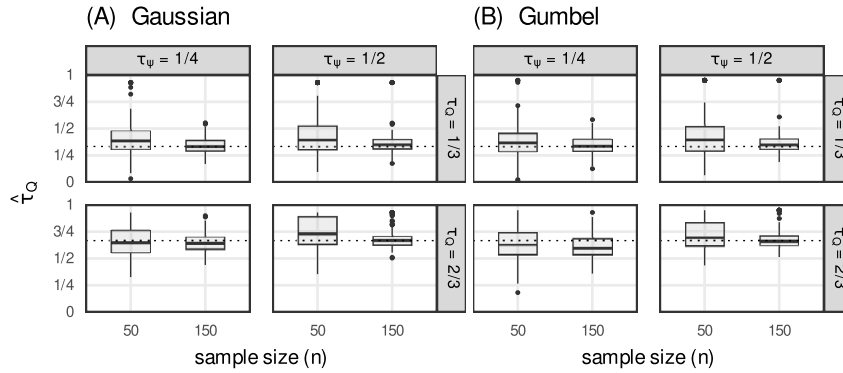


FIG E.3. Boxplots of the fitted values of Kendall's Tau for the radial copula \bar{Q} in the cases where clustered Archimax copula contained both Clayton and Joe clusters. True values τ_Q are plotted in dotted lines. The simulation is based on 100 Monte-Carlo replicates.

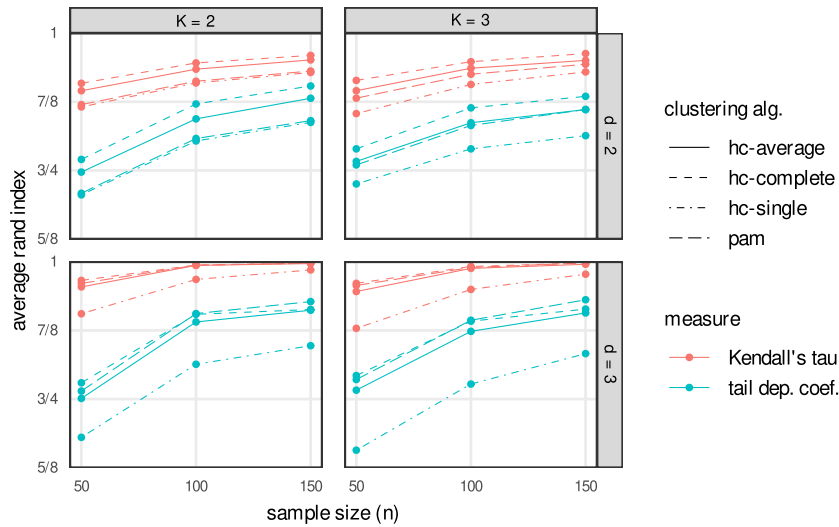


FIG E.4. Average Rand indices obtained for each cluster size and number configuration according to sample size. Four clustering algorithms were considered with dissimilarity matrices built from estimates of either Kendall's tau or the upper tail dependence coefficient. Notation: *hc* denotes hierarchical clustering with average, complete and single linkage while *pam* denotes the partition around medoids method.

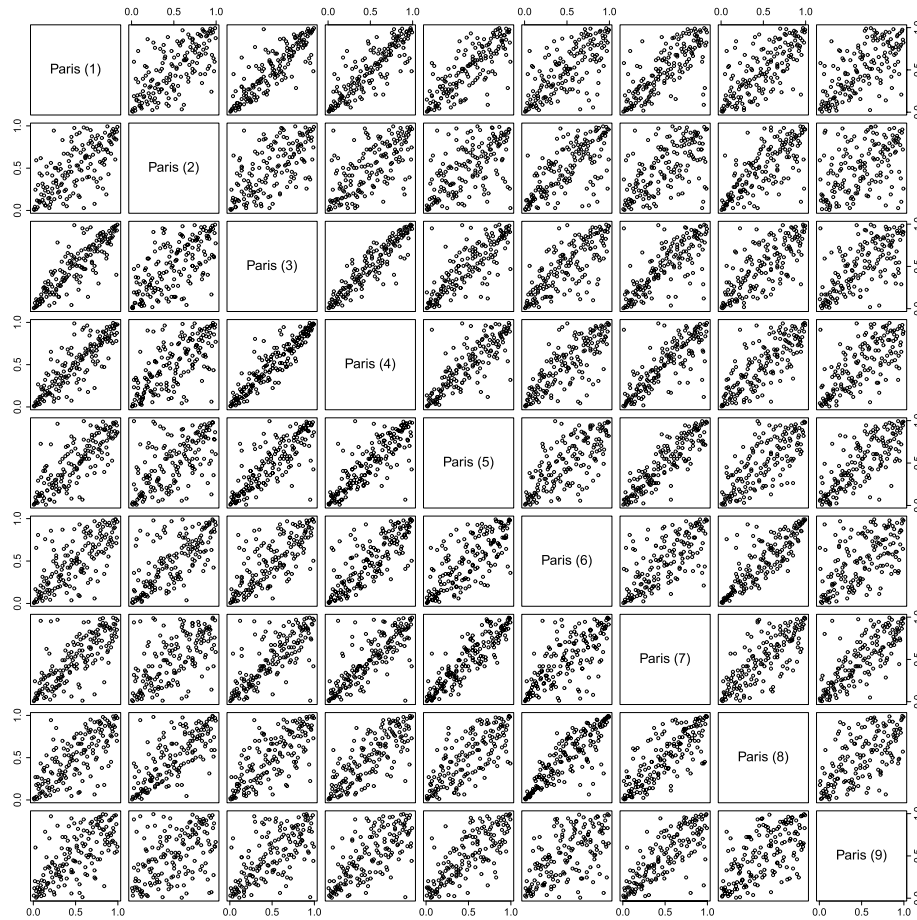
F. Additional figures for the data illustration of Section 7

FIG F.1. Pair plots of the scaled componentwise ranks of monthly maxima of precipitations for the nine stations of the Paris cluster considered in the application of Section 7.

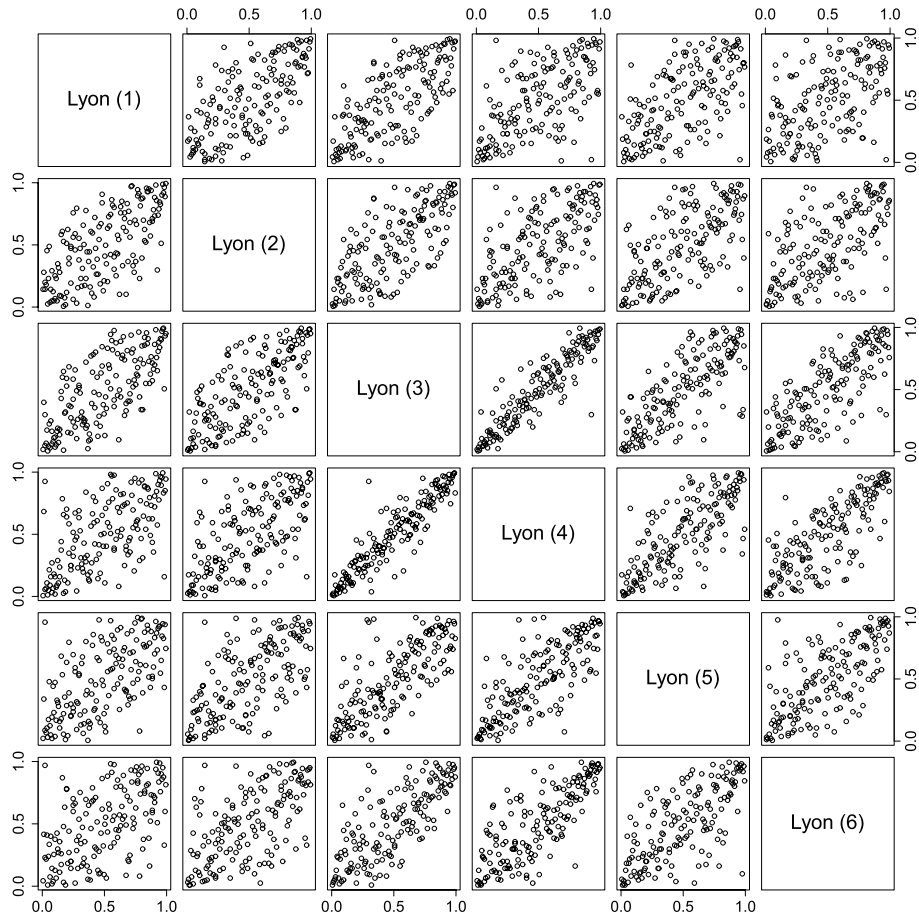


FIG F.2. Pair plots of the scaled componentwise ranks of monthly maxima of precipitations for the six stations of the Lyon cluster considered in the application of Section 7.

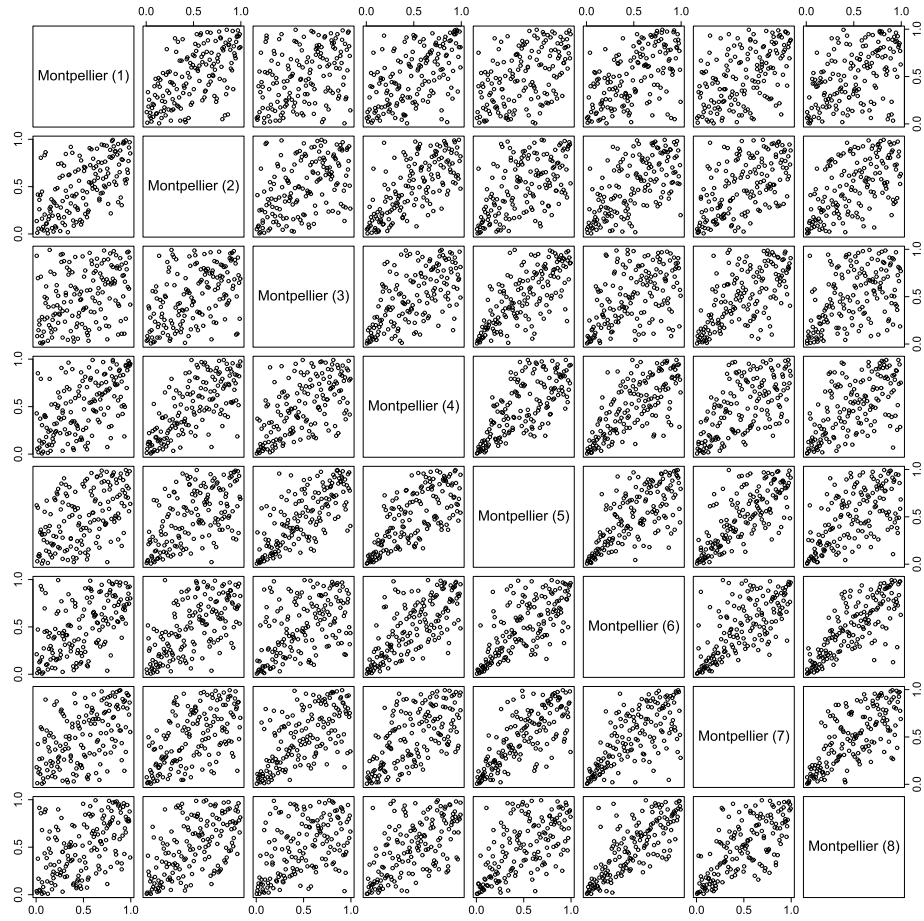


FIG F.3. Pair plots of the scaled componentwise ranks of monthly maxima of precipitations for the eight original stations of the Montpellier cluster considered in the application of Section 7.

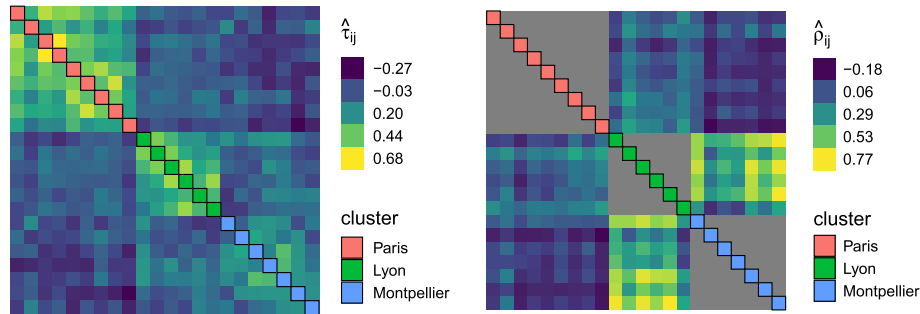


FIG F.4. Estimated quantities from Section 7.3 ($d = 22$ stations). Left: Matrix of pairwise Kendall correlations associated with the (yearly) data used to perform the test of [Kojadinovic and Holmes \(2009\)](#). Right: Matrix of pairwise Pearson correlations involved in the simpler estimation procedure for Q described at the end of Section 6.2.

Acknowledgments

The authors would like thank both referees and the associate editor for their valuable comments on an earlier version of the manuscript, as well as Prof. Patrick Brown for his generosity in sharing computational resources. Thanks are also due to Météo France for providing the data, and in particular Maxime Taillardat for numerous fruitful discussions.

Supplementary Material

Supplement to “Clustered Archimax copulas”

(doi: [10.1214/24-EJS2340SUPP](https://doi.org/10.1214/24-EJS2340SUPP); .zip). Folder containing the functions used in the application and simulation study.

References

- ARVESEN, J. N. (1969). Jackknifing U-Statistics. *Ann. Math. Statist.* **40** 2076–2100. [MR0264805](#)
- AULBACH, S., FALK, M. and ZOTT, M. (2015). The space of D -norms revisited. *Extremes* **18** 85–97. <https://doi.org/10.1007/s10687-014-0204-y>. [MR3317854](#)
- BELZILE, L. R. and NEŠLEHOVÁ, J. G. (2017). Extremal attractors of Liouville copulas. *J. Multivariate Anal.* **160** 68–92. <https://doi.org/10.1016/j.jmva.2017.05.008>. [MR3688690](#)
- BERNARD, E., NAVEAU, P., VRAC, M. and MESTRE, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *J. Clim.* **26** 7929–7937.
- BREIMAN, L. (1965). On some limit theorems similar to the arc-sin law. *Teor. Veroyatnost. i Primenen.* **10** 351–360. [MR0184274](#)

- CAPÉRAÀ, P., FOUGÈRES, A. L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. <https://doi.org/10.1093/biomet/84.3.567>. MR1603985
- CAPÉRAÀ, P., FOUGÈRES, A.-L. and GENEST, C. (2000). Bivariate distributions with given extreme value attractor. *J. Multivariate Anal.* **72** 30–49. MR1747422
- CHARPENTIER, A., FOUGÈRES, A.-L., GENEST, C. and NEŠLEHOVÁ, J. G. (2014). Multivariate Archimax copulas. *J. Multivariate Anal.* **126** 118–136. MR3173086
- CHATELAIN, S., FOUGÈRES, A.-L. and NEŠLEHOVÁ, J. G. (2020). Inference for Archimax copulas. *Ann. Statist.* **48** 1025–1051. <https://doi.org/10.1214/19-AOS1836>. MR4102686
- CHATELAIN, S., PERREAULT, S., FOUGÈRES, A.-L. and NEŠLEHOVÁ, J. G. (2025). Supplement to “Clustered Archimax copulas”. <https://doi.org/10.1214/24-EJS2340SUPP>.
- COLES, S., HEFFERNAN, J. and TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.
- COX, D. R. and REID, N. (2004). A Note on Pseudolikelihood Constructed from Marginal Densities. *Biometrika* **91** 729–737. MR2090633
- DAVISON, A. C., HUSER, R. and THIBAUD, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences* **45** 511–529. MR3079649
- DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **40** 317–337. <https://doi.org/10.1007/BF00533086>. MR0478290
- DURANTE, F. and SEMPI, C. (2010). *Copula theory: an introduction*. Springer. MR3051261
- EMBRECHTS, P. and GOLDIE, C. M. (1980). On closure and factorization properties of subexponential and related distributions. *J. Austral. Math. Soc. Ser. A* **29** 243–256. MR0566289
- EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, New York. MR1458613
- ENGELKE, S., OPITZ, T. and WADSWORTH, J. (2019). Extremal dependence of random scale constructions. *Extremes* **22** 623–666. MR4031852
- FANG, K., KOTZ, S. and NG, K. (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall/CRC. MR1071174
- FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737. <https://doi.org/10.3150/13-BEJ538>. MR3263087
- FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Math. Proc. Cambridge Philos. Soc.* **24** 180–190. Cambridge University Press.
- GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Ann. Math.* **44** 423–453. MR0008655
- GÓRECKI, J. and HOFERT, M. (2023). Composite pseudo-likelihood estimation for pair-tractable copulas such as Archimedean, Archimax and re-

- lated hierarchical extensions. *Journal of Statistical Computation and Simulation* **93** 2321–2355. <https://doi.org/10.1080/00949655.2023.2180511>. MR4617813
- HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325. MR0026294
- HOFERT, M., HUSER, R. and PRASAD, A. (2018). Hierarchical Archimax copulas. *J. Multivariate Anal.* **167** 195–211. <https://doi.org/10.1016/j.jmva.2018.05.001>. MR3830642
- HUANG, X. (1992). Statistics of bivariate extreme values, PhD thesis, Erasmus universiteit, Rotterdam.
- HUSER, R., OPITZ, T. and THIBAUD, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spat. Stat.* **21** 166–186. MR3692183
- HUSER, R. and WADSWORTH, J. L. (2019). Modeling Spatial Processes with Unknown Extremal Dependence Class. *J. Amer. Statist. Assoc.* **114** 434–444. <https://doi.org/10.1080/01621459.2017.1411813>. MR3941266
- JOE, H. (2015). *Dependence Modeling With Copulas*. CRC Press. MR3328438
- KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer. <https://doi.org/10.1007/978-1-4757-4015-8>. MR1876169
- KOJADINOVIC, I. and HOLMES, M. (2009). Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process. *J. Multivariate Anal.* **100** 1137–1154. <https://doi.org/10.1016/j.jmva.2008.10.013>. MR2508377
- KOJADINOVIC, I., SEGERS, J. and YAN, J. (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canad. J. Statist.* **39** 703–720. MR2860835
- LARSSON, M. and NEŠLEHOVÁ, J. (2011). Extremal behavior of Archimedean copulas. *Adv. in Appl. Probab.* **43** 195–216. MR2761154
- LJUNG, G. M. and BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65** 297–303. <https://doi.org/10.1093/biomet/65.2.297>.
- MAI, J.-F. (2022). About the exact simulation of bivariate (reciprocal) Archimax copulas. *Depend. Model.* **10** 29–47. MR4414493
- MCNEIL, A. J. and NEŠLEHOVÁ, J. (2009). Multivariate Archimedean copulas, d -monotone functions and ℓ_1 -norm symmetric distributions. *Ann. Statist.* **37** 3059–3097. MR2541455
- MCNEIL, A. J. and NEŠLEHOVÁ, J. (2010). From Archimedean to Liouville copulas. *J. Multivariate Anal.* **101** 1772–1790. <https://doi.org/10.1016/j.jmva.2010.03.015>. MR2651954
- MESJAR, R. and JÁGR, V. (2013). d -dimensional dependence functions and Archimax copulas. *Fuzzy Sets Syst.* **228** 78–87. MR3092417
- NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer. MR2197664
- NG, Y., HASAN, A. and TAROKH, V. (2022). Inference and Sampling for Archimax Copulas. *arXiv preprint arXiv:2205.14025*.
- PERREAULT, S., NEŠLEHOVÁ, J. G. and DUCHESNE, T. (2022). Hypothesis

- Tests for Structured Rank Correlation Matrices. *J. Amer. Statist. Assoc.* **0** 1–12. <https://doi.org/10.1080/01621459.2022.2096619>. MR4681628
- PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667
- PICKANDS, J. III (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)* **49** 859–878, 894–902. MR0820979
- RESSEL, P. (2013). Homogeneous distributions—and a spectral representation of classical mean values and stable tail dependence functions. *J. Multivariate Anal.* **117** 246–256. MR3053546
- SAUNDERS, K., STEPHENSON, A. and KAROLY, D. (2021). A regionalisation approach for rainfall based on extremal dependence. *Extremes* **24** 215–240. MR4246276
- SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229–231. MR0125600
- VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Stat. Sin.* **21** 5–42. MR2796852
- VATAN, P. (1985). Max-infinite divisibility and max-stability in infinite dimensions. In *Probability in Banach spaces, V (Medford, Mass., 1984)*, **1153** 400–425. Springer. <https://doi.org/10.1007/BFb0074963>. MR0821994
- WADSWORTH, J. L., TAWN, J. A., DAVISON, A. C. and ELTON, D. M. (2017). Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 149–175. MR3597968