



**HAL**  
open science

# Estimation of the Spectral Measure from Convex Combinations of Regularly Varying Random Vectors

Marco Oesting, Olivier Wintenberger

► **To cite this version:**

Marco Oesting, Olivier Wintenberger. Estimation of the Spectral Measure from Convex Combinations of Regularly Varying Random Vectors. *Annals of Statistics*, In press. hal-02958799v3

**HAL Id: hal-02958799**

**<https://hal.science/hal-02958799v3>**

Submitted on 29 May 2024

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

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

---

# ESTIMATION OF THE SPECTRAL MEASURE FROM CONVEX COMBINATIONS OF REGULARLY VARYING RANDOM VECTORS

---

A PREPRINT

**Marco Oesting**

Stuttgart Center for Simulation Science and Institute of Stochastics and Applications,  
University of Stuttgart, 70569 Stuttgart, Germany,  
marco.oesting@mathematik.uni-stuttgart.de

**Olivier Wintenberger**

LPSM, Sorbonne Université,  
4 place Jussieu, 75005 Paris, France  
Wolfgang Pauli Institut, c/o Fakultät für Mathematik,  
Universität Wien, 1090 Vienna, Austria  
olivier.wintenberger@sorbonne-universite.fr

May 29, 2024

## ABSTRACT

The extremal dependence structure of a regularly varying random vector  $\mathbf{X}$  is fully described by its limiting spectral measure. In this paper, we investigate how to recover characteristics of the measure, such as extremal coefficients, from the extremal behaviour of convex combinations of components of  $\mathbf{X}$ . Our considerations result in a class of new estimators of moments of the corresponding combinations for the spectral vector. We show asymptotic normality by means of a functional limit theorem and, focusing on the estimation of extremal coefficients, we verify that the minimal asymptotic variance can be achieved by a plug-in estimator using subsampling bootstrap. We illustrate the benefits of our approach on simulated and real data.

**Keywords** Asymptotic statistics; extremal coefficient; moment estimator; multivariate extremes.

## 1 Introduction

One of the main challenges in multivariate extreme value analysis is the inference of the dependence among the extremes of a random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  from a sample of independent copies  $\mathbf{X}_1, \dots, \mathbf{X}_n$ . In risk management applications, particular interest lies in the case that  $\mathbf{X}$  possesses asymptotically dependent components, i.e. extremes tend to appear jointly in various components, with heavy tails. Many of these cases are covered by the framework of multivariate regular variation. We fix an orthant where the extremes occur, and restrict ourselves to the first one with no loss of generality; thus, we consider  $\mathbf{X} = (X_1, \dots, X_d)^\top \in [0, \infty)^d$ .

It is common practice in extreme value analysis to consider the extremal dependence structure of  $\mathbf{X}$  separately from the marginal properties of the components  $X_1, \dots, X_d$ . Thus, focus is often put on the joint tail behaviour of some standardized vector  $\mathbf{X}^*$  obtained from  $\mathbf{X}$  via marginal transformations. More precisely, the margins are standardized such that the tails are regularly varying with index 1 and asymptotically identical in the sense that

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(X_i^* > x)}{\mathbb{P}(X_j^* > x)} = 1 \quad \text{for all } i, j \in \{1, \dots, d\}.$$

By such a transformation, the property of multivariate regular variation is inherited from  $\mathbf{X}$  to  $\mathbf{X}^*$ . Thus, extremal dependence of  $\mathbf{X}^*$  can be fully described via its so-called spectral measure, that is, the limit distribution of the angular

component  $\mathbf{X}^*/\|\mathbf{X}^*\|_\infty$  conditional on  $\|\mathbf{X}^*\|_\infty > u$  as  $u \rightarrow \infty$ . We denote the vector distributed as the spectral measure by  $\Theta$ .

In the past decades, the estimation of the spectral measure and related concepts has attracted a lot of attention, starting from the bivariate setting ( $d = 2$ ), where estimation of the spectral measure is equivalent to inference for the so-called Pickands' dependence function. Here, seminal estimation procedures come from Pickands [1975] with refinements such as Capéraà et al. [1997] using a copula approach assuming known margins. These procedures have been extended by Genest and Segers [2009] to the setting with unknown margins of  $\mathbf{X}$  using the rank transform. All the above mentioned works were assuming that the observations are max-stable, i.e. that the asymptotic extreme regime is achieved, a typical assumption for observations that correspond to maxima over certain temporal blocks such as annual maxima. The pioneer works of Einmahl et al. [1993] and Einmahl et al. [1997] deal with the estimation of the cumulative distribution function of the spectral measure with observations in the domain of attraction of max-stable distribution, including the case of regularly varying distributions. The two-step procedure in Einmahl et al. [1997] for dealing with unknown margins is replaced by a more efficient rank transform in Einmahl et al. [2001]. An alternative approach due to Drees and Huang [1998] focuses on the estimation of the stable tail dependence function

$$L(\mathbf{x}) = \mathbb{E} \left[ \max_{i=1, \dots, d} x_i \Theta_i \right], \quad \mathbf{x} = (x_1, \dots, x_d) \in (0, \infty)^d,$$

using the rank transform for  $d = 2$ . This work was extended to the multivariate case  $d > 2$  in Einmahl et al. [2012]. We refer to Chapter 7 of the monograph De Haan and Ferreira [2006] for a general overview of non-parametric approaches.

When  $d$  is large, any inference of the complete spectral distribution suffers from the curse of dimensionality. Thus, it is more reasonable in practice to focus on the joint tail behaviour of the components of lower-dimensional subvectors  $\mathbf{X}_I^* = (X_i^*)_{i \in I}$  for sufficiently many ‘‘directions’’  $I$  that are non-empty subsets of  $\{1, \dots, d\}$ . In this framework, one often makes use of summary statistics such as the so-called extremal coefficients [cf. Smith, 1990, Schlather and Tawn, 2003] given by

$$\tau_I = \lim_{u \rightarrow \infty} \frac{\mathbb{P}(\max_{i \in I} X_i^* > u)}{\mathbb{P}(X_j^* > u)}, \quad j \in I, \quad (1)$$

a number between 1 and  $|I|$ . Such parameter has attracted a lot of attention as it quantifies the effective number of asymptotically independent components among  $I$  [cf. Schlather and Tawn, 2003]. Some properties, such as self-consistency for sets of extremal coefficients, and various procedures for estimating extremal coefficients were introduced by Schlather and Tawn [2003]. For max-stable random fields, Cooley et al. [2006] showed that inference can also be based on the madogram, a graphical tool originally developed in geostatistics.

In this paper, we also consider the extremal dependence structure of subvectors  $\mathbf{X}_I^*$  described in terms of the corresponding spectral vector  $\Theta^I$ , that is, the limit in distribution of  $\mathbf{X}_I^*/\|\mathbf{X}_I^*\|_\infty$  conditional on  $\ell_I(\mathbf{X}^*) := \|\mathbf{X}_I^*\|_\infty > u$  as  $u \rightarrow \infty$ . While existing works typically focus on a direct estimation of the distribution of  $\Theta^I$ , i.e. the spectral measure, we focus on moment estimators of  $\Theta^I$ . Our approach further differs from previous works since we impose the assumption that the original observations  $\mathbf{X}$  is such that  $\mathbf{X}^* = (r_1^{-1}X_1^{\alpha_1}, \dots, r_d^{-1}X_d^{\alpha_d})$  is standard regularly varying with index 1 and with tail equivalent margins. This restriction allows us to consider a two-step procedure in the spirit of Einmahl et al. [1997] with a simpler first standardization step relying on the componentwise normalization via upper order statistics and the estimation of the coefficients  $\alpha_1, \dots, \alpha_d > 0$  via Hill type estimators. Our second step is then based on the estimation of moments of convex combinations of the spectral vector, i.e. estimators of  $\mathbb{E}[(\mathbf{v}^\top \Theta^I)^p]$  with  $p \geq 1$  and  $\mathbf{v}$  on the simplex of the direction  $I$ . We prove that the collection of such moments  $\mathbb{E}[(\mathbf{v}^\top \Theta^I)^p]$ ,  $p \geq 1$ , characterizes the spectral measure of  $\mathbf{X}_I^*$ . Perceiving the resulting estimators as functions of the vector  $\mathbf{v}$  on the simplex, we obtain functional central limit theorems with precise asymptotic (co-)variances. Here, we consider both the case that the margins of  $\mathbf{X}$  are known, i.e. we can directly work with exactly standardized observations  $\mathbf{X}^*$ , and the case that the margins are unknown and a preliminary empirical standardization step has to be included.

In accordance to the above-mentioned paradigm that, in large dimensions  $d$ , emphasis is often put on simpler summary statistics, we suggest to mainly focus on small  $p$  for which our approach provides simple functions that contain important information on the dependence among extremes. For instance, by tail equivalence of the marginals of  $\mathbf{X}^*$  we have that  $\mathbb{E}[\Theta_i^I] = \mathbb{E}[\Theta_j^I] = 1/\tau_I$  for any  $i, j \in I$  and, consequently, for any positive weights  $\mathbf{v}$  such that  $\sum_{i \in I} v_i = 1$  we have that  $\mathbb{E}[\mathbf{v}^\top \Theta^I] = 1/\tau_I$ . Thus, the moment of order  $p = 1$  of convex combinations of  $\Theta^I$  does not depend on the weight vector  $\mathbf{v}$ . This observation motivates an additional third step of our procedure. More precisely, similarly to Mainik and Rüschemdorf [2010] who estimate the optimal portfolio of jointly regularly varying assets, we use the functional central limit theorem to estimate the weight vector  $\mathbf{v}^*$  that minimizes the asymptotic variance and show that this minimal asymptotic variance can be achieved by a plug-in estimator. We note that the determination of  $\mathbf{v}^*$  is interesting on its own and provides important information on the dependence among extremes. The determination of a prominent direction in the data is related to recent developments on extreme PCA in Cooley and Thibaud [2019] and

Drees and Sabourin [2019]. Moreover we discuss how the weights  $v^*$  are impacted by the fact that the margins are known or not.

The paper is organized as follows. In Section 2, we gather several results on convex combinations of the spectral components of a regularly varying vector. The asymptotic normality of empirical moment estimators is given in Section 3 both for the case of known and the case of unknown margins. These results specified to first order moments are then applied in Section 4 in order to obtain estimators for extremal coefficients with minimal asymptotic variance. Some comparison to benchmark estimators based on the definition of  $\tau_I$  and some illustrations on simulated and real data are provided in Section 5. We conclude with a short discussion on our results in Section 6.

## 2 Extremes of convex combinations of regularly varying random vectors

Let  $\mathbf{X} = (X_1, \dots, X_d)^\top$  be a non-standard regularly varying  $E = [0, \infty)^d$ -valued random vector, with different tail indices  $\alpha_1, \dots, \alpha_d > 0$ . We assume the existence of scaling factors  $r_1, \dots, r_d > 0$  such the transformed vector  $\mathbf{X}^* = (r_1^{-1}X_1^{\alpha_1}, \dots, r_d^{-1}X_d^{\alpha_d})$  is regularly varying with index 1 and satisfies

$$a^*(u)\mathbb{P}(u^{-1}\mathbf{X}^* \in \cdot) \xrightarrow{v} \mu^*(\cdot), \quad u \rightarrow \infty, \quad (2)$$

for some normalized exponent measure  $\mu^*$  which is homogeneous of order  $-1$  and

$$\mu^*(\{\mathbf{x} \in E : x_i > 1\}) = \lim_{u \rightarrow \infty} a^*(u)\mathbb{P}(X_i^* > u) = 1, \quad (3)$$

i.e.  $a^*(u) \sim \mathbb{P}(X_i^* > u)^{-1}$  for all  $i = 1, \dots, d$ . This verifies that the tail behaviour of all the components of  $\mathbf{X}^*$  is asymptotically identical, i.e. its marginal distribution functions  $F_1^*, \dots, F_d^*$  satisfy

$$\frac{1 - F_i^*(u)}{1 - F_j^*(u)} \rightarrow 1, \quad u \rightarrow \infty,$$

for all  $i, j \in \{1, \dots, d\}$ , which also implies

$$\frac{(F_i^*)^{-1}(p)}{(F_j^*)^{-1}(p)} \rightarrow 1, \quad p \rightarrow 1. \quad (4)$$

Eq. (2) can be equivalently expressed in terms of a spectral component for an arbitrary norm  $\|\cdot\|$  on  $\mathbb{R}^d$ . More precisely, we obtain the weak convergence

$$\mathcal{L}(u^{-1}\mathbf{X}^* \mid \|\mathbf{X}^*\| > u) \xrightarrow{w} \mathcal{L}(Y\Theta), \quad u \rightarrow \infty, \quad (5)$$

where  $Y$  is a unit Pareto random variable and, independently from  $Y$ ,  $\Theta$  is a  $[0, \infty)^d$ -valued random vector satisfying  $\|\Theta\| = 1$  almost surely, i.e.

$$\mathbb{P}(\Theta \in S_{d-1}^+) = 1, \quad \text{where } S_{d-1}^+ = \{\mathbf{x} \in [0, \infty)^d : \|\mathbf{x}\| = 1\}.$$

In the following, we will focus on the maximum norm  $\|\cdot\| = \|\cdot\|_\infty$  which is not a serious restriction as all norms on  $\mathbb{R}^d$  are equivalent.

There is a one-to-one correspondence between the normalized exponent measure  $\mu^*$  and the distribution of  $\Theta$ , the so-called (normalized) *spectral measure* (w.r.t. the norm  $\|\cdot\|$ ) via

$$\mu^*(\{\mathbf{x} \in E : \|\mathbf{x}\| > r, \mathbf{x}/\|\mathbf{x}\| \in B\}) = \tau r^{-1}\mathbb{P}(\Theta \in B), \quad r > 0, B \subset S_{d-1}^+ \quad (6)$$

for some constant  $1 \leq \tau \leq d$  given by

$$\tau = \mu^*(\{\mathbf{x} \in E : \|\mathbf{x}\| > 1\}) = \lim_{u \rightarrow \infty} a^*(u)\mathbb{P}(\|\mathbf{X}^*\| > u). \quad (7)$$

The extremal dependence structure of the components of the normalized random vector  $\mathbf{X}^*$ , and, thus, also of the original vector  $\mathbf{X}$ , is fully described by the distribution of the spectral vector  $\Theta$  which we will estimate in the following.

More precisely, we now describe the tail behaviour of convex combinations of  $X_1^*, \dots, X_d^*$  in terms of the corresponding spectral measure. To this end, we will make use of the following well-known lemma which is a direct consequence of the weak convergence in Eq. (5).

**Lemma 1.** Let  $\mathbf{X}^*$ ,  $Y$  and  $\Theta$  be as above. Furthermore, let  $g : (0, \infty) \times S_{d-1}^+ \rightarrow \mathbb{R}$  be a continuous bounded function. Then, we have

$$\lim_{u \rightarrow \infty} \mathbb{E} [g(u^{-1} \|\mathbf{X}^*\|, \mathbf{X}^*/\|\mathbf{X}^*\|) \mid \|\mathbf{X}^*\| > u] = \mathbb{E}[g(Y, \Theta)].$$

In particular, the lemma implies that

$$\lim_{u \rightarrow \infty} \mathbb{E}[\tilde{g}(\mathbf{X}^*/\|\mathbf{X}^*\|) \mid \|\mathbf{X}^*\| > u] = \mathbb{E}[\tilde{g}(\Theta)] \quad (8)$$

for every continuous function  $\tilde{g} : S_{d-1}^+ \rightarrow \mathbb{R}$ .

Furthermore, the lemma extends to indicator functions of sets that are continuous with respect to the measure of  $(Y, \Theta)$ , i.e. conditional probabilities of events that can be expressed in terms of  $u^{-1}\|\mathbf{X}^*\|$  and  $\mathbf{X}^*/\|\mathbf{X}^*\|$ . For instance, for convex combinations of  $X_1^*, \dots, X_d^*$  with coefficient vector  $\mathbf{v} \in B_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_1 \leq 1\}$ ,  $\mathbf{v} \neq \mathbf{0}$ , we obtain

$$\lim_{u \rightarrow \infty} \mathbb{P}(\mathbf{v}^\top \mathbf{X}^* > u \mid \|\mathbf{X}^*\| > u) = \mathbb{P}(Y \mathbf{v}^\top \Theta > 1) = \mathbb{E}[(\mathbf{v}^\top \Theta)_+ \wedge 1] = \mathbb{E}[(\mathbf{v}^\top \Theta)_+] \quad (9)$$

using that  $\mathbf{v}^\top \Theta \leq \|\mathbf{v}\|_1 \cdot \|\Theta\| = \|\mathbf{v}\|_1 \leq 1$  a.s. It can be shown that the law of  $\Theta$  is uniquely determined by the set of all the limits in (9) with  $\mathbf{v} \in \partial B_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_1 = 1\}$ , see Basrak et al. [2002], Klüppelberg and Pergamenchtchikov [2007], Boman and Lindskog [2009]. This fact motivates the use of extremes of convex combinations of observations for inference of the spectral measure. In this paper, we modify the above approach by assessing the limiting quantity on the RHS of Eq. (9) in an alternative way based on the following result which is a trivial consequence of Lemma 1.

**Lemma 2.** Let  $\mathbf{X}^*$  be a regularly varying random vector with standard Pareto margins. Then, for every  $\mathbf{v} \in \mathbb{R}^d$  and  $p \in \mathbb{N}$ , we have

$$\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{X}^*}{\|\mathbf{X}^*\|} \right)_+^p \mid \|\mathbf{X}^*\| > u \right] \rightarrow \mathbb{E}[(\mathbf{v}^\top \Theta)_+^p] \quad (u \rightarrow \infty). \quad (10)$$

*Proof.* The result directly follows by applying Eq. (8) to the continuous functions  $\tilde{g}_{\mathbf{v},p}(\theta) = (\mathbf{v}^\top \theta)_+^p$  with  $\mathbf{v} \in \mathbb{R}^d$  and  $p \in \mathbb{N}$ .  $\square$

Note that, for  $\mathbf{v} \not\geq \mathbf{0}$ , the term  $\mathbf{v}^\top \Theta$  is negative with positive probability, which hampers the calculation of closed-form expressions for the RHS of (10). In order to avoid these difficulties, we restrict our attention to non-negative vectors  $\mathbf{v} \in \partial B_1^+(\mathbf{0}) = \partial B_1(\mathbf{0}) \cap [0, \infty)^d$ , i.e. we consider only convex combinations of components of  $\mathbf{X}^*$  with positive coefficients, while allowing for arbitrary powers of these combinations. This ensures the identifiability of the spectral measure from the resulting limits as the following proposition shows.

**Proposition 3.** Let  $\mathbf{X}^*$  be a non-negative regularly varying random vector with index 1. Then, the spectral measure of  $\Theta$  is uniquely determined by the moments

$$\{\mathbb{E}[(\mathbf{v}^\top \Theta)_+^p] : \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in \mathbb{N}\} = \{\mathbb{E}[(\mathbf{v}^\top \Theta)^p] : \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in \mathbb{N}\}. \quad (11)$$

*Proof.* By a series expansion, for all  $\mathbf{s} \in [0, \infty)^d$ , the limits in (11) uniquely determine

$$\mathbb{E}[\exp(-\mathbf{s}^\top \Theta)] = \sum_{p=0}^{\infty} (-1)^p \cdot \frac{\|\mathbf{s}\|_1^p}{p!} \cdot \mathbb{E} \left[ \left( \frac{\mathbf{s}^\top}{\|\mathbf{s}\|_1} \Theta \right)^p \right], \quad \mathbf{s} \in [0, \infty),$$

i.e. the Laplace transform of the spectral measure.  $\square$

Let us list some properties of the function  $\mathbf{v} \mapsto m_p(\mathbf{v}) = \mathbb{E}[(\mathbf{v}^\top \Theta)^p]$  for  $p = 1, 2, \dots$  and  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ . We identify  $m_1(\mathbf{v}) = \sum_{i=1}^d v_i \mathbb{E}[\Theta_i]$ , i.e. the function is linear for  $p = 1$ . As, for all  $i = 1, \dots, d$ ,

$$\begin{aligned} \mathbb{E}[\Theta_i] &= \int_0^1 \mathbb{P}(\Theta_i > x) dx = \int_1^\infty y^{-2} \mathbb{P}(\Theta_i > 1/y) dy \\ &= \mathbb{P}(Y \Theta_i > 1) = \lim_{u \rightarrow \infty} \mathbb{P}(X_i^* > u \mid \|\mathbf{X}^*\| > u) = \lim_{u \rightarrow \infty} \frac{1}{a^*(u) \mathbb{P}(\|\mathbf{X}^*\| > u)} = \frac{1}{\tau}, \end{aligned} \quad (12)$$

where  $\tau$  is defined as in Eq. (7), we even obtain  $m_1(\mathbf{v}) = \frac{1}{\tau} \sum_{i=1}^d v_i$ , which implies that  $m_1$  is constant on  $\partial B_1^+(\mathbf{0})$ . For  $p \geq 2$ , the function  $m_p$  is strongly convex. Thus, one easily deduces from Jensen's inequality that  $m_p(\mathbf{v}) \leq \max_{1 \leq i \leq d} \mathbb{E}[\Theta_i^p]$ .

For many popular limit models, closed-form expressions for  $\mathbb{E}[(\mathbf{v}^\top \Theta)^p]$  are difficult to obtain as the spectral measure w.r.t. the maximum norm  $\|\cdot\| = \|\cdot\|_\infty$  is often difficult to handle in high dimensional setting. Instead, spectral vectors are often normalized w.r.t. some less complex functional  $\ell_I(\mathbf{x}) = \bigvee_{i \in I} x_i$  over some smaller subset  $I \subset \{1, \dots, d\}$ . Such functionals  $\ell_I$  play an important role when estimating extremal coefficients, see Section 4. Since  $\partial\{\mathbf{x} \in E : \ell_I(\mathbf{x}) > 1\} \subset \{\mathbf{x} \in E : \ell_I(\mathbf{x}) = 1\}$  is a  $\mu^*$ -null set, the limit

$$\tau_I := \mu^*(\{\mathbf{x} \in E : \ell_I(\mathbf{x}) > 1\}) = \lim_{u \rightarrow \infty} a^*(u) \mathbb{P}(\ell_I(\mathbf{X}^*) > u) \quad (13)$$

exists and, by Eq. (2), equals  $\mu^*(\{\mathbf{x} \in E : \ell_I(\mathbf{x}) > 1\})$ . Furthermore, for every  $I \neq \emptyset$ , it satisfies the relation

$$\tau_I = \mathbb{E}[\ell_I(\Theta)] \cdot \tau \in [1, |I|]. \quad (14)$$

Using the property  $\mathbb{E}[\ell_I(\Theta)] > 0$  for every  $I \neq \emptyset$ , we can study the behaviour of the  $\ell_I$ -spectral vector  $\Theta^I$  whose distribution can be defined from the original spectral measure via the relation

$$\mathbb{P}(\Theta^I \in A) = \frac{1}{\mathbb{E}[\ell_I(\Theta)]} \int_{[0, \infty)^d} \mathbb{1}\{\theta / \ell_I(\theta) \in A\} \ell_I(\theta) \mathbb{P}(\Theta \in d\theta), \quad (15)$$

where  $A \subset \{\mathbf{w} \in [0, \infty)^d : \ell_I(\mathbf{w}) = 1\}$ . Note that, from Eq. (14) and Eq. (15), we directly obtain an analogous result to Eq. (12), namely

$$\mathbb{E}[\Theta_i^I] = \frac{1}{\tau_I}, \quad i = 1, \dots, d. \quad (16)$$

Analogues to Eq. (9) and to Lemma 2 are given in the following proposition.

**Proposition 4.** *Let  $\mathbf{X}^*$  be a normalized regularly varying random vector with index 1. Then, for every  $\mathbf{v} \in \mathbb{R}^d$  and  $p \in \mathbb{N}$ , we have*

$$\mathbb{P}(\mathbf{v}^\top \mathbf{X}^* > u \mid \ell_I(\mathbf{X}^*) > u) \longrightarrow \mathbb{E} \left[ \left( \mathbf{v}^\top \Theta^I \right)_+ \wedge 1 \right] = \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \Theta \right)_+ \wedge \ell_I(\Theta) \right]}{\mathbb{E}[\ell_I(\Theta)]} \quad (17)$$

$$\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{X}^*}{\ell_I(\mathbf{X}^*)} \right)_+^p \mid \ell_I(\mathbf{X}^*) > u \right] \longrightarrow \mathbb{E} \left[ \left( \mathbf{v}^\top \Theta^I \right)_+^p \right] = \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\Theta}{\ell_I(\Theta)} \right)_+^p \cdot \ell_I(\Theta) \right]}{\mathbb{E}[\ell_I(\Theta)]} \quad (18)$$

as  $u \rightarrow \infty$ .

*Proof.* First, we notice that  $\ell_I(\mathbf{x}) \leq \|\mathbf{x}\|$  on  $[0, \infty)^d \setminus \{\mathbf{0}\}$  by definition of  $\ell_I$ . Thus  $\ell_I(\mathbf{X}^*) > u$  implies that  $\|\mathbf{X}^*\| > u$  and we have

$$\begin{aligned} \lim_{u \rightarrow \infty} \mathbb{P}(\mathbf{v}^\top \mathbf{X}^* > u \mid \ell_I(\mathbf{X}^*) > u) &= \lim_{u \rightarrow \infty} \frac{\mathbb{P}(\mathbf{v}^\top \mathbf{X}^* > u, \ell_I(\mathbf{X}^*) > u \mid \|\mathbf{X}^*\| > u)}{\mathbb{P}(\ell_I(\mathbf{X}^*) > u \mid \|\mathbf{X}^*\| > u)} \\ &= \frac{\mathbb{P}(Y(\mathbf{v}^\top \Theta) > 1, Y \ell_I(\Theta) > 1)}{\mathbb{P}(Y \ell_I(\Theta) > 1)} = \frac{\mathbb{E}[(\mathbf{v}^\top \Theta)_+ \wedge \ell_I(\Theta) \wedge 1]}{\mathbb{E}[\ell_I(\Theta) \wedge 1]} \\ &= \frac{1}{\mathbb{E}[\ell_I(\Theta)]} \cdot \mathbb{E}[(\mathbf{v}^\top \Theta)_+ \wedge \ell_I(\Theta)] = \mathbb{E}[(\mathbf{v}^\top \Theta^I)_+ \wedge 1], \end{aligned}$$

where we used the fact that  $\ell_I(\Theta) \leq \|\Theta\| = 1$  and we apply the change of measure (15). Using the same arguments and applying Lemma 1 to the bounded measurable function  $g(y, \theta) = (\mathbf{v}^\top \theta)_+^p \mathbb{1}\{\ell_I(y\theta) > 1\}$  whose discontinuity set  $\{\ell_I(y\theta) = 1\}$  has no limiting mass, i.e.  $\mathbb{P}(\ell_I(Y\Theta) = 1) = \mathbb{P}(Y \ell_I(\Theta) = 1) = 0$ , we further obtain

$$\begin{aligned} &\lim_{u \rightarrow \infty} \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{X}^*}{\ell_I(\mathbf{X}^*)} \right)_+^p \mid \ell_I(\mathbf{X}^*) > u \right] \\ &= \lim_{u \rightarrow \infty} \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{X}^*}{\ell_I(\mathbf{X}^*)} \right)_+^p \mathbb{1}\{\ell_I(\mathbf{X}^*) > u\} \mid \|\mathbf{X}^*\| > u \right]}{\mathbb{P}(\ell_I(\mathbf{X}^*) > u \mid \|\mathbf{X}^*\| > u)} \\ &= \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\Theta}{\ell_I(\Theta)} \right)_+^p \mathbb{1}\{Y \ell_I(\Theta) > 1\} \right]}{\mathbb{P}(Y \ell_I(\Theta) > 1)} = \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\Theta}{\ell_I(\Theta)} \right)_+^p \cdot (\ell_I(\Theta) \wedge 1) \right]}{\mathbb{E}[\ell_I(\Theta) \wedge 1]} \end{aligned}$$

$$= \frac{1}{\mathbb{E}[\ell_I(\Theta)]} \cdot \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\Theta}{\ell_I(\Theta)} \right)_+^p \cdot \ell_I(\Theta) \right] = \mathbb{E} \left[ (\mathbf{v}^\top \Theta^I)_+^p \right].$$

□

Similarly to Prop. 3, it can be shown that the distribution of  $\Theta^I$  is uniquely determined by

$$\{\mathbb{E}[(\mathbf{v}^\top \Theta^I)_+^p]; \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in \mathbb{N}\} = \{\mathbb{E}[(\mathbf{v}^\top \Theta^I)^p]; \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in \mathbb{N}\}.$$

While the  $\ell_I$ -spectral measure, i.e. the distribution of  $\Theta^I$ , can be written in terms of the original spectral measure as in (15), the original spectral measure can be uniquely recovered from the distribution of  $\Theta^I$  only if  $\mathbb{P}(\ell_I(\Theta) = 0) = 0$ . Then, the inverse transformation is

$$\mathbb{P}(\Theta \in A) = \frac{1}{\mathbb{E}[\|\Theta^I\|]} \int_{[0, \infty)^d} \mathbb{1}\{\theta / \|\theta\| \in A\} \|\theta\| \mathbb{P}(\Theta^I \in d\theta), \quad A \subset S_{d-1}^+. \quad (19)$$

Note that the discussion above can be easily extended to any positively homogeneous continuous functional  $\ell : [0, \infty)^d \rightarrow [0, \infty)$  with corresponding constant  $\tau_\ell$  which is positive.

### 3 Asymptotic normality

#### 3.1 Asymptotic normality for deterministic thresholds

Assume that we observe independent copies  $\mathbf{X}_l^* = (X_{l1}^*, \dots, X_{ld}^*)^\top, l = 1, \dots, n$ , of the normalized regularly varying random vector  $\mathbf{X}^*$ . Consider a non-empty subset  $I \subset \{1, \dots, d\}$ . One of the classical approaches of Pickands [1975] to estimate the distribution of the  $\ell_I$ -spectral vector  $\Theta^I$  is based on the conditional exceedance probabilities of convex combinations given in relation (17), i.e.

$$\lim_{u \rightarrow \infty} \mathbb{P}(\mathbf{v}^\top \mathbf{X}^* > u \mid \ell_I(\mathbf{X}^*) > u) = \mathbb{E}[(\mathbf{v}^\top \Theta^I)_+ \wedge 1]$$

for  $\mathbf{v} \in \partial B_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_1 = 1\}$ . The associated empirical estimators are of the form

$$\frac{\widehat{P}_{n,u,I}^{(\text{conv})}(\mathbf{v})}{\widehat{P}_{n,u,I}} = \frac{n^{-1} \sum_{l=1}^n \mathbb{1}\{\mathbf{v}^\top \mathbf{X}_l^* > u, \ell_I(\mathbf{X}_l^*) > u\}}{n^{-1} \sum_{l=1}^n \mathbb{1}\{\ell_I(\mathbf{X}_l^*) > u\}}, \quad \mathbf{v} \in \partial B_1(\mathbf{0}). \quad (20)$$

In view of Prop. 3, we promote in this paper the use of the alternative characterization of the distribution of the  $\ell_I$ -spectral vector  $\Theta^I$  based on the limits of moments of convex combinations provided by Eq. (18), i.e.

$$\lim_{u \rightarrow \infty} \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{X}^*}{\ell_I(\mathbf{X}^*)} \right)^p \mid \ell_I(\mathbf{X}^*) > u \right] = \mathbb{E}[(\mathbf{v}^\top \Theta^I)^p],$$

for every  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$  and  $p \in \mathbb{N}_0$ . The associated empirical estimators are of the form

$$\frac{\widehat{M}_{n,u,I}(\mathbf{v}, p)}{\widehat{P}_{n,u,I}} = \frac{n^{-1} \sum_{l=1}^n \left( \mathbf{v}^\top \frac{\mathbf{X}_l^*}{\ell_I(\mathbf{X}_l^*)} \right)^p \mathbb{1}\{\ell_I(\mathbf{X}_l^*) > u\}}{n^{-1} \sum_{l=1}^n \mathbb{1}\{\ell_I(\mathbf{X}_l^*) > u\}}, \quad \mathbf{v} \in \partial B_1^+(\mathbf{0}). \quad (21)$$

In the following, we will consider the new estimators  $\widehat{M}_{n,u,\ell}(\mathbf{v}, p) / \widehat{P}_{n,u,\ell}$  for any  $p \in \mathbb{N}_0$ . As the functional  $\ell_I$  puts emphasis on the components  $(X_i^*)_{i \in I}$  of  $\mathbf{X}^*$  only, we will focus on the same components in the convex combination, i.e. we will consider weight vectors  $\mathbf{v}_I$  where  $\mathbf{v}_I$  is the vector with components  $v_i$  for  $i \in I$  and 0 otherwise. Thus, we obtain  $\mathbf{v}_I^\top \mathbf{X}^* = \sum_{i \in I} v_i X_i^*$ .

The next theorem will provide a functional central limit theorem for the estimators  $\widehat{M}_{n,u,I}(\mathbf{v}_I, p)$  noticing that  $\widehat{P}_{n,u,I} = \widehat{M}_{n,u,I}(\mathbf{v}_I, 0)$  for all  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ . Then we can apply the functional delta method to obtain the asymptotic result for the ratio estimator  $\widehat{M}_{n,u,I}(\mathbf{v}_I, p) / \widehat{P}_{n,u,I}$ .

In this subsection, we assume observations from the normalized random vector  $\mathbf{X}^*$ , in Subsection 3.2 the more general case of observations from the regularly varying vector  $\mathbf{X}$  with unknown index  $\alpha > 0$  and unknown marginal scales is considered. To allow for this extension, in the next theorem, we do not only prove a functional central limit theorem for  $\widehat{M}_{n,u,I}(\mathbf{v}_I, p)$  with  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$  and  $p \in \mathbb{N}_0$ , but already for a more general version.

Indeed, it will turn out to be useful to allow for multiplication of single components of  $\mathbf{X}^*$  by a vector  $\mathbf{s} \in [0, \infty)^d$  of factors, and for raising them up to a power of  $1/\beta$  with  $\beta \in (0, \infty)^d$ . In the following, let  $(\mathbf{s} \circ \mathbf{X}^*)^{1/\beta} = (s_i^{1/\beta_i} X_i^*)_{1 \leq i \leq d}$ . With this notation, we can write

$$\ell_I(\mathbf{X}^*) = \max_{i \in I} |X_i^*| = \|(\mathbf{1}_I \circ \mathbf{X}^*)^{\mathbf{1}}\|,$$

where  $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^d$ , i.e., according to the notation introduced above,  $\mathbf{1}_I$  is the vector with the  $i$ th component being equal to 1 if  $i \in I$  and being equal to 0 otherwise. Furthermore,  $\mathbf{v}_I^\top \mathbf{X}^* = \mathbf{v}^\top (\mathbf{1}_I \circ \mathbf{X}^*)$ . Thus, we can rewrite

$$\begin{aligned} \widehat{M}_{n,u,I}(\mathbf{v}_I, p) &= \frac{1}{n} \sum_{l=1}^n \left( \mathbf{v}^\top \frac{\mathbf{1}_I \circ \mathbf{X}_l^*}{\|\mathbf{1}_I \circ \mathbf{X}_l^*\|} \right)^p \mathbb{1}\{\|\mathbf{1}_I \circ \mathbf{X}_l^*\| > u\} \\ &= \frac{1}{n} \sum_{l=1}^n \left( \mathbf{v}^\top \frac{\mathbf{1}_I \circ \mathbf{X}_l^*/u}{\|\mathbf{1}_I \circ \mathbf{X}_l^*/u\|} \right)^p \mathbb{1}\{\|\mathbf{1}_I \circ \mathbf{X}_l^*/u\| > 1\}, \end{aligned} \quad (22)$$

where the right-hand side of the equation depends on the components  $v_i$ ,  $i \in I$ , of  $\mathbf{v}$  only. In our extended version of the functional central limit theorem, we will not restrict our attention to the vector  $\mathbf{1}_I \circ \mathbf{X}^*$  being extreme, but a perturbed vector  $(\mathbf{s} \circ \mathbf{X}^*)^{1/\beta}$  with  $\mathbf{s}$  being close to  $\mathbf{1}_I$  and  $\beta$  being close to  $\mathbf{1}$ . More precisely, we consider  $\mathbf{s} \in A_{\delta,I}$  and  $\beta \in [(1+\delta)^{-1}, 1+\delta]^d$  where

$$A_{\delta,I} = \{\mathbf{x} \in (0, \infty)^d : x_i = 0 \text{ for all } i \notin I, (1+\delta)^{-1} \leq x_i \leq 1+\delta \text{ for all } i \in I\}$$

for some  $\delta \geq 0$ , i.e. we allow for a perturbation of a factor and a power between  $(1+\delta)^{-1}$  and  $1+\delta$  in each component of  $\mathbf{s}$  and in each component of  $\beta$ . We also denote the disjoint unions

$$A_\delta = \bigcup_{\emptyset \neq I \subset \{1, \dots, d\}} A_{\delta,I} \quad \text{and} \quad A'_\delta = \bigcup_{\emptyset \neq I \subset \{1, \dots, d\}} A_{\delta,I} \times [(1+\delta)^{-1}, 1+\delta]^d.$$

Replacing  $\mathbf{1}_I \circ \mathbf{X}^*/u$  in Eq. (22) by  $(\mathbf{s} \circ \mathbf{X}^*/u)^{1/\beta}$ , we obtain the generalized estimator

$$\begin{aligned} \widehat{M}_{n,u}(\mathbf{v}, \mathbf{s}, \beta, p) &= \frac{1}{n} \sum_{l=1}^n \left( \mathbf{v}^\top \frac{(\mathbf{s} \circ \mathbf{X}_l^*/u)^{1/\beta}}{\|(\mathbf{s} \circ \mathbf{X}_l^*/u)^{1/\beta}\|} \right)^p \mathbb{1}\{\|(\mathbf{s} \circ \mathbf{X}_l^*/u)^{1/\beta}\| > 1\} \\ &= \frac{1}{n} \sum_{l=1}^n \left( \mathbf{v}^\top \frac{(\mathbf{s} \circ \mathbf{X}_l^*/u)^{1/\beta}}{\|(\mathbf{s} \circ \mathbf{X}_l^*/u)^{1/\beta}\|} \right)^p \mathbb{1}\{\|\mathbf{s} \circ \mathbf{X}_l^*\| > u\} \end{aligned}$$

for  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ ,  $(\mathbf{s}, \beta) \in A'_\delta$  and  $p \in \mathbb{N}_0$ . By definition,  $\widehat{M}_{n,u,I}(\mathbf{v}_I, p) = \widehat{M}_{n,u}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p)$  for every  $I \subset \{1, \dots, d\}$ . In particular, the corresponding index set  $I$  of components being relevant for the maximum can be read off from the components of  $\mathbf{s}$  that are strictly positive. Analogously, with the convention  $0^0 = 1$  we define  $\widehat{P}_{n,u}(\mathbf{s}) = \widehat{M}_{n,u}(\mathbf{v}, \mathbf{s}, \beta, 0)$  which depends neither on  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$  nor on  $\beta \in [(1+\delta)^{-1}, 1+\delta]^d$ , and satisfies  $\widehat{P}_{n,u,I} = \widehat{P}_{n,u}(\mathbf{1}_I)$ .

Analogously to Prop. 4, we obtain

$$\begin{aligned} &a^*(u) \mathbb{E} \left[ \widehat{M}_{n,u}(\mathbf{v}, \mathbf{s}, \beta, p) \right] \\ &= a^*(u) \mathbb{P}(\|\mathbf{s} \circ \mathbf{X}^*\| > u) \cdot \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{(\mathbf{s} \circ \mathbf{X}^*/u)^{1/\beta}}{\|(\mathbf{s} \circ \mathbf{X}^*/u)^{1/\beta}\|} \right)^p \middle| \|\mathbf{s} \circ \mathbf{X}^*\| > u \right] \\ &\xrightarrow{u \rightarrow \infty} \tau \cdot \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{(Y \mathbf{s} \circ \Theta)^{1/\beta}}{\|(Y \mathbf{s} \circ \Theta)^{1/\beta}\|} \right)^p \mathbb{1}\{Y \|\mathbf{s} \circ \Theta\| > 1\} \right], \end{aligned} \quad (23)$$

where we used Eq. (14) and Eq. (15). Beyond the limiting behaviour of the expectation, the following theorem establishes asymptotic normality of  $\widehat{M}_{n,u}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p)$ . A detailed proof is given in Appendix A.

**Theorem 5.** *Let  $\mathbf{X}_l^*$ ,  $l \in \mathbb{N}$ , be independent copies of a regularly varying  $[0, \infty)^d$ -valued random vector  $\mathbf{X}^*$  with index  $\alpha = 1$  satisfying Eq. (3) and with spectral component denoted by  $\Theta$ . Furthermore, let  $u_n \rightarrow \infty$  such that*



$n/a^*(u_n) \rightarrow \infty$ . Then, for every finite set  $K_0 \subset \mathbb{N}_0$  and  $\delta \geq 0$ , the sequence of processes  $(\{G_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p); \mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K_0\})_{n \in \mathbb{N}}$  with

$$G_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) = \sqrt{\frac{n}{a^*(u_n)}} \left[ a^*(u_n) \widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) - a^*(u_n) \mathbb{E}[\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)] \right] \quad (24)$$

converges weakly in  $\ell^\infty(\partial B_1^+(\mathbf{0}) \times A'_\delta \times K_0)$  to a tight centered Gaussian process  $G$  with covariance

$$\begin{aligned} & \text{Cov}(G(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p_1), G(\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma}, p_2)) \\ &= \tau \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{(Y \mathbf{s} \circ \boldsymbol{\Theta})^{1/\beta}}{\|(Y \mathbf{s} \circ \boldsymbol{\Theta})^{1/\beta}\|} \right)^{p_1} \left( \mathbf{w}^\top \frac{(Y \mathbf{t} \circ \boldsymbol{\Theta})^{1/\gamma}}{\|(Y \mathbf{t} \circ \boldsymbol{\Theta})^{1/\gamma}\|} \right)^{p_2} \mathbf{1}\{Y(\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|) > 1\} \right]. \end{aligned} \quad (25)$$

A direct consequence of the weak convergence of  $G_n$  is that  $a^*(u_n) \widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)$  converges in probability to the limit expression in Eq. (23). Consequently, the ratio estimator  $\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) / \widehat{P}_{n,u_n}(\mathbf{s})$  is consistent:

$$\frac{\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)}{\widehat{P}_{n,u_n}(\mathbf{s})} = \frac{\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)}{\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, 0)} \rightarrow_p c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p), \quad n \rightarrow \infty, \quad (26)$$

for all  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ ,  $(\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta$  and  $p \in \mathbb{N}_0$  where

$$c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) := \frac{1}{\mathbb{E}[\|\mathbf{s} \circ \boldsymbol{\Theta}\|]} \cdot \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{(Y \mathbf{s} \circ \boldsymbol{\Theta})^{1/\beta}}{\|(Y \mathbf{s} \circ \boldsymbol{\Theta})^{1/\beta}\|} \right)^p \mathbf{1}\{Y \|\mathbf{s} \circ \boldsymbol{\Theta}\| > 1\} \right].$$

Applying Thm. 5 and using the functional Delta method, we can even establish the desired functional limit theorem for the ratio estimator  $\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) / \widehat{P}_{n,u_n}(\mathbf{s})$  provided that we can neglect the preasymptotic bias that arises from the fact that Eq. (23) yields an asymptotic relation only.

**Corollary 6.** *Let the assumptions of Thm. 5 hold and assume that there exist some  $\delta \geq 0$  and some finite subset  $K \subset \mathbb{N}$  such that, for all  $p \in K \cup \{0\}$ ,*

$$\sqrt{\frac{n}{a^*(u_n)}} \left| a^*(u_n) \mathbb{E}[\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)] - \tau \mathbb{E}[\|\mathbf{s} \circ \mathbf{X}^*\|] c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) \right| \rightarrow 0 \quad (27)$$

uniformly in  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$  and  $(\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta$ .

Then, the sequence of processes  $(\{\widetilde{G}_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p); \mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K\})_{n \in \mathbb{N}}$  with

$$\widetilde{G}_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) = \sqrt{\frac{n}{a^*(u_n)}} \left( \frac{\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)}{\widehat{P}_{n,u_n}(\mathbf{s})} - c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) \right),$$

converges weakly in  $\ell^\infty(\partial B_1^+(\mathbf{0}) \times A'_\delta \times K)$  to a tight centered Gaussian process  $\widetilde{G}$ .

The proof of Cor. 6 is deferred to Appendix B. The expression of the covariance of  $\widetilde{G}$  is provided there in Eq. (48).

*Remark 7.* 1. For the original estimator with deterministic threshold as given in (21), it is sufficient to restrict to the case that  $\mathbf{s} = \mathbf{t} = \mathbf{1}_I$  and  $\boldsymbol{\beta} = \mathbf{1}$  in Cor. 6. Then, we have  $\|\mathbf{s} \circ \boldsymbol{\Theta}\|_\infty = \|\mathbf{t} \circ \boldsymbol{\Theta}\|_\infty = \ell_I(\boldsymbol{\Theta})$  and the covariance of  $\widetilde{G}$  given in (48) simplifies to

$$\begin{aligned} & \text{Cov} \left( \widetilde{G}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p_1), \widetilde{G}(\mathbf{w}, \mathbf{1}_I, \mathbf{1}, p_2) \right) \\ &= \frac{1}{\tau \mathbb{E} \ell_I(\boldsymbol{\Theta})} \cdot \left( \frac{1}{\mathbb{E} \ell_I(\boldsymbol{\Theta})} \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\boldsymbol{\Theta}_I}{\ell_I(\boldsymbol{\Theta})} \right)^{p_1} \cdot \left( \mathbf{w}^\top \frac{\boldsymbol{\Theta}_I}{\ell_I(\boldsymbol{\Theta})} \right)^{p_2} \cdot \ell_I(\boldsymbol{\Theta}) \right] \right. \\ & \quad \left. - \frac{1}{(\mathbb{E} \ell_I(\boldsymbol{\Theta}))^2} \mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\boldsymbol{\Theta}_I}{\ell_I(\boldsymbol{\Theta})} \right)^{p_1} \cdot \ell_I(\boldsymbol{\Theta}) \right] \mathbb{E} \left[ \left( \mathbf{w}^\top \frac{\boldsymbol{\Theta}_I}{\ell_I(\boldsymbol{\Theta})} \right)^{p_2} \cdot \ell_I(\boldsymbol{\Theta}) \right] \right) \\ &= \frac{1}{\tau_I} \text{Cov} \left( (\mathbf{v}^\top \boldsymbol{\Theta}_I^I)^{p_1}, (\mathbf{w}^\top \boldsymbol{\Theta}_I^I)^{p_2} \right). \end{aligned}$$

2. Using the same techniques as in the proof of Cor. 6, one can easily show the joint weak convergence of  $(G_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p_1), \tilde{G}_n(\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma}, p_2))_{n \in \mathbb{N}}$   $(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p_1), (\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma}, p_2)$  in  $\partial B_1^+(\mathbf{0}) \times A'_\delta \times K$  to  $(G(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p_1), \tilde{G}(\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma}, p_2))$  which are jointly Gaussian. Then for  $\boldsymbol{\beta} = \boldsymbol{\gamma} = \mathbf{1}$ , the covariances express as

$$\begin{aligned} & \text{Cov}(G(\mathbf{v}, \mathbf{s}, \mathbf{1}, p_1), \tilde{G}(\mathbf{w}, \mathbf{t}, \mathbf{1}, p_2)) \\ &= \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{p_1} \left( \mathbf{w}^\top \frac{\mathbf{t} \circ \boldsymbol{\Theta}}{\|\mathbf{t} \circ \boldsymbol{\Theta}\|} \right)^{p_2} (\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|) \right]}{\mathbb{E}[\|\mathbf{t} \circ \boldsymbol{\Theta}\|]} \\ &= \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{p_1} (\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|) \right] \mathbb{E} \left[ \left( \mathbf{w}^\top \frac{\mathbf{t} \circ \boldsymbol{\Theta}}{\|\mathbf{t} \circ \boldsymbol{\Theta}\|} \right)^{p_2} \|\mathbf{t} \circ \boldsymbol{\Theta}\| \right]}{(\mathbb{E}[\|\mathbf{t} \circ \boldsymbol{\Theta}\|])^2}. \end{aligned}$$

3. From the proofs of Thm. 5 and Cor. 6, it follows that, for fixed  $p \in \mathbb{N}_0$ , the sequences of processes  $G_n(\cdot, \cdot, \cdot, p)$  and  $\tilde{G}_n(\cdot, \cdot, \cdot, p)$  are asymptotically uniformly equicontinuous on  $\partial B_1^+(\mathbf{0}) \times A'_\delta$  w.r.t. the semimetric

$$\rho((\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}), (\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma})) = \max\{\|\mathbf{v} - \mathbf{w}\|, \|\mathbf{s} - \mathbf{t}\|, \|\boldsymbol{\beta} - \boldsymbol{\gamma}\|\}.$$

Consequently, almost all sample paths of the tight limits  $G$  and  $\tilde{G}$ , respectively, are uniformly  $\rho$ -continuous [cf. Addendum 1.5.8 in Van der Vaart and Wellner, 1996]. Since  $\rho$  induces the standard topology on  $\partial B_1^+(\mathbf{0}) \times A'_\delta$  as subset of  $\mathbb{R}^d \times \mathbb{R}^{|\mathcal{I}|} \times \mathbb{R}^d$ , it follows that almost all sample paths of  $G(\cdot, \cdot, \cdot, p)$  and  $\tilde{G}(\cdot, \cdot, \cdot, p)$  are uniformly continuous.

### 3.2 Asymptotic normality for random thresholds via order statistics

Consider the general case of a random vector  $\mathbf{X}$  that is non standard regularly varying such that  $\mathbf{X}^* = (r_1^{-1}X_1^{\alpha_1}, \dots, r_d^{-1}X_d^{\alpha_d})$  is regularly varying with index 1 and satisfies (2). In practice, the marginal distributions of  $X_1, \dots, X_d$  are unknown. In particular, the indices  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$  of regular variation and the scaling factors  $r_1, \dots, r_d$  that are needed for the transformation to  $\mathbf{X}^*$  cannot be assumed to be given. Furthermore, the normalizing function  $a^*(\cdot) \sim \mathbb{P}(X_i^* > \cdot)$  is unknown. To cope with these difficulties, we propose to work with an estimator  $\hat{\boldsymbol{\alpha}}$  of  $\boldsymbol{\alpha}$  and order statistics of  $\mathbf{X}$ . More precisely, let  $X_{k:n,i}$  denote the  $k$ th upper order statistic of  $X_{1i}, \dots, X_{ni}$  for  $i = 1, \dots, d$  and let  $\mathbf{X}_{k:n} = (X_{k:n,1}, \dots, X_{k:n,d})$ . With no loss of generality we assume that  $X_{k:n,i} > 0$ ,  $1 \leq i \leq d$ . We propose the following estimator

$$\frac{\tilde{M}_{n,k,I}(\mathbf{v}, p)}{\tilde{P}_{n,k,I}} = \frac{n^{-1} \sum_{l=1}^n \left( \mathbf{v}_I^\top \frac{(\mathbf{X}_l / \mathbf{X}_{k:n})^{\hat{\boldsymbol{\alpha}}}}{\ell_I((\mathbf{X}_l / \mathbf{X}_{k:n})^{\hat{\boldsymbol{\alpha}}})} \right)^p \mathbf{1}\{\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}}{n^{-1} \sum_{l=1}^n \mathbf{1}\{\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}}$$

for some  $k < n$ , where the ratio of vectors  $\mathbf{X}_l / \mathbf{X}_{k:n}$  is to be interpreted componentwise. Expanding each of the ratios  $\mathbf{X}_l / \mathbf{X}_{k:n}$  by  $r_1^{-1}, \dots, r_d^{-1}$ , respectively, and taking the power of  $\boldsymbol{\alpha}$  coordinatewise, it can be easily seen that

$$\frac{\tilde{M}_{n,k,I}(\mathbf{v}, p)}{\tilde{P}_{n,k,I}} = \frac{n^{-1} \sum_{l=1}^n \left( \mathbf{v}_I^\top \frac{(\mathbf{X}_l^* / \mathbf{X}_{k:n}^*)^{\hat{\boldsymbol{\alpha}}/\boldsymbol{\alpha}}}{\ell_I((\mathbf{X}_l^* / \mathbf{X}_{k:n}^*)^{\hat{\boldsymbol{\alpha}}/\boldsymbol{\alpha}})} \right)^p \mathbf{1}\{\ell_I(\mathbf{X}_l^* / \mathbf{X}_{k:n}^*) > 1\}}{n^{-1} \sum_{l=1}^n \mathbf{1}\{\ell_I(\mathbf{X}_l^* / \mathbf{X}_{k:n}^*) > 1\}},$$

i.e. we can rewrite

$$\frac{\tilde{M}_{n,k,I}(\mathbf{v}, p)}{\tilde{P}_{n,k,I}} = \frac{\widehat{M}_{n,u_n}(\mathbf{v}_I, u_n / \mathbf{X}_{k:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}, p)}{\widehat{P}_{n,u_n}(u_n / \mathbf{X}_{k:n}^* \circ \mathbf{1}_I)}. \quad (28)$$

For an appropriate sequence  $\{k_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$  with  $k_n \rightarrow \infty$  such that  $k_n/n \rightarrow 0$ , we want to show that

$$\left( \left\{ \sqrt{k_n} \left( \frac{\tilde{M}_{n,k_n,I}(\mathbf{v}, p)}{\tilde{P}_{n,k_n,I}} - \mathbb{E}[(\mathbf{v}_I^\top \boldsymbol{\Theta}_I^I)^p] \right) \right\}; \mathbf{v} \in \partial B_1^+(\mathbf{0}) \right)_{n \in \mathbb{N}}$$

converges weakly to a Gaussian process. The functional central limit theorem will essentially be proven by plugging the random scaling factor  $\mathbf{s} = u_n / \mathbf{X}_{k:n}^* \circ \mathbf{1}_I$  and the random power  $\boldsymbol{\beta} = \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}$  into the results of Subsection 3.1 according to Eq. (28). Provided that both the random  $\mathbf{s}$  and the random  $\boldsymbol{\beta}$  are asymptotically normal, we can use the functional central limit theorem given in Cor. 6 to obtain the desired result.

To this end, we will first specify our estimators  $\hat{\boldsymbol{\alpha}}_{n,k} = (\hat{\alpha}_{n,k,1}, \dots, \hat{\alpha}_{n,k,d})$  of  $\boldsymbol{\alpha}$  which will be componentwise Hill estimators. Similarly to Section 3.1, we will first start with an estimator with a fixed sequence of thresholds  $\{u_n\}_{n \in \mathbb{N}}$

and normalized observations  $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ . To mimic the behaviour of the real non-normalized data  $\mathbf{X} = (\mathbf{r} \circ \mathbf{X}^*)^{1/\alpha}$  where  $\mathbf{r} = (r_1, \dots, r_d)$ , we also allow for a scaling factor  $\mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}}$ , that is, a scaling factor with exactly one non-zero component, and take the logarithm of the scaled vector to the coordinatewise power of  $1/\alpha$ . Thus, for  $\mathbf{s} \in A_{\delta, \{i\}}$ ,  $1 \leq i \leq d$ , we obtain the estimator

$$\frac{\widehat{L}_{n,u}(\mathbf{s})}{\widehat{P}_{n,u}(\mathbf{s})} = \frac{n^{-1} \sum_{l=1}^n \log \left( \left\| \left( \mathbf{s} \circ \frac{\mathbf{X}_l^*}{u_n} \right)^{1/\alpha} \right\| \right) \mathbb{1}\{\|\mathbf{s} \circ \mathbf{X}_l^*\| > u_n\}}{\widehat{P}_{n,u}(\mathbf{s})}. \quad (29)$$

Under the assumptions of Thm. 5, it can be shown that for  $\mathbf{s} \in A_{\delta, \{i\}}$ ,  $1 \leq i \leq d$ ,

$$a^*(u_n) \mathbb{E}[\widehat{L}_{n,u_n}(\mathbf{s})] = \frac{\mathbb{P}(s_i X_i^* > u_n)}{\mathbb{P}(X_i^* > u_n)} \cdot \mathbb{E} \left[ \frac{1}{\alpha_i} \log \left( \frac{s_i X_i^*}{u_n} \right) \middle| s_i X_i^* > u_n \right] \rightarrow \frac{s_i}{\alpha_i} \quad (n \rightarrow \infty).$$

In order to neglect the bias we assume the following second order condition from De Haan and Ferreira [2006]:

There exists an auxiliary positive function  $A_i^*$  such that  $A_i^*(t) \rightarrow 0$  as  $t \rightarrow \infty$  and

$$\lim_{x \rightarrow \infty} \frac{a^*(x) \mathbb{P}(X_i^* > xs) - s^{-1}}{A_i^*(a^*(x))} = K_i(s) \quad (30)$$

where  $K_i$  is not identically 0 for any  $1 \leq i \leq d$ .

Now if we choose  $u_n$  such that  $u_n \rightarrow \infty$  and  $\sqrt{n/a^*(u_n)} A_i^*(a^*(u_n)) \rightarrow 0$  as  $n \rightarrow \infty$  for some  $1 \leq i \leq d$  we obtain

$$\sup_{\mathbf{s} \in A_{\delta, \{i\}}} \sqrt{\frac{n}{a^*(u_n)}} \left| a^*(u_n) \mathbb{E}[\widehat{L}_{n,u_n}(\mathbf{1}_{\{i\}}, \mathbf{s})] - \frac{s_i}{\alpha_i} \right| \rightarrow 0, \quad n \rightarrow \infty, \quad (31)$$

as well as

$$\sup_{(1+\delta)^{-1} \leq s_i \leq 1+\delta} \sqrt{\frac{n}{a^*(u_n)}} |a^*(u_n) \mathbb{P}(s_i X_i^* > u_n) - s_i| \rightarrow 0, \quad n \rightarrow \infty, \quad (32)$$

which is nothing else than Eq. (27) for  $p = 0$  and  $\mathbf{s} \in A_{\delta, \{i\}}$ ,  $\delta > 0$ . As Eq. (31) and (32) ensure that the biases of the marginal Hill estimators become asymptotically negligible, the second-order condition (30) will be sufficient for us to establish asymptotic normality of our final marginal Hill estimators  $\widehat{\alpha}_{k_n, i}$  defined via order statistics

$$\frac{1}{\widehat{\alpha}_{n, k, i}} = \frac{n^{-1} \sum_{l=1}^n \log \left( \ell_{\{i\}}(\mathbf{X}_l / \mathbf{X}_{k:n}) \right) \mathbb{1}\{\ell_{\{i\}}(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}}{\widehat{P}_{n, k, \{i\}}} \quad (33)$$

in the following theorem. A detailed proof is given in Appendix C.

**Theorem 8.** *Let  $\mathbf{X}_l$ ,  $l \in \mathbb{N}$ , be independent copies of a non-standard regularly varying  $[0, \infty)^d$ -valued random vector  $\mathbf{X}$ . Assume that the vector  $\mathbf{X}^* = (r_1^{-1} X_1^\alpha, \dots, r_d^{-1} X_d^\alpha)$  satisfies Eq. (3) and (30). Let  $\{k_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$  be a sequence such that  $k_n \rightarrow \infty$ ,  $k_n/n \rightarrow 0$  and  $\sqrt{k_n} A_i^*(n/k_n) \rightarrow 0$  for all  $1 \leq i \leq d$ . Then we have*

$$\sqrt{k_n} \left( \widehat{\alpha}_{n, k_n}^{-1} - \boldsymbol{\alpha}^{-1} \right) \rightarrow_d \widetilde{\mathbf{H}}, \quad n \rightarrow \infty,$$

where  $\widetilde{\mathbf{H}}$  is a  $d$ -dimensional centered Gaussian random vector with covariances

$$\text{Cov}(\widetilde{H}_i, \widetilde{H}_j) = \frac{\tau}{\alpha_i \alpha_j} \mathbb{E}[\Theta_i \wedge \Theta_j] = \frac{2 - \tau_{ij}}{\alpha_i \alpha_j}.$$

This result is in line to results for the joint asymptotic behaviour of marginal Hill estimators obtained in Stupfler [2019].

*Remark 9.* We can show joint convergence of the different estimates to the limit processes  $G$  and  $\widetilde{G}$  and the variables  $\widetilde{H}_i$ ,  $i = 1, \dots, d$ , as obtained in Thm. 5, Cor. 6 and Thm. 8, respectively. The covariance between the processes  $G$  and  $\widetilde{G}$  is given in Rem. 7.2. Similar calculations as above reveal that, for all  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ ,  $J \subset \{1, \dots, d\}$  and  $p \in \mathbb{N}_0$ ,

$$\text{Cov}(\widetilde{H}_i, G(\mathbf{v}, \mathbf{1}_J, \mathbf{1}, p)) = \frac{\tau}{\alpha_i} \mathbb{E} \left[ -\log \left( 1 \wedge \frac{\|\Theta_J\|}{\Theta_i} \right) \cdot \left( \mathbf{v}^\top \frac{\Theta_J}{\|\Theta_J\|} \right)^p \cdot (\Theta_i \wedge \|\Theta_J\|) \right]. \quad (34)$$

Here, using similar arguments as above, we obtain

$$\begin{aligned} & \text{Cov}(\tilde{H}_i, \tilde{G}(\mathbf{v}, \mathbf{1}_J, \mathbf{1}, p)) \\ &= \frac{1}{\alpha_i \mathbb{E}[\|\Theta_J\|]} \left( \mathbb{E} \left[ -\log \left( 1 \wedge \frac{\|\Theta_J\|}{\Theta_i} \right) \left( v^\top \frac{\Theta_J}{\|\Theta_J\|} \right)^p (\Theta_i \wedge \Theta_J) \right] \right. \\ & \quad \left. - \mathbb{E} \left[ -\log \left( 1 \wedge \frac{\|\Theta_J\|}{\Theta_i} \right) (\Theta_i \wedge \Theta_J) \right] \frac{\mathbb{E}[\|\Theta_J\| \left( v^\top \frac{\Theta_J}{\|\Theta_J\|} \right)^p]}{\mathbb{E}[\|\Theta_J\|]} \right] \end{aligned} \quad (35)$$

for all  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$  and  $p \in \mathbb{N}_0$ . In particular, this implies that  $\text{Cov}(\tilde{H}_i, \tilde{G}(\mathbf{v}, \mathbf{1}_J, \mathbf{1}, p)) = 0$  if  $i \in J$ .

Making use of the asymptotic results for the Hill-type estimator in Thm. 8 and for the estimators in the case of deterministic thresholds given in Section 3.1, we will establish a functional central limit theorem for  $\tilde{M}_{n,k,I}(\mathbf{v}_I, p) / \tilde{P}_{n,k,I}$ . Compared to its analogue for deterministic thresholds, the estimator is blurred by the random marginal normalization based on the order statistics. To ease notation, we just write  $G^0(\mathbf{s}) = G(\cdot, \mathbf{s}, \cdot, 0)$ .

Imposing some additional conditions on the function  $c$ , we obtain the following main result.

**Theorem 10.** *Let  $\mathbf{X}_l$ ,  $l \in \mathbb{N}$ , be independent copies of a non-standard regularly varying  $[0, \infty)^d$ -valued random vector  $\mathbf{X}$  satisfying the assumptions of Cor. 6 and Thm. 8 for  $\alpha^*(u_n) \sim n/k_n$ , some  $\delta > 0$  and some  $K \subset \mathbb{N}$ . Assume that all the partial derivatives*

$$\begin{aligned} c_{s_i}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) &= \frac{\partial}{\partial s_i} c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p), & i \in \{1, \dots, d\}, p \in K, \\ c_{\beta_i}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) &= \frac{\partial}{\partial \beta_i} c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p), & i \in \{1, \dots, d\}, p \in K, \end{aligned}$$

exist and are continuous on  $\partial B_1^+(\mathbf{0}) \times A'_\delta$ . Then, for non-empty  $I \subset \{1, \dots, d\}$ , we have

$$\begin{aligned} & \left\{ \sqrt{k} \left( \frac{\tilde{M}_{n,k,I}(\mathbf{v}_I, p)}{\tilde{P}_{n,k,I}} - c(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p) \right); \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K \right\} \\ & \rightarrow \left\{ \tilde{G}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p) - \sum_{i \in I} \left( c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p) G^0(\mathbf{1}_{\{i\}}) + c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p) \alpha_i \tilde{H}_i \right) \right\} \end{aligned}$$

weakly in  $\ell^\infty(\partial B_1^+(\mathbf{0}) \times K)$  as  $n \rightarrow \infty$ .

The proof of Thm. 10 is deferred to Appendix D.

*Remark 11.* In the case  $p = 1$  we notice that the limiting process is linear in  $\mathbf{v}_I$  because the process  $\tilde{G}$  and the functions  $c_{s_i}$  and  $c_{\beta_i}$  are all linear in  $\mathbf{v}_I$ :

$$\begin{aligned} & \tilde{G}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, 1) - \sum_{i \in I} \left( c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, 1) G^0(\mathbf{1}_{\{i\}}) + c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, 1) \alpha_i \tilde{H}_i \right) \\ &= \sum_{j \in I} v_j \left( \tilde{G}(\mathbf{1}_j, \mathbf{1}_I, \mathbf{1}, 1) - \sum_{i \in I} \left( c_{s_i}(\mathbf{1}_j, \mathbf{1}_I, \mathbf{1}, 1) G^0(\mathbf{1}_{\{i\}}) + c_{\beta_i}(\mathbf{1}_j, \mathbf{1}_I, \mathbf{1}, 1) \alpha_i \tilde{H}_i \right) \right). \end{aligned}$$

## 4 Application: Estimation of extremal coefficients

One of the most popular summary statistics for extremal dependence in the (sub-)vector  $(X_i^*)_{i \in I}$  is the extremal coefficient  $\tau_I$  given by Eq. (13), i.e.

$$\tau_I = \lim_{u \rightarrow \infty} a^*(u) \mathbb{P} \left( \max_{i \in I} X_i^* > u \right) \sim \frac{\mathbb{P}(\max_{i \in I} X_i^* > u)}{\mathbb{P}(X_j^* > u)}, \quad \text{for all } j \in I.$$

From this representation and the union bound for  $\mathbb{P}(\max_{i \in I} X_i^* > u)$ , it can be easily seen that  $\tau_I \in [1, |I|]$  where  $\tau_I = 1$  corresponds to the case that all the components of  $(X_i^*)_{i \in I}$  are (asymptotically) fully dependent, while they are all asymptotically independent if  $\tau_I = |I|$ . This gives rise to the common interpretation of the extremal coefficient as the number of asymptotically independent variables among  $(X_i^*)_{i \in I}$  [cf. Schlather and Tawn, 2003]. Even though, in general, the spectral measure of the vector  $\mathbf{X} = (X_1, \dots, X_d)$  cannot be fully described by the set of extremal coefficients  $(\tau_I)_{I \subset \{1, \dots, d\}}$ , they allow for the identification of the parameters in many popular parametric families

and, consequently, statistical efficiency in the estimation of the extremal coefficients is expected to carry over to the estimation of the corresponding model parameters. Illustrations on the Hüsler-Reiss are given in Section 5.

From Eq. (16), we obtain that

$$\mathbb{E}[\mathbf{v}_I^\top \boldsymbol{\Theta}^I] = \mathbb{E}[(\mathbf{v}_I^\top \boldsymbol{\Theta}^I)_+ \wedge 1] = \frac{1}{\tau_I},$$

for all  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$ . This motivates the use of the two classes of estimators considered in Eq. (20) and Eq. (21), respectively. We will distinguish two cases; Section 4.1 collects the asymptotic normality results for the case of known marginal distributions where we can exploit the results for deterministic thresholds given in Section 3.1 with  $\mathbf{s} = \mathbf{1}_I$  and  $\beta = 1$ . The results derived in Section 3.2 can be used when the margins are unknown as we will investigate in Section 4.2.

#### 4.1 Known margins

In this section, we focus on the case where the margins are known, i.e.  $\mathbf{X}_I^*$  is observed.

##### 4.1.1 The inefficiency of the benchmark

For this case, as a first class of estimators, we can directly consider the estimators based on relative frequencies of threshold exceedances by convex combinations, see Eq. (20), namely

$$\frac{\widehat{P}_{n,u,I}^{(\text{conv})}(\mathbf{v}_I)}{\widehat{P}_{n,u,I}} = \frac{n^{-1} \sum_{l=1}^n \mathbf{1}\{\mathbf{v}_I^\top \mathbf{X}_l^* > u\}}{n^{-1} \sum_{l=1}^n \mathbf{1}\{\ell_I(\mathbf{X}_l^*) > u\}}, \quad \mathbf{v}_I \in \partial B_1^+(\mathbf{0}),$$

where we use the fact that  $\mathbf{v}_I^\top \mathbf{X}^* > u$  already implies that  $\ell_I(\mathbf{X}^*) > u$ . This type of estimator will serve as a benchmark. By standard arguments, for a sequence  $\{u_n\}_{n \in \mathbb{N}}$  of thresholds such that  $u_n \rightarrow \infty$  and  $n/a^*(u_n) \rightarrow \infty$  as  $n \rightarrow \infty$ , we obtain the asymptotic normality

$$\sqrt{\frac{n}{a^*(u_n)}} \begin{pmatrix} a^*(u_n) \widehat{P}_{n,u_n,I}^{(\text{conv})}(\mathbf{v}_I) - a^*(u_n) \mathbb{P}(\mathbf{v}_I^\top \mathbf{X}^* > u_n) \\ a^*(u_n) \widehat{P}_{n,u_n,I} - a^*(u_n) \mathbb{P}(\ell_I(\mathbf{X}^*) > u_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \tau_I \mathbb{E}[\mathbf{v}_I^\top \boldsymbol{\Theta}^I] & \tau_I \mathbb{E}[\mathbf{v}_I^\top \boldsymbol{\Theta}^I] \\ \tau_I \mathbb{E}[\mathbf{v}_I^\top \boldsymbol{\Theta}^I] & \tau_I \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & \tau_I \end{pmatrix}.$$

Assuming that the bias is negligible, i.e.

$$\sqrt{\frac{n}{a^*(u_n)}} (a^*(u_n) \mathbb{P}(\mathbf{v}_I^\top \mathbf{X}^* > u_n) - 1) \rightarrow 0$$

and

$$\sqrt{\frac{n}{a^*(u_n)}} (a^*(u_n) \mathbb{P}(\ell_I(\mathbf{X}^*) > u_n) - \tau_I) \rightarrow 0, \quad (36)$$

the Delta method yields asymptotic normality of the benchmark estimator for the case of known margins,

$$\frac{1}{\widehat{\tau}_{I,n}^{BK}(\mathbf{v}_I)} = \frac{\widehat{P}_{n,u_n,I}^{(\text{conv})}(\mathbf{v}_I)}{\widehat{P}_{n,u_n,I}}, \quad \mathbf{v} \in \partial B_1^+(\mathbf{0}),$$

i.e.

$$\sqrt{\frac{n}{a^*(u_n)}} \left( \frac{1}{\widehat{\tau}_{I,n}^{BK}(\mathbf{v}_I)} - \frac{1}{\tau_I} \right) \xrightarrow{d} \mathcal{N}(0, \tau_I^{-3}(\tau_I - 1)).$$

Here, we note that the asymptotic variance does not depend on the choice of  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ .

Secondly, we consider moment-based estimators as discussed in Section 3.1, i.e. estimators based on empirical means of convex combinations, namely  $\widehat{M}_{n,u,I}(\mathbf{v}_I, 1)/\widehat{P}_{n,u,I}$  with weights  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$  as defined in Eq. (21). Provided that the bias is negligible, i.e.

$$\sqrt{\frac{n}{a^*(u_n)}} \left( a^*(u_n) \mathbb{E} \left[ \mathbf{v}_I^\top \frac{\mathbf{X}^*}{\ell_I(\mathbf{X}^*)} \cdot \mathbf{1}\{\ell_I(\mathbf{X}^*) > u_n\} \right] - 1 \right) \rightarrow 0$$

and (36), we obtain the asymptotic normality

$$\sqrt{\frac{n}{a^*(u_n)}} \left( \frac{\widehat{M}_{n,u_n,I}(\mathbf{v}_I)}{\widehat{P}_{n,u_n,I}} - \frac{1}{\tau_I} \right) \xrightarrow{d} \mathcal{N} \left( 0, \tau_I^{-1} \text{Var}(\mathbf{v}_I^\top \Theta^I) \right),$$

cf. the results in Cor. 6 and the first part of Rem. 7.

Since  $\mathbf{v}_I^\top \Theta^I \leq 1$  a.s., we have that

$$\frac{1}{\tau_I} \text{Var}(\mathbf{v}_I^\top \Theta^I) = \frac{1}{\tau_I} \left( \mathbb{E}[(\mathbf{v}_I^\top \Theta^I)^2] - \frac{1}{\tau_I^2} \right) \leq \frac{1}{\tau_I} \left( \mathbb{E}[(\mathbf{v}_I^\top \Theta^I)] - \frac{1}{\tau_I^2} \right) = \frac{\tau_I - 1}{\tau_I^3}$$

and, therefore, the asymptotic variance of the moment-based approach is always less than or equal to the asymptotic variance of the benchmark estimator. Notice that the variances of both estimators are equal, i.e.  $\mathbb{E}[(\mathbf{v}_I^\top \Theta^I)^2] = \mathbb{E}[\mathbf{v}_I^\top \Theta^I] = 1/\tau_I$ , if and only if  $\Theta_i^I \equiv 1$  a.s. for all  $i$  with  $v_i > 0$ . This is the case if and only if  $\tau_I = 1$  and then both asymptotic variances are equal to 0. Thus, except for this degenerate case, the moment-based approach is always preferable to the benchmark in terms of the asymptotic variance.

Furthermore, in general, the asymptotic variance of the moment-based estimator depends on the vector  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$ . As we will see in Example 14, the minimal variance is typically not achieved for a standard vector of the type  $\mathbf{v}_{\{i\}}$  for  $i \in I$ , i.e. the empirical estimators of  $\mathbb{E}[\Theta_i^I]$ , but by a non-trivial mixture of different basis vectors. This improvement can be attributed to the fact that the moment-based estimator is linear in  $\mathbf{v}$ . Thus optimizing in  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$  can be interpreted as aggregating different estimators. It is well-known that the use of a combination of estimators can improve the accuracy. We note that we could also aggregate the benchmark estimator for different  $\mathbf{v}$ . However this strategy is much more intricate to study as the benchmark estimator is not linear in  $\mathbf{v}$ .

Based on this discussion, for the remainder of Section 4.1, we focus on convex combinations of moment-based estimators rather than on the benchmark. In the following, we optimize the asymptotic variance  $\tau_I^{-1} \text{Var}(\mathbf{v}_I^\top \Theta^I)$  w.r.t. the weight vectors  $\mathbf{v}_I$  and we find some approximation of this best aggregation. We should notice that the optimal asymptotic variance  $\text{Var}(\mathbf{v}_I^\top \Theta^I)$  may be degenerate even when  $\tau_I > 1$ . In these cases the results hereinafter still hold with a null limit.

*Remark 12.* The asymptotic variance of the moment-based estimator is zero if and only if there exists a weight vector  $\mathbf{v}_I$  such that  $\mathbf{v}_I^\top \Theta^I = 1/\tau_I$  almost surely. This happens in case of an asymptotically independent vector  $\mathbf{X}$  with  $\mathbf{v}$  satisfying  $v_i = \frac{1}{|I|}$  for  $i \in I$ . It might also happen in more complex cases. For instance, assume that the vector  $\Theta^I$  takes at only a finite number of different possible values  $\theta_I^{(j)}$ ,  $j = 1, \dots, m$ . Then,  $\mathbf{v}_I$  is given as the solution of the linear system  $\mathbf{v}_I^\top \theta_I^{(j)} = 1/\tau_I$ ,  $j = 1, \dots, m$ , provided that the solution is in the simplex.

#### 4.1.2 Variance minimization

We focus on the moment estimators  $\widehat{M}_{n,u,I}(\mathbf{v}_I, 1)/\widehat{P}_{n,u,I}$  for different  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$ . As all of these estimators are consistent, it is appealing to choose the estimator that provides the minimal asymptotic variance. Thus, in the following, we will focus on determining the weight vectors  $\mathbf{v}_I^* \in \partial B_1^+(\mathbf{0})$  that minimize the asymptotic variance

$$\tau_I^{-1} \text{Var}(\mathbf{v}_I^\top \Theta^I) = \tau_I^{-3} (\tau_I^2 \cdot \mathbb{E}[(\mathbf{v}_I^\top \Theta^I)^2] - 1),$$

or, equivalently, the function

$$f^{(I)} : \mathbf{v}_I \mapsto \mathbb{E}[(\mathbf{v}_I^\top \Theta^I)^2].$$

We will show that, under mild conditions, the minimizer of  $f^{(I)}$  in  $\partial B_1^+(\mathbf{0})$  is unique.

**Proposition 13.** *Let  $\Theta$  be a spectral vector. Then, for non-empty  $I \subset \{1, \dots, d\}$ , it holds: If the matrix  $\mathbf{V}_I = (\mathbb{E}[\Theta_i^I \Theta_j^I])_{i,j \in I}$  is conditionally positive definite, i.e.*

$$\mathbf{a}^\top \mathbf{V}_I \mathbf{a} > 0,$$

for all  $\mathbf{a} \in \mathbb{R}^{|I|} \setminus \{\mathbf{0}\}$  such that  $\sum_{i \in I} a_i = 0$ , then the function  $\mathbf{v}_I \mapsto f^{(I)}(\mathbf{v}_I)$  attains its minimum on  $\partial B_1^+(\mathbf{0})$  at some unique  $\mathbf{v}_I^* \in \partial B_1^+(\mathbf{0})$ .

*Proof.* First, we note that

$$f^{(I)}(\mathbf{v}_I) = \sum_{i \in I} \sum_{j \in I} v_i v_j \mathbb{E}(\Theta_i^I \Theta_j^I) = (\mathbf{v}_i)_{i \in I}^\top \mathbf{V}_I (\mathbf{v}_i)_{i \in I}.$$

By a slight abuse of notation, henceforth, we will write  $f^{(I)}(\mathbf{v}_I) = \mathbf{v}_I^\top \mathbf{V}_I \mathbf{v}_I$ . Now, assume that  $f^{(I)}$  attains its minimum at two distinct points  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$  and  $\mathbf{w}_I \in \partial B_1^+(\mathbf{0})$ . Then, for all  $\lambda \in (0, 1)$ , we have  $\lambda \mathbf{v}_I + (1 - \lambda) \mathbf{w}_I \in \partial B_1^+(\mathbf{0})$ ,  $\sum_{i \in I} (v_i - w_i) = 0$  and, therefore,

$$\begin{aligned} f^{(I)}(\lambda \mathbf{v}_I + (1 - \lambda) \mathbf{w}_I) &= (\mathbf{w}_I + \lambda(\mathbf{v}_I - \mathbf{w}_I))^\top \mathbf{V}_I (\mathbf{v}_I + (1 - \lambda)(\mathbf{w}_I - \mathbf{v}_I)) \\ &= -\lambda(1 - \lambda)(\mathbf{v}_I - \mathbf{w}_I)^\top \mathbf{V}_I (\mathbf{v}_I - \mathbf{w}_I) + \lambda \mathbf{v}_I^\top \mathbf{V}_I \mathbf{v}_I + (1 - \lambda) \mathbf{w}_I^\top \mathbf{V}_I \mathbf{w}_I \\ &< \lambda f^{(I)}(\mathbf{v}_I) + (1 - \lambda) f^{(I)}(\mathbf{w}_I), \end{aligned}$$

which is a contradiction to the choice of  $\mathbf{v}_I$  and  $\mathbf{w}_I$  as minimizers of  $f^{(I)}$ .  $\square$

*Example 14.* In the case of a pairwise extremal coefficient  $\tau_I$  with  $I = \{i, j\}$ , we obtain an explicit expression for the minimizer  $\mathbf{v}_{\{i,j\}}^*$  with  $v_i^*, v_j^* \geq 0$  and  $v_i^* + v_j^* = 1$  of the function  $f^{(I)}(\mathbf{v}_{\{i,j\}}) = \mathbb{E}[(v_i \Theta_i^I + v_j \Theta_j^I)^2] = (v_i, v_j) \mathbf{V}_I (v_i, v_j)^\top$ . To this end, we note that the function

$$v_i \mapsto (v_i, 1 - v_i) \mathbf{V}_I \begin{pmatrix} v_i \\ 1 - v_i \end{pmatrix} = v_i^2 \mathbb{E}[(\Theta_i^I)^2] + 2v_i(1 - v_i) \mathbb{E}[\Theta_i^I \Theta_j^I] + (1 - v_i)^2 \mathbb{E}[(\Theta_j^I)^2]$$

is a quadratic function which possesses a unique minimizer  $v_i^* \in [0, 1]$  if and only if  $\mathbb{E}[(\Theta_i^I)^2] - 2\mathbb{E}[\Theta_i^I \Theta_j^I] + \mathbb{E}[(\Theta_j^I)^2] > 0$ , i.e. if and only if  $\mathbf{V}_I$  is conditionally positive definite, cf. Prop. 13. From

$$\begin{aligned} \mathbb{E}[\Theta_i^I \cdot \Theta_j^I] &= \frac{1}{\mathbb{E}[\ell_I(\Theta)]} \cdot \mathbb{E}\left[\frac{\Theta_i \Theta_j}{\Theta_i \vee \Theta_j}\right] = \frac{1}{\mathbb{E}[\ell_I(\Theta)]} \cdot \mathbb{E}[\Theta_i \wedge \Theta_j] \\ &= \frac{1}{\mathbb{E}[\ell_I(\Theta)]} \cdot \mathbb{E}[\Theta_i + \Theta_j - \Theta_i \vee \Theta_j] = \frac{\tau}{\tau_I} \cdot \left(\frac{2}{\tau} - \frac{\tau_I}{\tau}\right) = \frac{2}{\tau_I} - 1, \end{aligned}$$

we obtain that

$$\mathbb{E}[(\Theta_i^I)^2] \geq [\mathbb{E}(\Theta_i^I)]^2 = \frac{1}{\tau_I^2} \geq \frac{2}{\tau_I} - 1 = \mathbb{E}[\Theta_i^I \cdot \Theta_j^I] \quad (37)$$

and, analogously,

$$\mathbb{E}[(\Theta_j^I)^2] \geq \mathbb{E}[\Theta_i^I \cdot \Theta_j^I]. \quad (38)$$

Note that equality in Eq. (37) and in Equation (38) holds if and only if we have  $\mathbb{E}[(\Theta_i^I)^2] = \mathbb{E}[\Theta_i^I] = 1$  and  $\mathbb{E}[(\Theta_j^I)^2] = \mathbb{E}[\Theta_j^I] = 1$ , respectively. Thus, for  $\tau_I > 1$ , there exists a unique minimizer which can be easily computed:

$$\begin{aligned} v_i^* &= \frac{\mathbb{E}[(\Theta_j^I)^2] - \mathbb{E}[\Theta_i^I \Theta_j^I]}{\mathbb{E}[(\Theta_i^I)^2] - 2\mathbb{E}[\Theta_i^I \Theta_j^I] + \mathbb{E}[(\Theta_j^I)^2]} = \frac{\mathbb{E}[(\Theta_j^I)^2] - \mathbb{E}[\Theta_i^I \Theta_j^I]}{\mathbb{E}[(\Theta_i^I - \Theta_j^I)^2]} \\ \text{and } v_j^* &= 1 - v_i^* = \frac{\mathbb{E}[(\Theta_i^I)^2] - \mathbb{E}[\Theta_i^I \Theta_j^I]}{\mathbb{E}[(\Theta_i^I - \Theta_j^I)^2]}. \end{aligned}$$

By Eq. (37) and (38), we obtain  $v_i^*, v_j^* > 0$ , i.e.  $\mathbf{v}_{\{i,j\}}^* \in \partial B_1^+(\mathbf{0})$ . Some further calculations lead to the minimal asymptotic variance  $\tau_I^{-3}(\tau_I^2 \cdot f^{(I)}(\mathbf{v}_I) - 1)$  with

$$f^{(I)}(\mathbf{v}_I^*) = \frac{\mathbb{E}[(\Theta_i^I)^2] \cdot \mathbb{E}[(\Theta_j^I)^2] - (\mathbb{E}[\Theta_i^I \Theta_j^I])^2}{\mathbb{E}[(\Theta_i^I - \Theta_j^I)^2]}.$$

### 4.1.3 Asymptotic normality of the plug-in estimator

We use the functional central limit theorems Cor. 6 and Thm. 10 to show that estimators with minimal asymptotic variance can be obtained by plugging-in consistent estimators of the optimal weight vectors.

The variance minimizing vector  $\mathbf{v}_I^*$  can be estimated by taking the minimum of the empirical counterpart of  $f^{(I)}$ :

$$\widehat{\mathbf{v}}_{n,u,I}^* = \operatorname{argmin}_{\mathbf{v}_I \in \partial B_1^+(\mathbf{0})} \frac{\widehat{M}_{n,u,I}(\mathbf{v}_I, 2)}{\widehat{P}_{n,u,I}}.$$

It is important to notice that, being constructed as the empirical second moment of a convex combination,  $\mathbf{v}_i \mapsto \widehat{M}_{n,u,I}(\mathbf{v}_I, 2)/\widehat{P}_{n,u,I}$  is a positive semi-definite quadratic form. Thus, the evaluation of  $\widehat{M}_{n,u,I}(\mathbf{v}_I, 2)/\widehat{P}_{n,u,I}$  for a

finite number of vectors  $\mathbf{v}_I$  is sufficient to reconstruct the whole function  $\{\widehat{M}_{n,u,I}(\mathbf{v}_I, 2)/\widehat{P}_{n,u,I} : \mathbf{v}_I \in \partial B_1^+(\mathbf{0})\}$  and, thus, to calculate its minimizer.

Even though  $f^{(I)}$  is conditionally positive definite, i.e. the theoretical minimizer  $\mathbf{v}_I^*$  is unique, definiteness of the empirical counterpart  $\mathbf{v}_I \mapsto \widehat{M}_{n,u,I}(\mathbf{v}_I, 2)/\widehat{P}_{n,u,I}$  is not guaranteed for a finite sample, i.e. its minimizer  $\widehat{\mathbf{v}}_{n,u,I}^*$  might no longer be unique. In such a case, we just choose an arbitrary minimizer. Moreover, if  $\ell_I(\mathbf{X}_I^*) \leq u$  then  $\widehat{\mathbf{v}}_{n,u,I}^*$  is not well defined and can be an arbitrary vector of  $\partial B_1^+(\mathbf{0})$  for all  $l \in \{1, \dots, n\}$ . We first show consistency of this estimator.

**Proposition 15.** *Let  $\mathbf{X}_l^* = (X_{l1}^*, \dots, X_{ld}^*)^\top$ ,  $l \in \mathbb{N}$ , be independent copies of a  $d$ -dimensional random vector  $\mathbf{X}^*$  satisfying the assumptions of Prop. 13 for all non-empty set  $I \subset \{1, \dots, d\}$ . Furthermore, let  $\{u_n\}_{n \in \mathbb{N}}$  be a sequence such that  $u_n \rightarrow \infty$  and  $n/a^*(u_n) \rightarrow \infty$  as  $n \rightarrow \infty$  and assume that Eq. (27) holds for  $p = 0, 2$  and  $\delta = 0$ . Then,*

$$\widehat{\mathbf{v}}_{n,u_n,I}^* \xrightarrow{p} \mathbf{v}_I^*$$

for any sequence of minimizers  $\widehat{\mathbf{v}}_{n,u_n,I}^*$  and any  $I \subset \{1, \dots, d\}$ .

Furthermore,  $\widehat{M}_{n,u_n,I}(\widehat{\mathbf{v}}_{n,u_n,I}^*, 2) \rightarrow_p f^{(I)}(\mathbf{v}_I^*)$ .

*Proof.* We observe that, non-empty  $I \subset \{1, \dots, d\}$ , we have  $f^{(I)}(\mathbf{v}_I) = c(\mathbf{v}_I, \mathbf{1}_I, 2)$  for all  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$  and, thus, Cor. 6 yields

$$\sup_{\mathbf{v}_I \in \partial B_1^+(\mathbf{0})} |\widehat{M}_{n,u_n,I}(\mathbf{v}_I, 2) - f^{(I)}(\mathbf{v}_I)| \xrightarrow{p} 0. \quad (39)$$

As the continuous function  $f^{(I)}$  attains its unique minimum on  $\partial B_1^+(\mathbf{0})$  at  $\mathbf{v}_I^*$  (see Prop. 13), for every  $\varepsilon > 0$ , there exists some  $\eta = \eta(\varepsilon) > 0$  such that  $f^{(I)}(\mathbf{v}) - f^{(I)}(\mathbf{v}_I^*) > \eta$  for all  $\mathbf{v}$  with  $\|\mathbf{v} - \mathbf{v}_I^*\| > \varepsilon$ . Consequently, if  $\|\widehat{\mathbf{v}}_{n,u_n,I}^* - \mathbf{v}_I^*\| > \varepsilon$  then we have that  $f^{(I)}(\widehat{\mathbf{v}}_{n,u_n,I}^*) - f^{(I)}(\mathbf{v}_I^*) > \eta$ , which further implies

$$f^{(I)}(\widehat{\mathbf{v}}_{n,u_n,I}^*) - \widehat{M}_{n,u_n,I}(\widehat{\mathbf{v}}_{n,u_n,I}^*, 2) + \widehat{M}_{n,u_n,I}(\mathbf{v}_I^*, 2) - f^{(I)}(\mathbf{v}_I^*) > \eta$$

as  $\widehat{M}_{n,u_n,I}(\widehat{\mathbf{v}}_{n,u_n,I}^*, 2) \leq \widehat{M}_{n,u_n,I}(\mathbf{v}_I^*, 2)$  by definition of  $\widehat{\mathbf{v}}_{n,u_n,I}^*$ . In particular,

$$\sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0})} |\widehat{M}_{n,u_n,I}(\mathbf{v}, 2) - f^{(I)}(\mathbf{v})| > \eta/2.$$

Noticing that  $\widehat{\mathbf{v}}_{n,u_n,I}^*$  is well defined only if  $\bigvee_{1 \leq l \leq n} \ell_l(\mathbf{X}_l^*) > u_n$ , we obtain

$$\mathbb{P}(\|\widehat{\mathbf{v}}_{n,u_n,I}^* - \mathbf{v}_I^*\| > \varepsilon) \leq \mathbb{P}(\ell_I(\mathbf{X}^*) \leq u_n)^n + \mathbb{P}\left(\sup_{\mathbf{v}_I \in \partial B_1^+(\mathbf{0})} |\widehat{M}_{n,u_n}(\mathbf{v}_I, 2) - f^{(I)}(\mathbf{v}_I)| > \frac{\eta}{2}\right),$$

which tends to 0 as  $n \rightarrow \infty$  by Eq. (39).

The statement that

$$\widehat{M}_{n,u_n,I}(\widehat{\mathbf{v}}_{n,u_n,I}^*, 2) \rightarrow_p f^{(I)}(\mathbf{v}_I^*)$$

follows from the facts that  $f^{(I)}(\widehat{\mathbf{v}}_{n,u_n,I}^*) - \widehat{M}_{n,u_n,I}(\widehat{\mathbf{v}}_{n,u_n,I}^*, 2) \rightarrow_p 0$  by Eq. (39) and  $f^{(I)}(\widehat{\mathbf{v}}_{n,u_n,I}^*) - f^{(I)}(\mathbf{v}_I^*) \rightarrow_p 0$  due to  $\widehat{\mathbf{v}}_{n,u_n,I}^* \rightarrow_p \mathbf{v}_I^*$  and the continuity of  $f^{(I)}$ .

The second part can be shown analogously using the results from Thm. 10 instead of Cor. 16.  $\square$

Now, we are ready establish the desired limit theorem for the plug-in estimator, the moment-based estimator for the known margin case,

$$\frac{1}{\widehat{\tau}_{I,n}^{MK}} = \frac{\widehat{M}_{n,u_n,I}(\widehat{\mathbf{v}}_{n,u_n,I}^*)}{\widehat{P}_{n,u_n,I}},$$

by combining the results of Prop. 15 with Cor. 6.

**Corollary 16.** *Let  $\mathbf{X}_l^* = (X_{l1}^*, \dots, X_{ld}^*)^\top$ ,  $l \in \mathbb{N}$ , be independent copies of a  $d$ -dimensional random vector  $\mathbf{X}^*$  satisfying the assumptions of Prop. 13 for all non-empty  $I \subset \{1, \dots, d\}$ . Furthermore, let  $\{u_n\}_{n \in \mathbb{N}}$  be a sequence such that  $u_n \rightarrow \infty$  and  $n/a^*(u_n) \rightarrow \infty$  as  $n \rightarrow \infty$  and assume that Eq. (27) holds for  $p = 0, 1, 2$  and  $\delta = 0$ . Then, for all non-empty  $I \subset \{1, \dots, d\}$ ,*

$$\sqrt{\frac{n}{a^*(u_n)}} \left( \frac{1}{\widehat{\tau}_{I,n}^{MK}} - \frac{1}{\tau_I} \right) \xrightarrow{d} \mathcal{N}\left(0, \tau_I^{-1} \text{Var}(\mathbf{v}_I^{*\top} \boldsymbol{\Theta}^I)\right).$$



## 4.2 Unknown margins

Similarly, in case that the marginal distributions of  $X_1, \dots, X_d$  are unknown, i.e. Eq. (3) is no longer satisfied and the indices  $\alpha_1, \dots, \alpha_d$  of regular variation may be different from 1, we need to replace the deterministic threshold both in the benchmark estimator and in the moment-based estimator by a random one and include Hill type estimators as described in Subsection 3.2.

### 4.2.1 The benchmark and the moment-based estimator

In case of the benchmark defined in Eq. (20), the transformations above lead to the estimator

$$\frac{\tilde{P}_{n,k,I}^{(conv)}(\mathbf{v}_I)}{\tilde{P}_{n,k,I}} = \frac{n^{-1} \sum_{l=1}^n \mathbb{1}\{\mathbf{v}_I^\top(\mathbf{X}_l/\mathbf{X}_{k:n})^{\hat{\alpha}} > 1\}}{n^{-1} \sum_{l=1}^n \mathbb{1}\{\ell_I(\mathbf{X}_l/\mathbf{X}_{k:n}) > 1\}}, \quad \mathbf{v}_I \in \partial B_1^+(\mathbf{0}).$$

Of particular interest is the case that  $\mathbf{v}_I$  is a standard basis vector, i.e.  $\mathbf{v}_I = \mathbf{1}_{\{i\}}$  for some  $i \in I$ , since then the numerator satisfies  $\tilde{P}_{n,k,I}^{(conv)}(\mathbf{1}_{\{i\}}) = k/n$  a.s., i.e. it possesses variance zero. Thus, the estimator simplifies to

$$\frac{1}{\frac{n}{k} \tilde{P}_{n,k,I}} = \frac{1}{\hat{L}_{k,n}(\mathbf{1}_I)}$$

where  $\hat{L}_{k,n}$  is the well-known non-parametric estimator for the stable-tail dependence function [cf. Drees and Huang, 1998]. Provided that the bias is negligible and that the function  $\mathbf{s} \mapsto \mathbb{E}[\|\mathbf{s} \circ \Theta^I\|]$  is differentiable at  $\mathbf{s} = \mathbf{1}_I$ , this estimator is asymptotically normal:

$$\sqrt{k_n} \left( \frac{1}{\hat{L}_{k,n}(\mathbf{1}_I)} - \frac{1}{\tau_I} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\sigma_L^2}{\tau_I^4} \right)$$

where the asymptotic variance of the stable dependence function estimator  $\hat{L}_{k,n}(\mathbf{1}_I)$  is provided in Chapter 7.4 of De Haan and Ferreira [2006] by the expression

$$\sigma_L^2 = \tau_I^3 \sum_{i \in I} \sum_{j \in I} \frac{\partial}{\partial s_i} \mathbb{E}[\|\mathbf{s} \circ \Theta^I\|] \Big|_{\mathbf{s}=\mathbf{1}_I} \frac{\partial}{\partial s_j} \mathbb{E}[\|\mathbf{s} \circ \Theta^I\|] \Big|_{\mathbf{s}=\mathbf{1}_I} \mathbb{E}[\Theta_i^I \wedge \Theta_j^I] - \tau_I.$$

Using that  $\mathbb{E}[\Theta_i^I \wedge \Theta_j^I] \leq \mathbb{E}[\Theta_i^*] = 1/\tau_I$  for all  $i \in I$  and  $\sum_{i \in I} \frac{\partial}{\partial s_i} \mathbb{E}[\|\mathbf{s} \circ \Theta^I\|] \Big|_{\mathbf{s}=\mathbf{1}_I} = 1$  by Euler's homogeneous function theorem due to the positive 1-homogeneity of the function  $\mathbf{s} \mapsto \mathbb{E}[\|\mathbf{s} \circ \Theta^I\|]$ , we obtain that

$$\frac{\sigma_L^2}{\tau_I^4} \leq \frac{\tau_I^2 - \tau_I}{\tau_I^4} = \frac{\tau_I - 1}{\tau_I^3},$$

i.e. the asymptotic variance of  $1/\hat{L}_{k,n}(\mathbf{1}_I)$  is always smaller than the asymptotic variance of the benchmark estimators for known margins,  $1/\hat{\tau}_{I,n}(\mathbf{v}_I)$ . In particular, the estimator becomes more efficient by using order statistics even if the marginal distributions are known.

For the moment-based estimators, as already discussed in Subsection 3.2, the marginal transformations yield estimators of the type  $\tilde{M}_{n,k,I}(\mathbf{v}_I, 1)/\tilde{P}_{n,k,I}$ ,  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$ . Provided that the assumptions of Thm. 10 hold, we obtain the asymptotic normality

$$\sqrt{k_n} \left( \frac{\tilde{M}_{n,k,I}(\mathbf{v}_I, 1)}{\tilde{P}_{n,k,I}} - \frac{1}{\tau_I} \right) \xrightarrow{d} \mathcal{N}(0, \tilde{V}(\mathbf{v}_I)), \quad n \rightarrow \infty,$$

with  $\tilde{V}(\mathbf{v}_I, 1)$  is the variance of the limiting process given in Thm. 10.

### 4.2.2 Variance minimization

Analogously to Section 4.1, we aim at choosing the estimator that provides the minimal asymptotic variance, i.e. use the weight vector  $\tilde{\mathbf{v}}_I \in \partial B_1^+(\mathbf{0})$  that minimizes  $\tilde{V}(\cdot)$ , respectively. To this end, we first exploit the results given in Rem. 11; By linearity of this limiting process  $\tilde{V}(\mathbf{v}_I)$  is a quadratic form  $\mathbf{v}_I^\top \tilde{\Sigma}_I \mathbf{v}_I$  for the symmetric positive semidefinite covariance matrix

$$\tilde{V}_I = \text{Var} \left( \left( \tilde{G}(\mathbf{1}_j, \mathbf{1}_I, \mathbf{1}, 1) - \sum_{i \in I} [c_{s_i}(\mathbf{1}_j, \mathbf{1}_I, \mathbf{1}, 1) G^0(\mathbf{1}_{\{i\}}) + c_{\beta_i}(\mathbf{1}_j, \mathbf{1}_I, \mathbf{1}, 1) \alpha_i \tilde{H}_i] \right)_{j \in I} \right).$$

We obtain an analogous result to Prop. 13 and the analogous proof is omitted.

**Proposition 17.** Let  $\Theta$  be a spectral vector. Assume that all the partial derivatives

$$\begin{aligned} c_{s_i}(\mathbf{v}, \mathbf{s}, \beta, 1) &= \frac{\partial}{\partial s_i} c(\mathbf{v}, \mathbf{s}, \beta, 1), \quad i \in \{1, \dots, d\}, \\ c_{\beta_i}(\mathbf{v}, \mathbf{s}, \beta, 1) &= \frac{\partial}{\partial \beta} c(\mathbf{v}, \mathbf{s}, \beta, 1), \end{aligned}$$

exist and are continuous on  $\partial B_1^+(\mathbf{0}) \times A'_\delta$  and that all the partial derivatives of the function  $\mathbf{s} \mapsto \mathbb{E}(\|\mathbf{s} \circ \Theta^I\|)$  exist and are continuous in  $A_\delta$ .

Then, for all non-empty  $I \subset \{1, \dots, d\}$ : If the matrix  $\tilde{\mathbf{V}}_I$  is conditionally positive definite, then the function  $\mathbf{v}_I \mapsto \tilde{V}(\mathbf{v}_I)$  attains its minimum on  $\partial B_1^+(\mathbf{0})$  at some unique  $\tilde{\mathbf{v}}_I \in \partial B_1^+(\mathbf{0})$ .

### 4.2.3 Asymptotic normality of the plug-in estimator

Note that, in contrast to  $f^{(I)}$ , there is no direct empirical counterpart of  $\tilde{V}(\cdot)$ . To enable analytical minimization of that function, we consider an estimator that retains the quadratic form of  $\tilde{V}(\cdot)$  by estimating consistently the symmetric conditionally definite matrix  $\tilde{\mathbf{V}}_I$ . We consider a quadratic form  $\tilde{V}_{n,I}(\mathbf{v}_I) = \mathbf{v}_I^\top \tilde{\mathbf{V}}_{n,I} \mathbf{v}_I$  for a sequence of random symmetric matrices  $(\tilde{\mathbf{V}}_{n,I})_{n \in \mathbb{N}}$ . Then we define

$$\tilde{\mathbf{v}}_{n,I} = \operatorname{argmin}_{\mathbf{v}_I \in \partial B_1^+(\mathbf{0})} \tilde{V}_{n,I}(\mathbf{v}_I).$$

Here, without further conditions on the underlying estimators, no kind of (semi-)definiteness is guaranteed for the matrices  $\tilde{\mathbf{V}}_{n,I}$ . However, being a continuous function,  $\mathbf{v}_I \mapsto \tilde{V}_{n,I}(\mathbf{v}_I)$  attains its minimum on the compact set  $\partial B_1^+(\mathbf{0})$  at least once. Here, an arbitrary minimizer might be chosen for  $\tilde{\mathbf{v}}_{n,I}$  provided that it is not unique.

Analogously to Prop. 15, consistency of the estimator can be shown.

**Proposition 18.** Let  $\Theta$  be a spectral vector satisfying the assumptions of Prop. 17. Then, for all non-empty  $I \subset \{1, \dots, d\}$ , it holds: If  $\tilde{\mathbf{V}}_{I,n} \rightarrow_p \tilde{\mathbf{V}}_I$ , then we have that  $\tilde{\mathbf{v}}_{n,I} \rightarrow_p \tilde{\mathbf{v}}_I$  for any sequence of minimizers  $\tilde{\mathbf{v}}_{n,I}$ . Furthermore,  $\tilde{V}_{n,I}(\tilde{\mathbf{v}}_{n,I}) \rightarrow_p \tilde{V}(\tilde{\mathbf{v}}_I)$ .

*Proof.* Recall that, by construction,  $\tilde{V}_{n,I}(\mathbf{v}_I) = \mathbf{v}_I^\top \tilde{\mathbf{V}}_{n,I} \mathbf{v}_I$  while  $\tilde{V}(\mathbf{v}_I) = \mathbf{v}_I^\top \tilde{\mathbf{V}}_I \mathbf{v}_I$  for a conditionally positive definite matrix  $\tilde{\mathbf{V}}_I$ . The assumptions on the consistency of all the estimators involved in  $\tilde{\mathbf{V}}_{n,I}$  imply that  $\|\tilde{\mathbf{V}}_{n,I} - \tilde{\mathbf{V}}_I\|_\infty \rightarrow_p 0$ . Consequently,

$$\sup_{\mathbf{v}_I \in \partial B_1^+(\mathbf{0})} |\tilde{V}_{n,I}(\mathbf{v}_I) - \tilde{V}(\mathbf{v}_I)| \leq d^2 \|\tilde{\mathbf{V}}_{n,I} - \tilde{\mathbf{V}}_I\|_\infty \rightarrow_p 0,$$

which is the analogue to Eq. (39). The remainder of the proof runs analogously to the proof of Prop. 15.  $\square$

Combining the results of Prop. 18 and Thm. 10, we obtain the desired limit theorem for the plug-in estimator, the moment based estimator for the case of unknown margins

$$\frac{1}{\hat{\tau}_{I,n}^{MU}} = \frac{\tilde{M}_{n,k,I}(\tilde{\mathbf{v}}_{n,I}, 1)}{\tilde{P}_{n,k,I}}.$$

**Corollary 19.** Let  $\mathbf{X}_l, l \in \mathbb{N}$ , be independent copies of a regularly varying  $[0, \infty)^d$ -valued random vector  $\mathbf{X}$  satisfying the assumptions of Prop. 18 for all non-empty  $I \subset \{1, \dots, d\}$  and the assumptions of Thm. 10 for  $K = \{1\}$ . Then, the plug-in estimator satisfies

$$\sqrt{k} \left( \frac{1}{\hat{\tau}_{I,n}^{MU}} - \frac{1}{\tau_I} \right) \xrightarrow{d} \mathcal{N}(0, \tilde{V}(\tilde{\mathbf{v}}_I)), \quad n \rightarrow \infty.$$

In order to get a consistent estimator  $\tilde{\mathbf{V}}_{n,I}$  of  $\tilde{\mathbf{V}}_I$  we use the  $m$ -out-of- $n$  bootstrap procedure. From Thm. 10 it is very natural to consider the covariance matrix of the vector

$$\left( \sqrt{k_n} \left( \frac{\tilde{M}_{n,k_n,I}(\mathbf{1}_{\{i\}}, 1)}{\tilde{P}_{n,k_n,I}} - c(\mathbf{1}_{\{i\}}, \mathbf{1}_I, \mathbf{1}, 1) \right) \right)_{1 \leq i \leq d}$$

as an approximation of  $\tilde{\mathbf{V}}_I$ . In order to estimate each entry of this covariance matrix, we consider  $N_m$  subsamples  $(\mathbf{X}_1^{(j)}, \dots, \mathbf{X}_m^{(j)})$ ,  $1 \leq j \leq N_m$ , of the original sample  $(\mathbf{X}_1, \dots, \mathbf{X}_n)$  drawn uniformly without replacement. We denote  $\tilde{M}_{m,k_m,I}^{(j)}$  and  $\tilde{P}_{m,k_m,I}^{(j)}$  the statistics of interest using the subsample  $(\mathbf{X}_1^{(j)}, \dots, \mathbf{X}_m^{(j)})$ . Then we derive a bootstrapped distribution for any pair  $i, i' \in I$  such as

$$L_{n,i,i'}^{(m)}(x) = \frac{1}{N_m} \sum_{j=1}^{N_m} \mathbb{1} \left\{ \sqrt{k_m} \left( \frac{\tilde{M}_{m,k_m,I}^{(j)}(\mathbf{1}_{\{i\}}, 1) + \tilde{M}_{m,k_m,I}^{(j)}(\mathbf{1}_{\{i'\}}, 1)}{2\tilde{P}_{m,k_m,I}^{(j)}} - \frac{\tilde{M}_{n,k_n,I}(\mathbf{1}_{\{i\}}, 1) + \tilde{M}_{n,k_n,I}(\mathbf{1}_{\{i'\}}, 1)}{2\tilde{P}_{n,k_n,I}} \right) \leq x \right\}.$$

We then calculate the first and third quartiles  $q_{1,i,i'}^{(m)}$  and  $q_{3,i,i'}^{(m)}$  of the bootstrapped distribution and consider the standardized interquartile statistics

$$\tilde{\sigma}_{i,i'}^{(m)} = \frac{q_{3,i,i'}^{(m)} - q_{1,i,i'}^{(m)}}{q_3 - q_1}$$

where  $q_1$  and  $q_3$  are the first and third quartiles of a standard Gaussian random variable. Finally, we define the entries of our estimator  $\tilde{\mathbf{V}}_{n,I}^{(m)}$  such as

$$\tilde{\mathbf{V}}_{n,I,i,i'}^{(m)} = \begin{cases} \tilde{\sigma}_{i,i}^{(m)2} & \text{if } i = i', \\ 2\tilde{\sigma}_{i,i'}^{(m)2} - (\tilde{\sigma}_{i,i}^{(m)2} + \tilde{\sigma}_{i',i'}^{(m)2})/2 & \text{else.} \end{cases}$$

Notice that our estimation procedure also depends on the number of subsamples  $N_m$  that we do not fix at  $\binom{n}{m}$  for computational efficiency. We suppress this dependence for clarity.

**Proposition 20.** *Let  $\mathbf{X}_l$ ,  $l \in \mathbb{N}$ , be independent copies of a regularly varying  $[0, \infty)^d$ -valued random vector  $\mathbf{X}$  satisfying the assumptions of Cor. 19. Then choosing  $m_n$  such that  $m_n \rightarrow \infty$ ,  $m_n/n \rightarrow 0$  and  $k_n/k_{m_n} \rightarrow \infty$  and  $N_m \rightarrow \infty$  we have  $\tilde{\mathbf{V}}_{I,n}^{(m)} \rightarrow_p \tilde{\mathbf{V}}_I$  as  $n \rightarrow \infty$  for all  $I \subset \{1, \dots, d\}$ .*

*Proof.* We apply Theorem 2.1 (ii) of Politis and Romano [1994] to the statistics

$$\left( \sqrt{k_n} \left( \frac{\tilde{M}_{n,k_n,I}(\mathbf{1}_{\{i\}}, 1) + \tilde{M}_{n,k_n,I}(\mathbf{1}_{\{i'\}}, 1)}{2\tilde{P}_{n,k_n,I}} - \frac{1}{2}(c(\mathbf{1}_{\{i\}}, \mathbf{1}_I, \mathbf{1}, 1) + c(\mathbf{1}_{\{i'\}}, \mathbf{1}_I, \mathbf{1}, 1)) \right) \right)$$

remarking that the limiting centered Gaussian distribution with variance

$$\tilde{\sigma}_{i,i'}^2 = \begin{cases} \tilde{\mathbf{V}}_{I,i,i'} & \text{if } i = i', \\ \tilde{\mathbf{V}}_{I,i,i'}/2 + \tilde{\mathbf{V}}_{I,i,i'}/4 + \tilde{\mathbf{V}}_{I,i,i'}/4 & \text{else,} \end{cases} \quad (40)$$

is continuous for all  $i, i' \in I$ . Then  $L_{n,i,i'}^{(m)}(x)$  converges in probability to the limiting distributions,  $L_{i,i'}(x)$  say, uniformly for all  $x \in \mathbb{R}$ . The uniform convergence ensures the convergence in probability of the quantiles  $q_{j,i,i'}^{(m)}$  towards  $\tilde{\sigma}_{i,i'} q_j$  for all  $i, i' \in I$  and  $j = 1, 3$ . Thus  $\tilde{\sigma}_{i,i'}^{(m)}$  converges in probability to  $\tilde{\sigma}_{i,i'}$  and  $\tilde{\mathbf{V}}_{I,n}^{(m)}$  to  $\tilde{\mathbf{V}}_I$  as well because of the expression (40) of the limiting variances  $\tilde{\sigma}_{i,i'}^2$ ,  $i, i' \in I$ .  $\square$

Note that, by construction, the estimated matrices  $\tilde{\mathbf{V}}_{I,n}^{(m)}$  might be indefinite, which impedes the minimization of the quadratic form  $\mathbf{v}_I \mapsto \tilde{\mathbf{V}}_{I,n}^{(m)}(\mathbf{v}_I)$  on  $\partial B_1^+(\mathbf{0})$ . In practice, it turns out to define  $\tilde{\sigma}_{i,i'}^{(m)}$  as the sample variance of the bootstrap sample. Provided that the same subsamples  $(\mathbf{X}_1^{(j)}, \dots, \mathbf{X}_m^{(j)})$ ,  $1 \leq j \leq N_m$ , are used for each pair  $(i, i') \in \{1, \dots, d\}^2$ , this guarantees that the corresponding matrix  $\tilde{\mathbf{V}}_{n,I}^{(m)}$  is positive semi-definite. Numerical experiments illustrate that the resulting estimator performs as expected (see Section 5.1) in practice. However, a rigorous proof of weak consistency of  $\tilde{\mathbf{V}}_{n,I}^{(m)}$  would require convergence of second order moments of the bootstrapped distribution to the corresponding moments of the Gaussian limits, which we cannot show without imposing strong additional assumptions.

## 5 Numerical illustrations

### 5.1 Simulated example I: Hüsler-Reiss distribution

As a first simulated example, we consider a popular class of multivariate extreme value models, the Hüsler-Reiss distribution [Hüsler and Reiss, 1989] whose cumulative distribution is given by

$$\mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \exp \left( -\mathbb{E} \left[ \max_{i=1, \dots, d} \frac{1}{x_i} e^{W_i - \text{Var}(W_i/2)} \right] \right), \quad x_1, \dots, x_d > 0,$$

where  $(W_1, \dots, W_d)$  is a centered Gaussian random vector. By definition, the random vector  $\mathbf{X}$  is max-stable with unit Fréchet margins. Note that its distribution is uniquely determined by the variogram matrix

$$\Gamma = (\Gamma_{ij})_{1 \leq i, j \leq d}, \quad \Gamma_{ij} = \text{Var}(W_i - W_j).$$

#### 5.1.1 Bivariate Case

In the bivariate case, the extremal coefficient of the vector  $(X_1, X_2)$  is of the form

$$\tau = \tau_{\{1,2\}} = 2\Phi(\sqrt{\Gamma_{12}}/2) \in [1, 2],$$

i.e. the model can interpolate between full dependence ( $\Gamma_{12} = 0$ ) and asymptotic independence ( $\Gamma_{12} \rightarrow \infty$ ).

We compare the four types of estimators for  $\tau$ , namely the benchmark estimators and the moment-based estimators, both with known and unknown margins as described in Sections 4.1 and 4.2, respectively. More precisely, in the case of known margins where we can work with standardized observations  $\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_n^*$  and use

- the benchmark estimator

$$\frac{1}{\widehat{\tau}_{I,n}^{BK}(\mathbf{v})} = \frac{\widehat{P}_{n,u,\{1,2\}}^{(\text{conv})}(\mathbf{v})}{P_{n,u,\{1,2\}}};$$

- the moment-based plug-in estimator

$$\frac{1}{\widehat{\tau}_{I,n}^{MK}} = \frac{\widehat{M}_{n,u_n,\{1,2\}}(\widehat{\mathbf{v}}_{n,u_n,\{1,2\}}^*)}{\widehat{P}_{n,u_n,\{1,2\}}}$$

based on the estimated “optimal” weights  $\widehat{\mathbf{v}}_{n,u_n,\{1,2\}}^*$ ;

- and, for comparison, the perfect moment-based estimator

$$\frac{1}{\widehat{\tau}_{I,n,\text{opt}}^{MK}} = \frac{\widehat{M}_{n,u_n,\{1,2\}}(\mathbf{v}^*)}{\widehat{P}_{n,u_n,\{1,2\}}}$$

with theoretically optimal weight  $\mathbf{v}^*$ .

Note that, for simulated data from the bivariate Hüsler–Reiss model, no standardization is required as the (unit Fréchet) marginal distributions are already identical and regularly varying with index  $\alpha = 1$ . From Example 14, it can be seen that the variance minimizing linear combination is unique whenever  $0 < \Gamma_{12} < \infty$  and is given by  $\mathbf{v}^* = (1/2, 1/2)$ . As discussed in Section 4.1, the asymptotic distribution of the benchmark estimator does not depend on the choice of  $\mathbf{v}_I \in \partial B_1^+(\mathbf{0})$ . Here we choose  $\widehat{\tau}_{I,n}^{BK} = \widehat{\tau}_{I,n}^{BK}(\mathbf{v}^*)$  as benchmark.

Even though the margins are known, it might sometimes be beneficial to work with the rank transformations designed for the case of unknown margins as discussed in Section 4.2. Thus, we also consider

- the “benchmark” estimator

$$\frac{1}{\widehat{\tau}_{I,n}^{BU}} = \frac{1}{\frac{n}{k} \widetilde{P}_{n,k,\{1,2\}}} = \frac{1}{\widehat{L}_{k,n}(\mathbf{1})};$$

- the moment-based plug-in estimator

$$\frac{1}{\widehat{\tau}_{I,n}^{MU}} = \frac{\widetilde{M}_{n,k,\{1,2\}}(\widetilde{\mathbf{v}}_{n,\{1,2\}}, 1)}{\widetilde{P}_{n,k,\{1,2\}}}$$

based on estimated “optimal” weights  $\widetilde{\mathbf{v}}_{n,\{1,2\}}$ ;

$\Gamma_{12} = 0.1$  ( $\tau \approx 1.13$ , strong dependence)

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MU}$
bias	0.003	0.001	0.000	0.019	-0.001	0.001
std. deviation	0.030	0.007	0.007	0.019	0.012	0.012

 $\Gamma_{12} = 2$  ( $\tau \approx 1.52$ , moderate dependence)

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MU}$
bias	0.009	0.005	0.004	0.010	0.004	0.004
std. deviation	0.040	0.010	0.010	0.020	0.016	0.016

 $\Gamma_{12} = 10$  ( $\tau \approx 1.89$ , weak dependence)

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MU}$
bias	0.022	0.014	0.014	0.010	0.013	0.013
std. deviation	0.038	0.006	0.006	0.009	0.007	0.007

Table 1: Results for 500 simulations from the max-stable Hüsler–Reiss models in the three scenarios specified above with  $n = 5\,000$ ,  $k = 100$  and  $u_n = \Phi_1^{-1}(1 - k/n) = \Phi_1^{-1}(0.98)$ .

- and, again, the perfect moment-based plug-in estimator

$$\frac{1}{\widehat{\tau}_{I,n,\text{opt}}^{MU}} = \frac{\widetilde{M}_{n,k,\{1,2\}}(\widetilde{v}, 1)}{\widehat{P}_{n,k,\{1,2\}}}$$

with theoretically optimal weight  $\widetilde{v}$ .

From symmetry arguments, it can be easily seen that the “optimal” linear combination (if unique) is necessarily given by  $\widetilde{v} = (1/2, 1/2)$ .

We perform a simulation study to compare the estimators in a setting with finite sample sizes. Here, we consider three different scenarios:

- Scenario 1:  $\Gamma_{12} = 0.1$  (strong dependence;  $\tau \approx 1.13$ )
- Scenario 2:  $\Gamma_{12} = 2$  (medium dependence;  $\tau \approx 1.52$ )
- Scenario 3:  $\Gamma_{12} = 10$  (weak dependence;  $\tau \approx 1.89$ )

In each of these scenarios, we consider  $n = 5\,000$  independent observations of  $\mathbf{X} = (X_1, X_2)$  and apply the estimators with a threshold  $u_n = \Phi_1^{-1}(0.98)$  and  $k_n = 0.02 \cdot n$ , respectively. For the bootstrap procedure that is needed to estimate the optimal variance, we choose the subsample size  $m_n = 0.25 \cdot n$  and  $N_m = n$ . We determine the biases and standard deviations by 500 repetitions.

The results are displayed in Table 1. First, it can be seen that, at least in the cases of strong and moderate dependence, all the biases are much smaller than the standard deviations and are therefore negligible. Additional calculations of the asymptotic variances  $\text{AVar}^{BK}$ ,  $\text{AVar}_{\text{opt}}^{MK}$ ,  $\text{AVar}^{BU}$  and  $\text{AVar}_{\text{opt}}^{MU}$  according to the theoretical results given in Sections 4.1 and 4.2, respectively, show that the empirical standard deviations are very close to their theoretical counterparts even though the bootstrap procedure tends to slightly underestimate the true asymptotic standard deviation of the moment-based estimator in case of unknown margins (by roughly 10%). In particular, we obtain the expected ordering of the four methods if the optimal weights for the moment-based estimators are chosen:

$$\text{AVar}_{\text{opt}}^{MK} < \text{AVar}_{\text{opt}}^{MU} < \text{AVar}^{BU} < \text{AVar}^{BK}.$$

This ordering is kept if we replace the theoretically optimal weights by the estimated ones; in our simulation study the variances of the plug-in moment estimators are almost identical to the ones with optimal weights  $v^*$  and  $\widetilde{v}$ , which is well in line with the theoretical results. However, it has to be noted that taking non-optimal weights might result in significantly larger variances – in particular in the case of known margins. This is demonstrated in Figure 1 where the asymptotic variance is plotted as a function of the weight vector  $v$ . Compared to the case of known margins, the relative changes of the asymptotic variance for the moment-based estimator for unknown margins turn out to be rather small.

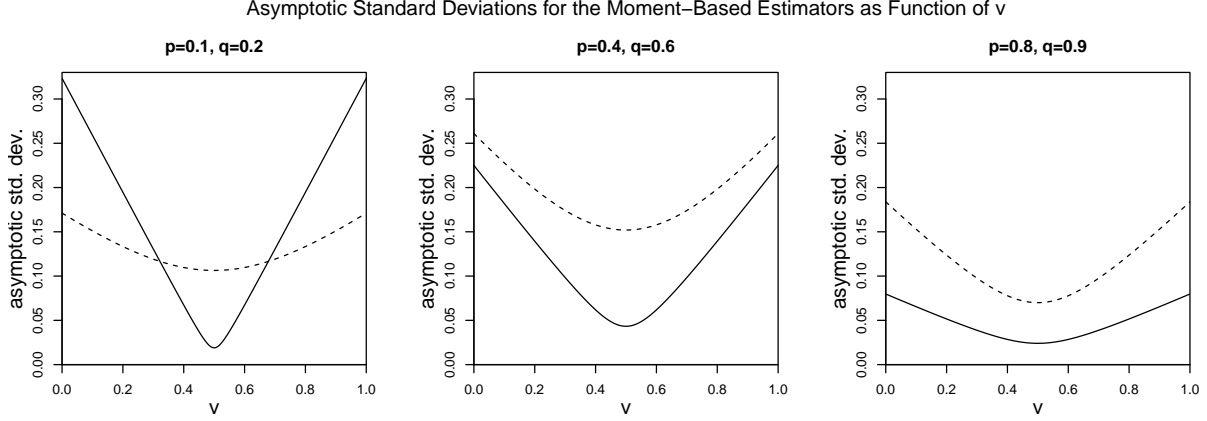


Figure 1: The asymptotic standard deviations  $\sqrt{V((v, 1 - v))}$  (solid lines) and  $\sqrt{\tilde{V}((v, 1 - v))}$  (dashed line) of the moment-based estimators with known and unknown margins, respectively, for each of the three scenarios reflecting strong, moderate and weak dependence (from left to right).

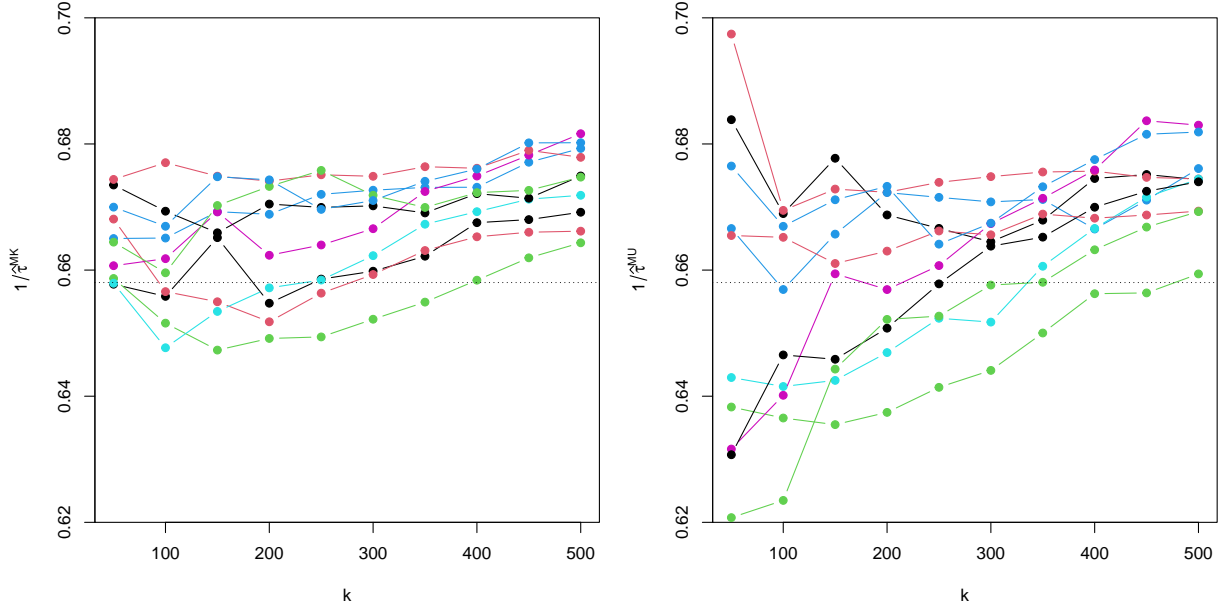


Figure 2: The results of the moment-based estimators,  $1/\hat{\tau}_{I,n}^{MK}$  (left) and  $1/\hat{\tau}_{I,n}^{MU}$  (right), as a function of  $k$  for ten realizations of the max-stable Hüsler-Reiss model with moderate dependence ( $\Gamma_{12} = 2$ ) and  $n = 5000$ .

We further analyze the sensitivity of the novel moment-based estimators with respect to the choice of  $k$ . To this end, the results of the estimators for ten realizations from the max-stable Hüsler-Reiss in the scenario of moderate dependence described above are displayed as a function of  $k$  are displayed in Figure 2. Here, we choose  $k \in \{50, 100, \dots, 500\}$  corresponding to the upper 1%, 2%, ..., 10% of the data. It can be seen that the results are rather stable with the estimator  $1/\hat{\tau}_{I,n}^{MU}$  being slightly more uncertain than  $1/\hat{\tau}_{I,n}^{MK}$ , which is in line with the results from Table 1.

As mentioned at the beginning of the section, there is a one-to-one correspondence between  $\tau = \tau_{\{1,2\}}$  and  $\Gamma_{12}$  via  $\tau = 2\Phi(\sqrt{\Gamma_{12}}/2)$  or, equivalently,  $\Gamma_{12} = [2\Phi^{-1}(\frac{\tau}{2})]^2$ . Thus, we obtain a plugin estimator  $\hat{\Gamma}_{12,n}^m$  from  $\hat{\tau}_{I,n}^m$  via

$$\hat{\Gamma}_{12,n}^m = \left[ 2\Phi^{-1}\left(\frac{\hat{\tau}_{I,n}^m}{2}\right) \right]^2, \quad m \in \{\text{BK}, \text{MK}, \text{BU}, \text{MU}\}.$$

By the delta method, the asymptotic variances are expected to be ordered in the same way as the variances for the estimators for  $\tau$ . To confirm this behaviour also in a finite sample setting, we apply the plugin estimators to the data

		$\hat{\Gamma}_{12,n}^{BK}$	$\hat{\Gamma}_{12,n}^{MK}$	$\hat{\Gamma}_{12,n}^{BU}$	$\hat{\Gamma}_{12,n}^{MU}$
$\Gamma_{12} = 0.1$	bias	0.007	-0.001	-0.016	0.004
	std. deviation	0.069	0.013	0.036	0.025
$\Gamma_{12} = 2$	bias	-0.001	-0.008	-0.016	-0.006
	std. deviation	0.889	0.207	0.381	0.323
$\Gamma_{12} = 10$	bias	NA	-2.123	-1.457	-1.983
	std. deviation	NA	0.760	1.413	0.956

Table 2: Results for the plugin estimators for  $\Gamma_{12}$  for 500 simulations from the max-stable Hüsler–Reiss models in the three scenarios specified above with  $n = 5\,000$ ,  $k = 100$  and  $u_n = \Phi_1^{-1}(1 - k/n) = \Phi_1^{-1}(0.98)$ .

		$\hat{\Gamma}_{12,n}^{BU}$	$\hat{\Gamma}_{12,n}^{MU}$
$\Gamma_{12} = 0.1$	bias	-0.011	0.005
	std. deviation	0.037	0.024
$\Gamma_{12} = 2$	bias	-0.071	0.081
	std. deviation	0.383	0.333
$\Gamma_{12} = 10$	bias	-0.720	0.317
	std. deviation	1.580	1.378

Table 3: Results for the plugin estimators for  $\Gamma_{12}$  for 500 simulations from the Hüsler–Reiss–Pareto models in the three scenarios specified above with  $n = 5\,000$ ,  $k = 100$  and  $u_n = \Phi_1^{-1}(1 - k/n) = \Phi_1^{-1}(0.98)$ .

from the simulation study above. The results are displayed in Table 2. Similarly to the study above, in the cases of strong and moderate dependence, the bias is negligible and the standard deviations show the expected ordering. In the case of weak dependence, the biases of the of the estimators for  $\Gamma_{12}$  are even more pronounced than the biases of the estimators for  $1/\tau$ . In particular, in some simulations, the benchmark estimator  $\hat{\tau}_{1,n}^{MK}$  even exceeds the theoretical maximum of 2 – consequently, the corresponding value for  $\Gamma_{12}$  cannot be calculated, i.e., we obtain invalid estimates. For the other three estimators, the standard deviations again exhibit the expected ordering.

Here, it is important to note that neither the benchmark nor the moment-based estimator are restricted to the setting of limiting multivariate extreme value distributions, but allow for estimation of the extremal coefficients, or, equivalently, the estimation of the corresponding dependence parameters, for any distribution from the max-domain of attraction. Thus, we duplicate the setting of the simulation study above simulating from the Hüsler–Reiss–Pareto distributions from the max-domain of attraction instead. As the marginal distributions are no longer unit Fréchet, but only asymptotically equivalent to that distribution, we do not consider the estimators for known margins, but only their variants for the case of unknown margins. The results are displayed in Table 3. While the results for the cases of strongly and moderately dependent data are very similar to the max-stable setting, the biases in the weakly dependent case are much smaller now, coming with the price of slightly larger standard deviations. This phenomenon can be explained by the fact that the angular part of a Hüsler–Reiss–Pareto random vector exactly follows the spectral measure, while the spectral measure still provides an approximation in the max-stable setting. In any case, the newly developed moment-based estimators provide smaller variances than the corresponding benchmark estimators.

### 5.1.2 Multivariate case

We repeat our study for the multivariate case. Here, we consider the case of a fully symmetric variogram matrix

$$\Gamma = \begin{pmatrix} 0 & 1 & \dots & \dots & 1 \\ 1 & 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 & 1 \\ 1 & \dots & \dots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{d \times d}.$$

We consider three different scenarios with dimensions  $d = 3$ ,  $d = 5$  and  $d = 8$ , respectively. For each of these scenarios, we compare the same six estimators as in the previous subsection. Again we can argue that, for symmetry reasons, the optimal weight vector, if unique, is necessarily given by  $\mathbf{v}^* = \tilde{\mathbf{v}} = (1/d, \dots, 1/d)$ . The results for  $n = 5\,000$ ,  $k = 100$  and 500 repetitions, i.e., the same setup as in the bivariate case, are displayed in Table 4.

$d = 3 :$

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MU}$
bias	0.006	0.003	0.002	0.010	0.002	0.001
std. deviation	0.038	0.010	0.010	0.024	0.020	0.020

$d = 5 :$

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MU}$
bias	0.008	0.003	0.002	0.010	0.002	0.003
std. deviation	0.034	0.009	0.009	0.022	0.021	0.022

$d = 8 :$

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$	$1/\widehat{\tau}_{I,n,\text{opt}}^{MU}$
bias	0.008	0.004	0.003	0.009	0.000	0.001
std. deviation	0.034	0.007	0.007	0.021	0.023	0.023

Table 4: Simulation results for the three multivariate scenarios specified above with  $n = 5\,000$ ,  $k = 100$  and  $u_n = \Phi_1^{-1}(1 - k/n) = \Phi_1^{-1}(0.98)$ .

The ordering of the estimators is similar to the bivariate setting with the exception that the asymptotic variances of the estimators with unknown margins are very close to each other:

$$\text{AVar}_{\text{opt}}^{MK} < \text{AVar}_{\text{opt}}^{MU} \approx \text{AVar}^{BU} < \text{AVar}^{BK}.$$

Again, the observed standard deviations correspond well to theoretical ones apart from the fact that the asymptotic standard deviation of the moment-based estimator for the case of unknown margins is again slightly underestimated (again by roughly 10 %) by the bootstrap procedure with  $m_n = 0.2 \cdot n$  and  $N_m = n$ .

## 5.2 Simulated example II: Multivariate Gaussian distribution

In order to demonstrate the performance of our estimators in the case of asymptotic independence, we also consider the case that  $\mathbf{X}$  follows a distribution with unit Fréchet margins, but a multivariate Gaussian copula. We consider the symmetric case with

$$\begin{pmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & \rho \\ \rho & \dots & \dots & \rho & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$$

being the covariance matrix of the underlying Gaussian copula. Thus, we obtain the theoretical extremal coefficient  $\tau_{\{1,\dots,d\}} = d$  whenever  $\rho < 1$ . Here, we choose  $\rho = 0.5$  and consider the 3-, 5- and 8-dimensional case, respectively. In the case of asymptotic independence, the matrices  $\mathbf{V}$  and  $\widetilde{\mathbf{V}}$  might be degenerate, which hampers the optimization of the asymptotic variance. Therefore, we just fix the weights to the default choice  $\mathbf{v}^* = \widetilde{\mathbf{v}} = (1/d, \dots, 1/d)$ . The results for  $n = 5\,000$ ,  $k = 100$  and 500 repetitions are displayed in Table 5.

Here, it is important to note that all the estimators exhibit a significant bias which can be explained by the fact that the true value is on the boundary of the estimation interval. However, it can be seen that the variance of the moment estimators is still smaller than the one of the benchmark estimators, which indicates that the moment estimators still show a reasonably good performance in the case of asymptotic independence.

## 5.3 Application to Precipitation Data in France

We apply our estimators to estimate extremal dependence among different weather stations in France recording heavy precipitation. More precisely, we use daily precipitation data for the fall seasons (September, October, November) in the years 1976 to 2015 at nine weather stations in France, grouped into three regions (northwest, south and northeast) consisting of three stations per region. This dataset has also been analyzed in [Buriticá and Naveau, 2021].

We start with an explanatory analysis. Based on block maxima over 15 days, we fit generalized extreme value distributions via maximum likelihood separately for each station. All the extreme value indices are estimated to be positive



$d = 3 :$ 

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$
bias	0.087	0.075	0.049	0.071
std. deviation	0.017	0.005	0.008	0.005

 $d = 5 :$ 

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$
bias	0.084	0.077	0.047	0.072
std. deviation	0.013	0.003	0.005	0.003

 $d = 8 :$ 

	$1/\widehat{\tau}_{I,n}^{BK}$	$1/\widehat{\tau}_{I,n}^{MK}$	$1/\widehat{\tau}_{I,n}^{BU}$	$1/\widehat{\tau}_{I,n}^{MU}$
bias	0.072	0.071	0.041	0.067
std. deviation	0.009	0.002	0.003	0.002

Table 5: Simulation results for the three asymptotically independent multivariate scenarios specified above with  $n = 5000$ ,  $k = 100$  and  $u_n = \Phi_1^{-1}(1 - k/n) = \Phi_1^{-1}(0.98)$  based on Gaussian copulas with  $\rho = 0.5$ .

region	based on different tail indices	based on single tail index
northwest	1.89 (0.020)	1.85 (0.018)
south	1.98 (0.019)	1.89 (0.018)
northeast	1.95 (0.020)	1.95 (0.019)

Table 6: Estimated extremal coefficients for extreme precipitation within each region. The estimated standard deviations via bootstrap are given in brackets.

(between 0.01 and 0.60). Even though, due to the rather small number of blocks, the positivity of the estimates is not significant, these results are in line with Bernard et al. [2013] who analyzed heavy precipitation in France for the fall season based on a similar data set. Furthermore, we empirically estimate the pairwise tail dependence coefficients for each pair of stations. The table estimates that there is moderate extremal dependence within each region, while extremal dependence across different regions is much weaker (though not vanishing), see Appendix E for details. Overall, the explanatory analysis suggests that the data are heavy-tailed and (at least within each region) asymptotically dependent. Therefore, it is reasonable to assume that they are covered by the framework of regular variation investigated in this work.

In order to avoid marginal standardization we make use of our flexible moment estimator for unknown margins  $\widehat{\tau}_{I,n}^{MU}$ . While this estimator allows for varying tail behaviour in the different components of the random vector  $\mathbf{X}$ , in particular for different tail indices, we also consider a modified version  $\widetilde{\tau}_{I,n}^{MU}$  based on the assumption of a single tail index:

$$\widetilde{\tau}_{I,n}^{MU} := \frac{n^{-1} \sum_{l=1}^n \left( \mathbf{v}_I^\top \frac{(\mathbf{X}_l / \mathbf{X}_{k:n})^{\widehat{\alpha}}}{\ell_I((\mathbf{X}_l / \mathbf{X}_{k:n})^{\widehat{\alpha}})} \right)^p \mathbb{1}\{\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}}{n^{-1} \sum_{l=1}^n \mathbb{1}\{\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}},$$

using a unique tail index estimator over each region  $I$

$$\frac{1}{\widehat{\alpha}_{n,k}} = \frac{n^{-1} \sum_{l=1}^n \log(\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n})) \mathbb{1}\{\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}}{n^{-1} \sum_{l=1}^n \mathbb{1}\{\ell_I(\mathbf{X}_l / \mathbf{X}_{k:n}) > 1\}}.$$

We first consider each region separately, i.e. we consider three different 3-dimensional vectors  $\mathbf{X}^{(northwest)}$ ,  $\mathbf{X}^{(south)}$  and  $\mathbf{X}^{(northeast)}$ . In order to avoid effects of temporal clustering, we consider block maxima over 5 days leading to a (nearly uncorrelated) time series of 720 observations at each station. For both estimators, we choose  $k = 40$ . The results for the estimated extremal coefficients within each region are displayed in Table 6.

It can be seen that there is moderate dependence within each region with extremal indices between 1.85 and 2 given a theoretical range of  $[1, 3]$ . This is in line with the results of the explanatory analysis in Appendix E. Interestingly, for both regions in the north, there are only little differences between the two estimators, while there are larger differences in the south. There the use of a single tail index would yield spurious dependence among the stations and reduce artificially the estimated tail dependence coefficient. The estimated standard deviations of both estimators are similar. Therefore, the more general estimator  $\widehat{\tau}_{I,n}^{MU}$  appears to be preferable in this context. For this estimator, the effect of the choice of  $k$  is also further analyzed in Appendix E.

Furthermore, we consider extremal dependence across the different regions. To this end, we consider the vector  $\mathbf{Y} = (\max\{\mathbf{X}^{(northwest)}\}, \max\{\mathbf{X}^{(south)}\}, \max\{\mathbf{X}^{(northeast)}\}) \in \mathbb{R}^3$  and estimate the extremal coefficient using the same estimators as above. This results in estimates of 2.62 (0.007) and 2.60 (0.007), respectively, showing that there is only little dependence across the different regions as also suggested by the results of the explanatory analysis in Appendix E.

## 6 Conclusion

We show some benefits of estimating extremal properties with moments of convex combination of spectral components. An analysis of the asymptotic variances yields a natural strategy consisting in plugging in the asymptotically optimal weights. This methodology is shown to efficiently reduce the variance when estimating the extremal coefficient by first order moments. As demonstrated in the case of Hüsler–Reiss models, for many popular parametric models in multivariate extremes there is a one-to-one correspondence between the extremal coefficients and the parameters. Consequently, our approach will result in more accurate parameter estimation for these models.

For more sophisticated parametric models, moments of higher orders  $\mathbb{E}[(\mathbf{v}_I^T \Theta^I)^p]$ ,  $p \geq 1$ , are also meaningful and can be used for asserting other properties of the spectral distribution. They represent interesting alternatives to the stable tail dependence function  $L$  as, for small  $p$ , such functions of  $\mathbf{v}$  have nice regularity properties (constant for  $p = 1$ , quadratic for  $p = 2$ ) that have proven to be very useful for inference purpose.

We focus on reducing the variance under ad-hoc assumptions implying bias negligibility. In our numerical illustrations, the biases of our procedures were empirically small. A systematic bias analysis and potential bias reduction for this new class of estimators should be considered in the future. Remind that, in the case of unknown margins, our procedure requires some variance estimation. To this end, we propose a subsampling step which could also be useful for other inference methods in extremes. In our simulation study, we found empirically that a subsampling ratio of 1/4 provides sufficiently accurate estimates of the asymptotic standard deviation. Larger ratios lead to significant underestimation of the standard deviation. A more detailed analysis of this phenomenon in a general setting, however, is beyond the scope of this paper.

Finally the first two steps of our procedure in the unknown margin settings are specific to tail-equivalence of transformed margins  $X_i^* = r_i X_i^{\alpha_i}$ ,  $1 \leq i \leq d$ . A one-step procedure based on rank transforms similar to the one of Einmahl et al. [2001] could be investigated. However the ratio-type statistics involved in moment estimators have not yet been proven to relate well with rank transforms.

## Acknowledgements

The authors are grateful to three anonymous referees and an associate editor for their valuable comments and suggestions. The work of MO was partially supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2075 – 390740016. OW acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR20-CE40-0025-01 (T-REX project).

## References

- B. Basrak, R. A. Davis, and T. Mikosch. A characterization of multivariate regular variation. *Ann. Appl. Probab.*, 12(3):908–920, 2002.
- E. Bernard, P. Naveau, M. Vrac, and O. Mestre. Clustering of maxima: Spatial dependencies among heavy rainfall in france. *Journal of Climate*, 26(20):7929–7937, 2013.
- J. Boman and F. Lindskog. Support theorems for the Radon transform and Cramér–Wold theorems. *J. Theor. Probab.*, 22(3):683–710, 2009.
- G. Buriticá and P. Naveau. Stable sums to infer high return levels of multivariate rainfall time series. *arXiv preprint arXiv:2112.02878*, 2021.
- P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 1997.
- D. Cooley and E. Thibaud. Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604, 2019.
- D. Cooley, P. Naveau, and P. Poncet. Variograms for spatial max-stable random fields. In *Dependence in probability and statistics*, pages 373–390. Springer, 2006.

- L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006.
- H. Drees and X. Huang. Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64(1):25–46, 1998.
- H. Drees and A. Sabourin. Principal component analysis for multivariate extremes. *arXiv preprint arXiv:1906.11043*, 2019.
- J. H. Einmahl, L. De Haan, and X. Huang. Estimating a multidimensional extreme-value distribution. *Journal of Multivariate Analysis*, 47(1):35–47, 1993.
- J. H. Einmahl, L. de Haan, and A. K. Sinha. Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications*, 70(2):143–171, 1997.
- J. H. Einmahl, L. De Haan, and V. I. Piterbarg. Nonparametric estimation of the spectral measure of an extreme value distribution. *Annals of Statistics*, pages 1401–1423, 2001.
- J. H. J. Einmahl, A. Krajina, and J. Segers. An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.*, 40(3):1764–1793, 2012.
- C. Genest and J. Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B): 2990–3022, 2009.
- J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286, 1989.
- C. Klüppelberg and S. Pergamenchtchikov. Extremal behaviour of models with multivariate random recurrence representation. *Stoch. Proc. Appl.*, 117(4):432–456, 2007.
- G. Mainik and L. Rüschendorf. On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, 14(4):593–623, 2010.
- J. Pickands, III. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131, 1975.
- D. N. Politis and J. P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, pages 2031–2050, 1994.
- S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York, 2007.
- M. Schlather and J. A. Tawn. A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156, 2003.
- R. L. Smith. Max-stable processes and spatial extremes. Unpublished manuscript, 1990.
- G. Stupfler. On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes*, 22(4):749–769, 2019.
- A. W. Van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.

## A Proof of Theorem 5

In order to prove functional convergence, we apply the functional Central Limit Theorem 2.11.9 in Van der Vaart and Wellner [1996]. Let us first check the convergence of the covariances.

For  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$ ,  $(\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta$  and  $p \in K_0$ , define the function

$$f_{\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p} : (0, \infty)^d \rightarrow [0, \infty), \quad f_{\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p}(\mathbf{x}) = \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/\boldsymbol{\beta}} \right)^p \mathbf{1}\{\|\mathbf{s} \circ \mathbf{x}\| > 1\}.$$

Define the function space  $\mathcal{F} = \{f_{\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p} : \mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K_0\}$  equipped with the semimetric  $\rho(f_{\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p_1}, f_{\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma}, p_2}) = \max\{\|\mathbf{v} - \mathbf{w}\|, \|\mathbf{s} - \mathbf{t}\|, \|\boldsymbol{\beta} - \boldsymbol{\gamma}\|, |p_1 - p_2|\}$ . Furthermore, for  $l = 1, \dots, n$ ,  $n \in \mathbb{N}$ , let

$$Z_{nl}(f) = \sqrt{\frac{a^*(u_n)}{n}} f(u_n^{-1} \mathbf{X}_l^*), \quad f \in \mathcal{F}.$$

Then, each element of the sequence  $(\{G_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p); \mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K_0\})_{n \in \mathbb{N}}$  can be rewritten as

$$\{G_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)\}_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K} = \left\{ \sum_{l=1}^n (Z_{nl}(f) - \mathbb{E}Z_{nl}(f)) \right\}_{f \in \mathcal{F}}, \quad n \in \mathbb{N}.$$

Analogously to the results from Prop. 4, we can calculate

$$\begin{aligned}
& \text{Cov} \left( \sum_{l=1}^n Z_{nl}(f_{\mathbf{v}, \mathbf{s}, \beta, p_1}), \sum_{l=1}^n Z_{nl}(f_{\mathbf{w}, \mathbf{t}, \gamma, p_2}) \right) \\
&= a^*(u_n) \mathbb{E} [f_{\mathbf{v}, \mathbf{s}, \beta, p_1}(u_n^{-1} \mathbf{X}^*) f_{\mathbf{w}, \mathbf{t}, \gamma, p_2}(u_n^{-1} \mathbf{X}^*)] \\
&\quad - a^*(u_n) \mathbb{E} [f_{\mathbf{v}, \mathbf{s}, \beta, p_1}(u_n^{-1} \mathbf{X}^*)] \mathbb{E} [f_{\mathbf{w}, \mathbf{t}, \gamma, p_2}(u_n^{-1} \mathbf{X}^*)] \\
&= \mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \mathbf{X}^*}{\|\mathbf{s} \circ \mathbf{X}^*\|} \right)^{1/\beta} \right)^{p_1} \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \mathbf{X}^*}{\|\mathbf{t} \circ \mathbf{X}^*\|} \right)^{1/\gamma} \right)^{p_2} \mathbb{1}_{\|\mathbf{s} \circ \mathbf{X}^*\| \wedge \|\mathbf{t} \circ \mathbf{X}^*\| > u_n} \right] \\
&\quad \cdot a^*(u_n) \mathbb{P}(\|\mathbf{s} \circ \mathbf{X}^*\| \wedge \|\mathbf{t} \circ \mathbf{X}^*\| > u_n) \\
&\quad - \frac{1}{\sqrt{a^*(u_n)}} \mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \mathbf{X}^*}{\|\mathbf{s} \circ \mathbf{X}^*\|} \right)^{1/\beta} \right)^{p_1} \mathbb{1}_{\|\mathbf{s} \circ \mathbf{X}^*\| > u_n} \right] \cdot a^*(u_n) \mathbb{P}(\|\mathbf{s} \circ \mathbf{X}^*\| > u_n) \\
&\quad \cdot \frac{1}{\sqrt{a^*(u_n)}} \mathbb{E} \left[ \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \mathbf{X}^*}{\|\mathbf{t} \circ \mathbf{X}^*\|} \right)^{1/\gamma} \right)^{p_2} \mathbb{1}_{\|\mathbf{t} \circ \mathbf{X}^*\| > u_n} \right] \cdot a^*(u_n) \mathbb{P}(\|\mathbf{t} \circ \mathbf{X}^*\| > u_n).
\end{aligned}$$

This sequence of covariances converges to, as  $n \rightarrow \infty$ ,

$$\begin{aligned}
& \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \Theta}{\|\mathbf{s} \circ \Theta\|} \right)^{1/\beta} \right)^{p_1} \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \Theta}{\|\mathbf{t} \circ \Theta\|} \right)^{1/\gamma} \right)^{p_2} (\|\mathbf{s} \circ \Theta\| \wedge \|\mathbf{t} \circ \Theta\|) \right]}{\mathbb{E}[\|\mathbf{s} \circ \Theta\| \wedge \|\mathbf{t} \circ \Theta\|]} \cdot \tau \cdot \mathbb{E}[\|\mathbf{s} \circ \Theta\| \wedge \|\mathbf{t} \circ \Theta\|] \\
&= \tau \cdot \mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \Theta}{\|\mathbf{s} \circ \Theta\|} \right)^{1/\beta} \right)^{p_1} \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \Theta}{\|\mathbf{t} \circ \Theta\|} \right)^{1/\gamma} \right)^{p_2} (\|\mathbf{s} \circ \Theta\| \wedge \|\mathbf{t} \circ \Theta\|) \right].
\end{aligned}$$

Furthermore, as  $\sup_{\mathbf{x} \in (0, \infty)^d} |f(\mathbf{x})| \leq 1$  for all  $f \in \mathcal{F}$ , we have that

$$\|Z_{nl}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |Z_{nl}(f)| \leq \sqrt{\frac{a^*(u_n)}{n}} \text{ a.s.}$$

for all  $l = 1, \dots, n$  and  $n \in \mathbb{N}$ .

Consequently, we check the Lindeberg condition: for any  $k \in \mathbb{N}$ , we have

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}[\|Z_{ni}\|_{\mathcal{F}}^k \mathbb{1}_{\{\|Z_{ni}\|_{\mathcal{F}} > \eta\}}] \leq \lim_{n \rightarrow \infty} n \left( \frac{a^*(u_n)}{n} \right)^{k/2} \mathbb{1}_{\{\sqrt{a^*(u_n)/n} > \eta\}} = 0, \quad (41)$$

where we used that, for any  $\eta > 0$ , eventually  $a^*(u_n)/n < \eta^2$  since  $\lim_{n \rightarrow \infty} n/a^*(u_n) = \infty$ . Making use of (41) for  $k = 2$  one can easily see that, for all  $a_1, \dots, a_m \in \mathbb{R}$ ,  $f_1, \dots, f_m \in \mathcal{F}$  and  $m \in \mathbb{N}$ , the triangular arrays  $(\sum_{i=1}^m a_i Z_{ni}(f_i))_{1 \leq l \leq n, n \in \mathbb{N}}$  satisfy a Lindeberg condition and, therefore,  $G_n$  converges to  $G$  in terms of finite-dimensional distributions. Setting  $k = 1$  in (41) gives the Lindeberg type condition of Theorem 2.11.9 in Van der Vaart and Wellner [1996].

We further check the equi-continuity condition, i.e. the second condition of that theorem. To this end, we consider two elements  $f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta, p} \in \mathcal{F}$  and  $f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \gamma, p} \in \mathcal{F}$  with distance  $\max\{\|\mathbf{v}^{(1)} - \mathbf{v}^{(2)}\|, \|\mathbf{s}^{(1)} - \mathbf{s}^{(2)}\|, \|\beta - \gamma\|\} < \eta$  for some  $0 < \eta < \delta$ . Notice that for sufficiently small  $\eta$ , we have that  $\mathbf{s}^{(1)}$  and  $\mathbf{s}^{(2)}$  necessarily belong to  $A_{\delta, I}$  for the same non-empty index set  $I$ , i.e. that the same components of both vectors are different from zero. Moreover we assume with no loss of generality that there exists some  $\mathbf{y} \in A_{\delta, I}$  satisfying  $s_i^{(1)}, s_i^{(2)} \in [y_i, y_i + \eta]$  for all  $i \in I$ .

We distinguish three cases to obtain an upper bound for  $|f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x}) - f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})|$ :

1.  $\|\mathbf{s}^{(1)} \circ \mathbf{x}\| \leq 1$  and  $\|\mathbf{s}^{(2)} \circ \mathbf{x}\| \leq 1$ :

In this case, we have

$$|f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x}) - f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})| = |0 - 0| = 0.$$

2.  $\|\mathbf{s}^{(1)} \circ \mathbf{x}\| \geq 1$  and  $\|\mathbf{s}^{(2)} \circ \mathbf{x}\| \geq 1$ :

In this case, both the indicator functions in  $f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x})$  and in  $f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})$  are equal to 1. The

fact that  $b^p - a^p = \int_a^b p x^{p-1} dx \leq p(b - a)$  for  $0 \leq a \leq b \leq 1$  and  $p \in \mathbb{N}_0$ , leads to the bound

$$|f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x}) - f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})| \leq p \cdot |f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, 1}(\mathbf{x}) - f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, 1}(\mathbf{x})|.$$

Thus, with  $\mathbf{v} = \mathbf{v}^{(1)} \wedge \mathbf{v}^{(2)}$ ,  $\mathbf{s} = \mathbf{s}^{(1)} \wedge \mathbf{s}^{(2)}$  and  $\beta = \beta^{(1)} \wedge \beta^{(2)}$  denoting componentwise minima and the fact that  $\mathbf{s} + \eta \mathbf{1}_I \leq (1 + (1 + \delta)\eta)\mathbf{s}$  for  $\mathbf{s} \in A_{\delta, I}$ , we obtain

$$\begin{aligned} & |f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x}) - f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})| \\ & \leq p(\mathbf{v} + \eta \mathbf{1}_I)^\top \left[ \left( \frac{(\mathbf{s} + \eta \mathbf{1}_I) \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} \vee \left( \frac{(\mathbf{s} + \eta \mathbf{1}_I) \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/\beta} \right] \\ & \quad - p\mathbf{v}^\top \left[ \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|(\mathbf{s} + \eta \mathbf{1}_I) \circ \mathbf{x}\|} \right)^{1/\beta} \wedge \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|(\mathbf{s} + \eta \mathbf{1}_I) \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} \right] \\ & \leq p(\mathbf{v} + \eta \mathbf{1}_I)^\top \left[ \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} (1 + (1 + \delta)\eta) \right)^{1/(\beta + \eta \mathbf{1})} \vee \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} (1 + (1 + \delta)\eta) \right)^{1/\beta} \right] \\ & \quad - p\mathbf{v}^\top \left[ \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \frac{1}{1 + (1 + \delta)\eta} \right)^{1/\beta} \wedge \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \frac{1}{1 + (1 + \delta)\eta} \right)^{1/(\beta + \eta \mathbf{1})} \right] \\ & \leq p(\mathbf{v} + \eta \mathbf{1})^\top \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} (1 + (1 + \delta)\eta)^{1 + \delta} - p\mathbf{v}^\top \left( \frac{\mathbf{s} \circ \mathbf{x}}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/\beta} \frac{1}{(1 + (1 + \delta)\eta)^{1 + \delta}} \\ & \leq p\eta \mathbf{1}^\top \mathbf{1}_I (1 + (1 + \delta)\eta)^{1 + \delta} \\ & \quad + p\mathbf{v}^\top \mathbf{1}_I \max_{i \in I} \left( \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} (1 + (1 + \delta)\eta)^{1 + \delta} - \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/\beta} \frac{1}{(1 + (1 + \delta)\eta)^{1 + \delta}} \right) \\ & \leq p\eta d (1 + (1 + \delta)\eta)^{1 + \delta} + p \max_{i \in I} \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} \cdot \left[ (1 + (1 + \delta)\eta)^{1 + \delta} - \frac{1}{(1 + (1 + \delta)\eta)^{1 + \delta}} \right] \\ & \quad + p \max_{i \in I} \left( \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} - \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/\beta} \right) \cdot \frac{1}{(1 + (1 + \delta)\eta)^{1 + \delta}} \\ & \leq pd (1 + (1 + \delta)\eta)^{1 + \delta} \eta + p \frac{(1 + (1 + \delta)\eta)^{2 + 2\delta} - 1}{(1 + (1 + \delta)\eta)^{1 + \delta}} \\ & \quad + p \max_{i \in I} \left( \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/(\beta + \eta \mathbf{1})} - \left( \frac{s_i x_i}{\|\mathbf{s} \circ \mathbf{x}\|} \right)^{1/\beta} \right) \\ & \leq pd (1 + (1 + \delta)\eta)^{1 + \delta} \eta + p2(1 + \delta)^2 (1 + (1 + \delta)\eta)^\delta \eta + p(1 + \delta)^2 \left[ \sup_{z \in [0, 1]} \log(z) z^{1/(1 + \delta)} \right] \eta. \end{aligned}$$

Bounding  $p$  by  $p_{\max} = \max_{p \in K_0} p$  and noting that  $\sup_{z \in [0, 1]} \log(z) z^{1/(1 + \delta)} < \infty$ , we have an overall bound of the form  $C\eta$  some constant  $C = C(\delta, p_{\max}, d) > 0$  for sufficiently small  $\eta$ .

This case occurs only if  $(1 + \delta)\|\mathbf{x}\| > 1$  as  $\mathbf{s}^{(1)}$  and  $\mathbf{s}^{(2)}$  belong to  $A_{\delta, I}$ .

3.  $\min\{\|\mathbf{s}^{(1)} \circ \mathbf{x}\|, \|\mathbf{s}^{(2)} \circ \mathbf{x}\|\} \leq 1 < \max\{\|\mathbf{s}^{(1)} \circ \mathbf{x}\|, \|\mathbf{s}^{(2)} \circ \mathbf{x}\|\}$ :

In this case one of the functions  $f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x})$  and  $f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})$  equals zero, while the other one is bounded by 1. Consequently, noting that this case does not appear if  $\mathbf{s}^{(1)} = \mathbf{s}^{(2)}$ , we have

$$\begin{aligned} & |f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \beta^{(1)}, p}(\mathbf{x}) - f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \beta^{(2)}, p}(\mathbf{x})| \\ & \leq \mathbf{1} \left\{ \bigcup_{i \in I, s_i^{(1)} \neq s_i^{(2)}} \{u_n / (s_i^{(1)} \vee s_i^{(2)}) \leq x_i < u_n / (s_i^{(1)} \wedge s_i^{(2)})\} \right\}. \end{aligned} \quad (42)$$

As, by construction,  $y_i \leq \min\{s_i^{(1)}, s_i^{(2)}\} \leq \max\{s_i^{(1)}, s_i^{(2)}\} \leq y_i + \eta$  this case occurs only if  $y_i x_i \leq 1 < (y_i + \eta)x_i$  for at least one  $i \in I$ .

Combining the three cases, we obtain the following uniform bound on  $\|\mathbf{v}^{(1)} - \mathbf{v}^{(2)}\| < \eta$ ,  $\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\| < \eta$  and  $s_i^{(1)}, s_i^{(2)} \in [y_i, y_i + \eta]$ ,  $i \in I$ , for some  $\mathbf{y} \in A_{\delta, I}$ :

$$\begin{aligned} & (Z_{nl}(f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \boldsymbol{\beta}^{(1)}, p}) - Z_{nl}(f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \boldsymbol{\beta}^{(2)}, p}))^2 \\ & \leq \frac{a^*(u_n)}{n} C^2 \eta^2 \mathbf{1} \{ \|\mathbf{X}^*\| > (1 + \delta)^{-1} u_n \} \\ & \quad + \frac{a^*(u_n)}{n} \cdot \mathbf{1} \left\{ \bigcup_{i \in I} \{ u_n (y_i + \eta)^{-1} \leq X_{li}^* < u_n y_i^{-1} \} \right\} \\ & =: Q_{nl}^{(1)}(\eta) + Q_{nl}^{(2)}(\eta, \mathbf{y}). \end{aligned} \tag{43}$$

In the following, we will consider the two summands separately. From the definition of  $Q_{nl}^{(1)}$  and regular variation of the function  $a^*$ , it directly follows that

$$\begin{aligned} \sum_{l=1}^n \mathbb{E} \left[ Q_{nl}^{(1)}(\eta) \right] & \leq a^*(u_n) C^2 \eta^2 \mathbb{P}(\|\mathbf{X}^*\| > (1 + \delta)^{-1} u_n) \\ & \sim \tau \frac{a^*(u_n)}{a^*((1 + \delta)^{-1} u_n)} C^2 \eta^2 \sim \tau (1 + \delta) C^2 \eta^2, \quad n \rightarrow \infty. \end{aligned}$$

As  $\tau \leq d$  we obtain for  $n$  sufficiently large

$$\sum_{l=1}^n \mathbb{E} \left[ Q_{nl}^{(1)}(\eta) \right] \leq 2d(1 + \delta) C^2 \eta^2. \tag{44}$$

For  $Q_{nl}^{(2)}$  we have that, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \sum_{l=1}^n \mathbb{E} \left[ Q_{nl}^{(2)}(\eta, \mathbf{y}) \right] & \leq a^*(u_n) \sum_{i \in I} \mathbb{P}(u_n (y_i + \eta)^{-1} \leq X_i^* < u_n y_i^{-1}) \\ & \sim \sum_{i \in I} \left( \frac{\mathbb{P}(X_i^* > u_n (y_i + \eta)^{-1})}{\mathbb{P}(X_i^* > u_n)} - \frac{\mathbb{P}(X_i^* > u_n y_i^{-1})}{\mathbb{P}(X_i^* > u_n)} \right). \end{aligned}$$

with  $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}$  in  $A_{\delta, I}$ . As the function  $u \mapsto \mathbb{P}(X_i^* > u)$  is regular varying with index  $-1$  for each  $i = 1, \dots, d$ , the convergence of  $\mathbb{P}(X_i^* > u_n y) / \mathbb{P}(X_i^* > u_n) \rightarrow y^{-1}$  is uniform in  $y \in [(1 + \delta)^{-1}, 1 + \delta]$ , i.e. there exists a sequence of positive numbers  $\{f_n\}_{n \in \mathbb{N}}$  with  $f_n \rightarrow 0$  such that

$$(1 - f_n) y^{-1} \leq \frac{\mathbb{P}(X_i^* > u_n y)}{\mathbb{P}(X_i^* > u_n)} \leq (1 + f_n) y^{-1}$$

for all  $n \in \mathbb{N}$ ,  $i \in \{1, \dots, d\}$  and  $y \in [(1 + \delta)^{-1}, 1 + \delta]$ . Thus, we obtain, as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \sup_{\mathbf{y} \in A_{\delta, I}} \sum_{i \in I} \left( \frac{\mathbb{P}(X_i^* > u_n (y_i + \eta)^{-1})}{\mathbb{P}(X_i^* > u_n)} - \frac{\mathbb{P}(X_i^* > u_n y_i^{-1})}{\mathbb{P}(X_i^* > u_n)} \right) \\ & \leq |I| \cdot (1 + f_n) \cdot \eta + 2f_n \cdot \sup_{\mathbf{y} \in A_{\delta, I}} \sum_{i \in I} y_i \leq 2d\eta + 2df_n(1 + \delta) \end{aligned}$$

for sufficiently large  $n$ .

Combining the estimates obtained for  $Q_{nl}^{(1)}$  and  $Q_{nl}^{(2)}$  and denoting  $\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in [\mathbf{y}, \mathbf{y} + \eta]$  for  $\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in A_{\delta, I}$  satisfying  $s_i^{(1)}, s_i^{(2)} \in [y_i, y_i + \eta]$  for all  $i \in I$ ,  $I$  uniquely determined by  $\mathbf{y} \in A_{\delta, I}$  for any  $\mathbf{y} \in A_{\delta}$ , we obtain for  $n$  large enough

$$\begin{aligned} & \sup_{\mathbf{y} \in A_{\delta}} \sum_{l=1}^n \mathbb{E} \left[ \sup_{\|\mathbf{v}^{(1)} - \mathbf{v}^{(2)}\| < \eta, (\mathbf{s}^{(1)}, \mathbf{s}^{(2)}) \in [\mathbf{y}, \mathbf{y} + \eta]^2, \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\| < \eta} (Z_{nl}(f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \boldsymbol{\beta}^{(1)}, p}) - Z_{nl}(f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \boldsymbol{\beta}^{(2)}, p}))^2 \right] \\ & \leq 2d \left( (1 + \delta) C^2 \eta^2 + \eta + (1 + \delta) f_n \right). \end{aligned}$$

Reminding that  $f_n \rightarrow 0$ , we obtain

$$\sup_{\rho(f, g) < \eta_n} \sum_{l=1}^n \mathbb{E} \left[ (Z_{nl}(f) - Z_{nl}(g))^2 \right] \leq 2d \left( (1 + \delta) C^2 \eta_n^2 + \eta_n + (1 + \delta) f_n \right) \rightarrow 0$$

as  $\eta_n \rightarrow 0$ .

Now, we are checking the entropy condition of Theorem 2.11.9 in Van der Vaart and Wellner [1996] by constructing and counting  $\varepsilon$ -brackets for  $\varepsilon > 0$ . To this end, for each  $i \in I$  and  $n \in \mathbb{N}$ , we consider the conditional distribution of  $u_n/X_i^* \mid X_i^* > (1 + \delta)^{-1}u_n$  which, by definition is supported on  $(0, 1 + \delta)$ .

Putting all the atoms  $\{m_j^{(i,n)}\}$  of mass larger than  $\varepsilon^2/(4(1 + \delta)d)$  separately into singletons (there are at most  $1 \leq j \leq \lceil 4(1 + \delta)d/\varepsilon^2 \rceil$  of them), the set  $[(1 + \delta)^{-1}, 1 + \delta] \cup \{m_j^{(i,n)}\}$  can be divided in at most  $\lceil 4(1 + \delta)d/\varepsilon^2 \rceil$  sets  $J_j^{(i,n)}$ ,  $\lceil 4(1 + \delta)d/\varepsilon^2 \rceil + 1 \leq j \leq 2\lceil 4(1 + \delta)d/\varepsilon^2 \rceil$  such that

$$\mathbb{P}\left(\frac{u_n}{X_i^*} \in J_j^{(i,n)} \mid X_i^* > (1 + \delta)^{-1}u_n\right) \leq \frac{\varepsilon^2}{4(1 + \delta)d},$$

for  $\lceil 4(1 + \delta)d/\varepsilon^2 \rceil + 1 \leq j \leq 2\lceil 4(1 + \delta)d/\varepsilon^2 \rceil$ . Thus, denoting  $J_j^{(i,n)} = \{m_j^{(i,n)}\}$  for any  $1 \leq j \leq \lceil 4(1 + \delta)d/\varepsilon^2 \rceil$ , the set  $A_\delta \subset [0, 1 + \delta]^d$  can be divided into  $2^d \lceil 4(1 + \delta)d/\varepsilon^2 \rceil^d$  sets of the form

$$J = J_{j_1, \dots, j_d}^{(n)} = J_{j_1}^{(1,n)} \times J_{j_2}^{(2,n)} \times \dots \times J_{j_d}^{(d,n)}, \quad 1 \leq j_1, \dots, j_d \leq 2\lceil 4(1 + \delta)d/\varepsilon^2 \rceil. \quad (45)$$

Choosing  $n$  sufficiently large such that  $a^*(u_n) \leq 2(1 + \delta)\mathbb{P}(X_i^* > u_n(1 + \delta)^{-1})^{-1}$  for all  $i = 1, \dots, d$ , we obtain that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in J} a^*(u_n) \mathbf{1} \left\{ \bigcup_{i \in I, s_i^{(1)} \neq s_i^{(2)}} \{u_n/(s_i^{(1)} \vee s_i^{(2)}) \leq X_i^* < u_n/(s_i^{(1)} \wedge s_i^{(2)})\} \right\} \right] \\ & \leq 2(1 + \delta) \sum_{i \in I, j_i > \lceil 4(1 + \delta)d/\varepsilon^2 \rceil} \mathbb{P}\left(\frac{u_n}{X_i^*} \in J_{j_i}^{(i,n)} \mid X_i^* > (1 + \delta)^{-1}u_n\right) \leq \frac{\varepsilon^2}{2} \end{aligned}$$

Setting  $\eta_0 = \varepsilon/(2C\sqrt{d(1 + \delta)})$ , Eq. (42) and (44) imply that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\|\mathbf{v}^{(1)} - \mathbf{v}^{(2)}\| < \eta_0, \|\mathbf{s}^{(1)} - \mathbf{s}^{(2)}\| < \eta_0, \mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in J, \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\| < \eta_0} (Z_{nl}(f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \boldsymbol{\beta}^{(1)}, p}) - Z_{nl}(f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \boldsymbol{\beta}^{(2)}, p}))^2 \right] \\ & \leq \sum_{l=1}^n \mathbb{E} \left[ Q_{nl}^{(1)}(\eta_0) \right] \\ & + \mathbb{E} \left[ \sup_{\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in J} a^*(u_n) \mathbf{1} \left\{ \bigcup_{i \in I, s_i^{(1)} \neq s_i^{(2)}} \{u_n/(s_i^{(1)} \vee s_i^{(2)}) \leq X_i^* < u_n/(s_i^{(1)} \wedge s_i^{(2)})\} \right\} \right] \\ & \leq \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2} = \varepsilon^2. \end{aligned}$$

Now, we note that  $\partial B_1^+(\mathbf{0}) \subset [0, 1]^d$  can be covered by  $\lceil 1/\eta_0 \rceil^d$  hyperrectangles  $B_1, B_2, \dots$  of side length  $\eta_0$ , while  $A'_\delta \subset [0, 1 + \delta]^{2d}$  can be covered by at most  $\lceil (1 + \delta)/\eta_0 \rceil^{2d}$  hyperrectangles  $C_1, C_2, \dots$  of the same side length: Combining each set  $B_k$  with each intersection of a set  $C_j$  with each set  $J_{j_1, \dots, j_d}^{(n)}$  from (45) extended to  $J_{j_1, \dots, j_d}^{(n)} \times [(1 + \delta)^{-1}, 1 + \delta]^d$ , we obtain at most

$$\left\lceil \frac{1}{\eta_0} \right\rceil^d \cdot \left\lceil \frac{1 + \delta}{\eta_0} \right\rceil^{2d} \cdot \left\lceil \frac{4(1 + \delta)}{\varepsilon^2} \right\rceil^d = \left\lceil \frac{2C\sqrt{d(1 + \delta)}}{\varepsilon} \right\rceil^d \cdot \left\lceil \frac{2C\sqrt{d(1 + \delta)^3}}{\varepsilon} \right\rceil^{2d} \cdot \left\lceil \frac{4(1 + \delta)}{\varepsilon^2} \right\rceil^d$$

sets of the type  $D_{k, j, j_1, \dots, j_d}^{(n)} = B_k \times (C_j \cap (J_{j_1, \dots, j_d}^{(n)} \times [(1 + \delta)^{-1}, 1 + \delta]^d))$  satisfying

$$\mathbb{E} \left[ \sup_{(\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \boldsymbol{\beta}^{(1)}, p), (\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \boldsymbol{\beta}^{(2)}, p) \in D_{k, j, j_1, \dots, j_d}^{(n)}} (Z_{nl}(f_{\mathbf{v}^{(1)}, \mathbf{s}^{(1)}, \boldsymbol{\beta}^{(1)}, p}) - Z_{nl}(f_{\mathbf{v}^{(2)}, \mathbf{s}^{(2)}, \boldsymbol{\beta}^{(2)}, p}))^2 \right] \leq \varepsilon^2.$$

As this number are the same for each  $p \in K_0$ , the function space  $\mathcal{F}$  can be partitioned into at most

$$N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2) = |K_0| \left\lceil \frac{2C\sqrt{d(1 + \delta)}}{\varepsilon} \right\rceil^d \cdot \left\lceil \frac{2C\sqrt{d(1 + \delta)^3}}{\varepsilon} \right\rceil^{2d} \cdot \left\lceil \frac{4(1 + \delta)}{\varepsilon^2} \right\rceil^d \in \mathcal{O}(\varepsilon^{-5d})$$

$\varepsilon$ -brackets. Consequently,

$$\int_0^{\varepsilon_n} \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2)} d\varepsilon \rightarrow 0$$

as  $\varepsilon_n \rightarrow 0$ . Now, weak convergence follows from Theorem 2.11.9 in Van der Vaart and Wellner [1996].

## B Proof of Cor. 6

From Thm. 5 and Eq. (27), we deduce that

$$\left\{ \sqrt{\frac{n}{a^*(u_n)}} \left( a^*(u_n) \widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) - \tau \mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{1/\beta} \right)^p \|\mathbf{s} \circ \boldsymbol{\Theta}\| \right] \right) \right\} \rightarrow G \quad (46)$$

weakly in  $\ell^\infty(\partial B_1^+(\mathbf{0}) \times A'_\delta \times (K \cup \{0\}))$  as  $n \rightarrow \infty$ .

For the case  $p = 0$  where  $\widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, 0) = \widehat{P}_{n,u_n}(\mathbf{s})$ , we can use the monotonicity of  $\widehat{P}_{n,u_n}$  to conclude that, for all  $\mathbf{s} \in A_{\delta,I}$ ,

$$a^*(u_n) \widehat{P}_{n,u_n,\ell}(\mathbf{s}) \geq a^*(u_n) \widehat{P}_{n,u_n,\ell}((1+\delta)^{-1} \mathbf{1}_I) \rightarrow_p \tau \cdot \mathbb{E}(\|(1+\delta)^{-1} \cdot \boldsymbol{\Theta}\|) = \tau_I (1+\delta)^{-1},$$

due to Eq. (13). Thus, for any  $\varepsilon \in (0, \tau_I (1+\delta)^{-1})$ , we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(a^*(u_n) \widehat{P}_{n,u_n}(\mathbf{s}) > \varepsilon \text{ for all } \mathbf{s} \in A_{\delta,I}) \\ & \geq \lim_{n \rightarrow \infty} \mathbb{P}(a^*(u_n) \widehat{P}_{n,u_n}((1+\delta)^{-1} \mathbf{1}_I) > \varepsilon) = 1. \end{aligned} \quad (47)$$

This can be strengthened to

$$\lim_{n \rightarrow \infty} \mathbb{P}(a^*(u_n) \widehat{P}_{n,u_n}(\mathbf{s}) > \varepsilon \text{ for all } \mathbf{s} \in A_\delta) = 1$$

as they are only finitely many  $\emptyset \neq I \subset \{1, \dots, d\}$ . This implies that, for  $p = 0$  in Eq. (46), the term  $a^*(u_n) \widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, 0) = a^*(u_n) \widehat{P}_{n,u_n}(\mathbf{s})$  may be replaced by the pointwise maximum  $\varepsilon \vee \{a^*(u_n) \widehat{P}_{n,u_n}(\mathbf{s})\}$ . As the map

$$\begin{aligned} \phi : \ell^\infty(\partial B_1^+(\mathbf{0}) \times A'_\delta) \times \left( \ell^\infty(\partial B_1^+(\mathbf{0}) \times A'_\delta) \cap [\varepsilon, \infty)^{\partial B_1^+(\mathbf{0}) \times A'_\delta} \right) & \rightarrow \ell^\infty(\partial B_1^+(\mathbf{0}) \times A'_\delta), \\ \phi(f, g)(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}) & = f(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}) / g(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}) \end{aligned}$$

is Hadamard-differentiable [cf. Lemma 3.9.25 in Van der Vaart and Wellner, 1996], weak convergence of the process

$$\left( \left\{ \sqrt{\frac{n}{a(u_n)}} \left( \frac{a^*(u_n) \widehat{M}_{n,u_n}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p)}{\varepsilon \vee \{a^*(u_n) \widehat{P}_{n,u_n}(\mathbf{s})\}} - c(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) \right); \mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K \right\} \right)_{n \in \mathbb{N}}$$

to  $\widetilde{G}$  is obtained by the functional delta method [cf. Theorem 3.9.4 in Van der Vaart and Wellner, 1996, for instance] with covariance of the form

$$\begin{aligned} & \text{Cov}(\widetilde{G}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p_1), \widetilde{G}(\mathbf{w}, \mathbf{t}, \boldsymbol{\gamma}, p_2)) \\ & = \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{1/\beta} \right)^{p_1} \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \boldsymbol{\Theta}}{\|\mathbf{t} \circ \boldsymbol{\Theta}\|} \right)^{1/\gamma} \right)^{p_2} (\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|) \right]}{\tau \cdot \mathbb{E}[\|\mathbf{s} \circ \boldsymbol{\Theta}\|] \cdot \mathbb{E}[\|\mathbf{t} \circ \boldsymbol{\Theta}\|]} \\ & \quad - \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{1/\beta} \right)^{p_1} \|\mathbf{s} \circ \boldsymbol{\Theta}\| \right] \cdot \mathbb{E} \left[ \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \boldsymbol{\Theta}}{\|\mathbf{t} \circ \boldsymbol{\Theta}\|} \right)^{1/\gamma} \right)^{p_2} (\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|) \right]}{\tau \cdot (\mathbb{E}[\|\mathbf{s} \circ \boldsymbol{\Theta}\|])^2 \cdot \mathbb{E}[\|\mathbf{t} \circ \boldsymbol{\Theta}\|]} \\ & \quad - \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{1/\beta} \right)^{p_1} (\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|) \right] \cdot \mathbb{E} \left[ \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \boldsymbol{\Theta}}{\|\mathbf{t} \circ \boldsymbol{\Theta}\|} \right)^{1/\gamma} \right)^{p_2} \|\mathbf{t} \circ \boldsymbol{\Theta}\| \right]}{\tau \cdot \mathbb{E}[\|\mathbf{s} \circ \boldsymbol{\Theta}\|] \cdot (\mathbb{E}[\|\mathbf{t} \circ \boldsymbol{\Theta}\|])^2} \\ & \quad + \frac{\mathbb{E} \left[ \left( \mathbf{v}^\top \left( \frac{\mathbf{s} \circ \boldsymbol{\Theta}}{\|\mathbf{s} \circ \boldsymbol{\Theta}\|} \right)^{1/\beta} \right)^{p_1} \|\mathbf{s} \circ \boldsymbol{\Theta}\| \right] \cdot \mathbb{E} \left[ \left( \mathbf{w}^\top \left( \frac{\mathbf{t} \circ \boldsymbol{\Theta}}{\|\mathbf{t} \circ \boldsymbol{\Theta}\|} \right)^{1/\gamma} \right)^{p_2} \|\mathbf{t} \circ \boldsymbol{\Theta}\| \right] \cdot \mathbb{E}[\|\mathbf{s} \circ \boldsymbol{\Theta}\| \wedge \|\mathbf{t} \circ \boldsymbol{\Theta}\|]}{\tau \cdot (\mathbb{E}[\|\mathbf{s} \circ \boldsymbol{\Theta}\|])^2 \cdot (\mathbb{E}[\|\mathbf{t} \circ \boldsymbol{\Theta}\|])^2}. \end{aligned} \quad (48)$$

Using again Eq. (47), the assertion follows.



## C Proof of Thm. 8

Analogously to Cor. 6, We first show a functional limit theorem for the estimator  $\widehat{L}_{n,u}(\mathbf{s})/\widehat{P}_{n,u_n}(\mathbf{s})$  based on normalized observations  $\mathbf{X}_i^*$ .

**Proposition 21.** *Let the assumptions of Thm. 5 hold and assume that there exists some  $\delta > 0$  such that Eq. (31) is satisfied and Eq. (27) holds for  $p = 0$ . Then, the sequence of processes  $(\{\widetilde{H}_n(\mathbf{s}) : \mathbf{s} \in \bigcup_{i=1}^d A_{\delta,\{i\}}\})_{n \in \mathbb{N}}$  with*

$$\widetilde{H}_n(\mathbf{s}) = \sqrt{\frac{n}{a^*(u_n)}} \left( \frac{\widehat{L}_{n,u}(\mathbf{s})}{\widehat{P}_{n,u_n}(\mathbf{s})} - \frac{1}{\alpha_i} \right), \quad \mathbf{s} \in A_{\delta,\{i\}}, 1 \leq i \leq d,$$

converges weakly in  $\ell^\infty(A_\delta)$  to a tight centered Gaussian process  $\widetilde{H}$  with covariance

$$\text{Cov}(\widetilde{H}(\mathbf{s}), \widetilde{H}(\mathbf{t})) = \frac{1}{\tau \alpha_i \alpha_j} \cdot \frac{\mathbb{E}[(s_i \Theta_i) \wedge (t_j \Theta_j)]}{\mathbb{E}[s_i \Theta_i] \cdot \mathbb{E}[t_j \Theta_j]}, \quad \mathbf{s} \in A_{\delta,\{i\}}, \mathbf{t} \in A_{\delta,\{j\}}, 1 \leq i, j \leq d. \quad (49)$$

*Proof.* First, we will show that the sequence of processes  $(\{H_n(\mathbf{s}) : \mathbf{s} \in \bigcup_{i=1}^d A_{\delta,\{i\}}\})_{n \in \mathbb{N}}$  with

$$H_n(\mathbf{s}) = \sqrt{\frac{n}{a^*(u_n)}} \left( a^*(u_n) \widehat{L}_{n,u_n}(\mathbf{s}) - a^*(u_n) \mathbb{E}[\widehat{L}_{n,u_n}(\mathbf{s})] \right)$$

converges weakly in  $\ell^\infty(\bigcup_{i=1}^d A_{\delta,\{i\}})$  to a centered Gaussian process  $H$ . Analogously to the proof of Thm. 5, we define the function space  $\mathcal{F}' = \{f_{\mathbf{s}} : \mathbf{s} \in \bigcup_{i=1}^d A_{\delta,\{i\}}\}$  equipped with the semi-metric  $\rho'(f_{\mathbf{s}^{(1)}}, f_{\mathbf{s}^{(2)}}) = \|\mathbf{s}^{(1)} - \mathbf{s}^{(2)}\|$  where

$$f_{\mathbf{s}} : (0, \infty)^d \rightarrow [0, \infty), \quad f_{\mathbf{s}}(\mathbf{x}) = \log(\|(\mathbf{s} \circ \mathbf{x})^{1/\alpha}\|) \mathbf{1}\{\|\mathbf{s} \circ \mathbf{x}\| > 1\}.$$

For  $l = 1, \dots, n$ ,  $n \in \mathbb{N}$ , we set

$$Z_{nl}(f) = \sqrt{\frac{a^*(u_n)}{n}} f(u_n^{-1} \mathbf{X}_l^*)$$

such that

$$\{H_n(\mathbf{s})\}_{\mathbf{s} \in \bigcup_{i=1}^d A_{\delta,\{i\}}} = \left\{ \sum_{l=1}^n (Z_{nl}(f) - \mathbb{E}Z_{nl}(f)) \right\}_{f \in \mathcal{F}'}$$

By construction, each  $H_n$  is a centered process and it can be shown that its covariance function converges pointwise.

Now, using the Potter bound, for every  $\varepsilon > 0$ , there exists some  $C_\varepsilon > 0$  such that

$$\frac{\mathbb{P}(\|\mathbf{X}_I^*\| > u_n t)}{\mathbb{P}(\|\mathbf{X}_I^*\| > u_n)} \leq C_\varepsilon t^{-(1-\varepsilon)}, \quad t \geq 1,$$

for sufficiently large  $n$ . Thus, for every  $\eta > 0$  and  $k \in \{1, 2\}$ , we obtain

$$\begin{aligned} & \sum_{l=1}^n \mathbb{E}[\|Z_{n,l}\|_{\mathcal{F}'}^k \mathbf{1}\{\|Z_{n,l}\|_{\mathcal{F}'} > \eta\}] \\ & \leq n \left( \frac{a^*(u_n)}{n} \right)^{k/2} \left\| \frac{1}{\alpha} \right\|^k \mathbb{E} \left[ \log^k((1+\delta)\|\mathbf{X}_I^*\|/u_n) \mathbf{1}\{\log((1+\delta)\|\mathbf{X}_I^*\|/u_n) > \sqrt{n/a^*(u_n)}\eta\} \right] \\ & = n \left( \frac{a^*(u_n)}{n} \right)^{k/2} \left\| \frac{1}{\alpha} \right\|^k \int_0^\infty \mathbb{P} \left( \log^k((1+\delta)\|\mathbf{X}_I^*\|/u_n) \mathbf{1}\{\log((1+\delta)\|\mathbf{X}_I^*\|/u_n) > \sqrt{n/a^*(u_n)}\eta\} > x \right) dx \end{aligned}$$

$$\begin{aligned}
&= n \left( \frac{a^*(u_n)}{n} \right)^{k/2} \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^k \left[ \mathbb{P} \left( \|\mathbf{X}_I^*\| > \frac{u_n}{1+\delta} \exp \left( \sqrt{\frac{n}{a^*(u_n)}} \eta \right) \right) \cdot \left( \sqrt{\frac{n}{a^*(u_n)}} \eta \right)^k \right. \\
&\quad \left. + \int_{(\sqrt{n/a^*(u_n)})^k}^{\infty} \mathbb{P} \left( \|\mathbf{X}_I^*\| > \frac{u_n}{1+\delta} \exp(x^{1/k}) \right) dx \right] \\
&= n \left( \frac{n}{a^*(u_n)} \right)^{k/2} \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^k \eta^k \cdot \left[ \mathbb{P} \left( \|\mathbf{X}_I^*\| > \frac{u_n}{1+\delta} \exp \left( \sqrt{\frac{n}{a^*(u_n)}} \eta \right) \right) \right. \\
&\quad \left. + \int_1^{\infty} \mathbb{P} \left( \|\mathbf{X}_I^*\| > \frac{u_n}{1+\delta} \exp \left( \sqrt{\frac{n}{a^*(u_n)}} \eta u^{1/k} \right) \right) du \right] \\
&\leq C_\varepsilon n \mathbb{P}(\|\mathbf{X}_I^*\| > u_n) \left( \frac{n}{a^*(u_n)} \right)^{k/2} \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^k \frac{\eta^k}{(1+\delta)^{1-\varepsilon}} \\
&\quad \cdot \left[ \exp \left( -(1-\varepsilon)\eta \sqrt{\frac{n}{a^*(u_n)}} \right) + \int_1^{\infty} \exp \left( -(1-\varepsilon)\eta \sqrt{\frac{n}{a^*(u_n)}} u^{1/k} \right) du \right] \\
&\sim C_\varepsilon \tau_I \left( \frac{n}{a^*(u_n)} \right)^{1+k/2} \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^k \frac{\eta^k}{(1+\delta)^{1-\varepsilon}} \cdot \left[ \exp \left( -(1-\varepsilon)\eta \sqrt{\frac{n}{a^*(u_n)}} \right) \right. \\
&\quad \left. + \int_1^{\infty} \exp \left( -(1-\varepsilon)\eta \sqrt{\frac{n}{a^*(u_n)}} u^{1/k} \right) du \right] \rightarrow 0
\end{aligned}$$

because of  $n/a^*(u_n) \rightarrow 0$  as  $n \rightarrow \infty$ . For  $k = 2$ , this results in a Lindeberg condition that guarantees convergence of  $H_n$  to a centered Gaussian process  $H$  in terms of finite-dimensional distributions.

Analogously to the proof of Thm. 5, we verify the conditions from Theorem 2.11.9 in Van der Vaart and Wellner [1996] in order to ensure functional convergence. Thus, we consider two functions  $f_{\mathbf{s}^{(1)}}, f_{\mathbf{s}^{(2)}} \in \mathcal{F}'$  such that  $\|\mathbf{s}^{(1)} - \mathbf{s}^{(2)}\| < \eta$  for some  $0 < \eta < \delta$  and  $\eta < 1$  so that  $\mathbf{s}^{(1)} \in A_{\delta, \{i\}}$  and  $\mathbf{s}^{(2)} \in A_{\delta, \{i\}}$  for the same index  $i \in \{1, \dots, d\}$ , that is, there exists some  $\mathbf{y} \in A_{\delta, \{i\}}$  such that  $s_i^{(1)}, s_i^{(2)} \in [y_i, y_i + \eta]$ . Again, denoting  $\underline{\alpha} = \min_{1 \leq i \leq d} \alpha_i$ , we distinguish three cases:

1.  $\|\mathbf{s}^{(1)} \circ \mathbf{x}\| \leq 1$  and  $\|\mathbf{s}^{(2)} \circ \mathbf{x}\| \leq 1$ :  
In this case, we have

$$|f_{\mathbf{s}^{(1)}}(\mathbf{x}) - f_{\mathbf{s}^{(2)}}(\mathbf{x})| = |0 - 0| = 0.$$

2.  $\|\mathbf{s}^{(1)} \circ \mathbf{x}\| \geq 1$  and  $\|\mathbf{s}^{(2)} \circ \mathbf{x}\| \geq 1$ :

In this case, both the indicator functions in  $f_{\mathbf{s}^{(1)}}(\mathbf{x})$  and in  $f_{\mathbf{s}^{(2)}}(\mathbf{x})$  are equal to 1. Choosing  $0 < \beta < 1/2$  and using that  $|\log(x) - \log(y)| < \beta^{-1}|x - y|^\beta$ , we obtain that

$$|f_{\mathbf{s}^{(1)}}(\mathbf{x}) - f_{\mathbf{s}^{(2)}}(\mathbf{x})| \leq \frac{1}{\alpha_i \beta} \eta^\beta x_i^\beta \leq \left\| \frac{1}{\boldsymbol{\alpha}} \right\| \cdot \frac{1}{\beta} \cdot \eta^\beta \cdot \|\mathbf{x}\|^\beta.$$

This case occurs only if  $(1+\delta)\|\mathbf{x}\| > 1$  as  $\mathbf{s}^{(1)}$  and  $\mathbf{s}^{(2)}$  belong to  $\bigcup_{i=1}^d A_{\delta, \{i\}} \subset A_\delta$ .

3.  $\min\{\|\mathbf{s}^{(1)} \circ \mathbf{x}\|, \|\mathbf{s}^{(2)} \circ \mathbf{x}\|\} \leq 1 < \max\{\|\mathbf{s}^{(1)} \circ \mathbf{x}\|, \|\mathbf{s}^{(2)} \circ \mathbf{x}\|\}$ :

In this case one of the functions  $f_{\mathbf{s}^{(1)}}(\mathbf{x})$  and  $f_{\mathbf{s}^{(2)}}(\mathbf{x})$  equals zero, while the other one is bounded by 1. Here, it is important to use that  $\|\mathbf{s} \circ \mathbf{X}\| \leq u_n$  for some  $\mathbf{s} \in A_{\delta, \{i\}}$  implies that  $\|u_n^{-1} X_i\| \leq 1 + \delta$ . Thus, we will have  $\max_{j=1,2} f_{\mathbf{s}^{(j)}}(\mathbf{x}) \leq \alpha_i^{-1} \log((1+\delta)^2)$  in this case, and, thus,

$$|f_{\mathbf{s}^{(1)}}(\mathbf{x}) - f_{\mathbf{s}^{(2)}}(\mathbf{x})| \leq \frac{1}{\alpha_i} \log(1+\delta) \mathbf{1} \left\{ u_n / (s_i^{(1)} \vee s_i^{(2)}) \leq x_i < u_n / (s_i^{(1)} \wedge s_i^{(2)}) \right\}.$$

This case occurs only if  $y_i x_i \leq 1 < (y_i + \eta) x_i$ .

Thus, for each  $i \in \{1, \dots, d\}$  we obtain a uniform bound on  $s_i^{(1)}, s_i^{(2)} \in [y_i, y_i + \eta]$ ,  $i \in I$ , for some  $\mathbf{y} \in A_{\delta, \{i\}}$

$$\begin{aligned} (Z_{nl}(f_{\mathbf{s}^{(1)}}) - Z_{nl}(f_{\mathbf{s}^{(2)}}))^2 &\leq \frac{a^*(u_n)}{n} \left\| \frac{1}{\boldsymbol{\alpha}\beta} \right\|^2 \cdot \|u_n^{-1} \mathbf{X}^*\|^{2\beta} \eta^{2\beta} \mathbf{1}\{\|\mathbf{X}_l^*\| > (1+\delta)^{-1}u_n\} \\ &\quad + \frac{a^*(u_n)}{n} \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^2 \log^2(1+\delta) \cdot \mathbf{1}\{u_n(y_i\eta)^{-1} \leq X_{li}^* \leq u_n y_i^{-1}\} \\ &=: \tilde{Q}_{nl}^{(1)}(\eta) + \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^2 \log^2(1+\delta) \cdot Q_{nl}^{(2)}(\eta, \mathbf{y}) \end{aligned}$$

where  $Q_{nl}^{(2)}(\eta, \mathbf{y})$  is as in (43).

The definition of  $\tilde{Q}_{nl}^{(1)}(\eta)$  and regular variation of the function  $a^*$  imply that

$$\begin{aligned} \sum_{l=1}^n \mathbb{E} \left[ \tilde{Q}_{nl}^{(1)}(\eta) \right] &\leq a^*(u_n) \left\| \frac{1}{\boldsymbol{\alpha}\beta} \right\|^2 \eta^{2\beta} \mathbb{E} (\|u_n^{-1} \mathbf{X}^*\|^{2\beta} \mathbf{1}\{\|\mathbf{X}^*\| > (1+\delta)^{-1}u_n\}) \\ &\sim \tau \frac{a^*(u_n)}{a^*((1+\delta)^{-1}u_n)} \left\| \frac{1}{\boldsymbol{\alpha}\beta} \right\|^2 \eta^{2\beta} \mathbb{E} (\|u_n^{-1} \mathbf{X}^*\|^{2\beta} \mid \|\mathbf{X}^*\| > (1+\delta)^{-1}u_n) \\ &\sim \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^2 \frac{\tau(1+\delta)^{1-2\beta}}{\beta^2} \mathbb{E}(\|Y\boldsymbol{\Theta}^*\|^{2\beta}) \cdot \eta^{2\beta} \leq \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^2 \frac{\tau(1+\delta)^{1-2\beta}}{(1-2\beta)\beta^2} \cdot \eta^{2\beta}, \quad n \rightarrow \infty, \end{aligned}$$

since  $\mathbb{E}(Y^{2\beta}) = 1/(1-2\beta) < \infty$  for all  $\beta < 1/2$ . Thus, analogously to (44), for  $n$  sufficiently large, we have that

$$\sum_{l=1}^n \mathbb{E} \left[ \tilde{Q}_{nl}^{(1)}(\eta) \right] \leq 2 \left\| \frac{1}{\boldsymbol{\alpha}} \right\|^2 \frac{d(1+\delta)^{1-2\beta}}{(1-2\beta)\beta^2} \cdot \eta^{2\beta}$$

Combining this bound with the bound with the bounds for  $\sum_{l=1}^n \mathbb{E} \left[ Q_{nl}^{(2)}(\eta) \right]$  provided in the proof of Thm. 5, the equicontinuity condition and the entropy condition of Theorem 2.11.9 in Van der Vaart and Wellner [1996] can be verified, following the lines of the proof of Thm. 5. Thus, we obtain weak convergence of  $\{H_n(\mathbf{s}) : \mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}}\}$  to  $H$  in  $\ell^\infty(\bigcup_{i=1}^d A_{\delta, \{i\}})$ .

Making use of (31), the bias can be neglected asymptotically, i.e.

$$\left\{ \sqrt{\frac{n}{a^*(u_n)}} \left( a^*(u_n) \hat{L}_{n, u_n}(\mathbf{s}) - \frac{s_i}{\alpha_i} \right); \mathbf{s} \in A_{\delta, \{i\}}, i = 1, \dots, d \right\} \rightarrow \left\{ H(\mathbf{s}); \mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}} \right\}$$

weakly in  $\ell^\infty(A_\delta)$ . By Thm. 5 and (27) for  $p = 0$ , we also obtain

$$\left\{ \sqrt{\frac{n}{a^*(u_n)}} \left( a^*(u_n) \hat{P}_{n, u_n}(\mathbf{s}) - s_i \right); \mathbf{s} \in A_{\delta, \{i\}}, i = 1, \dots, d \right\} \rightarrow \left\{ G(\mathbf{v}, \mathbf{s}, \beta, 0); \mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}} \right\}$$

weakly  $\ell^\infty(\bigcup_{i=1}^d A_{\delta, \{i\}})$  for any fixed  $\mathbf{v} \in \partial B_1^+(\mathbf{0})$  and  $\beta \in [(1+\delta)^{-1}, 1+\delta]$  as the RHS does not depend on these variables. Note that both convergences hold jointly. Thus, analogously to the proof of Cor. 6, we can apply the functional Delta method to obtain the desired convergence

$$\left\{ \sqrt{\frac{n}{a^*(u_n)}} \left( \frac{\hat{L}_{n, u}(\mathbf{s})}{\hat{P}_{n, u_n}(\mathbf{s})} - \frac{s_i}{\alpha_i} \right); \mathbf{s} \in \times A_{\delta, \{i\}}, i = 1, \dots, d \right\} \rightarrow_w \left\{ \tilde{H}(\mathbf{s}); \mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}} \right\}$$

for some centered Gaussian process  $\tilde{H}$ . The covariance structure of  $\tilde{H}$  follows from classical computation.  $\square$

Using this result, we are ready to prove Thm. 8.

*Proof of Thm. 8.* First, we note that

$$\begin{aligned} \frac{1}{\hat{\alpha}_{n, k_n, \{i\}}} &= \frac{n^{-1} \sum_{l=1}^n \log(\|\mathbf{1}_{\{i\}} \circ \mathbf{X}_l / \mathbf{X}_{k_n:n}\|) \mathbf{1}\{\|\mathbf{1}_{\{i\}} \circ \mathbf{X}_l / \mathbf{X}_{k_n:n}\| > 1\}}{\tilde{P}_{n, k_n, \{i\}}} \\ &= \frac{n^{-1} \sum_{l=1}^n \log(\|\mathbf{1}_{\{i\}} \circ \mathbf{X}_l^* / \mathbf{X}_{k_n:n}^*\|^{1/\alpha_i}) \mathbf{1}\{\|\mathbf{1}_{\{i\}} \circ \mathbf{X}_l^* / \mathbf{X}_{k_n:n}^*\| > 1\}}{\tilde{P}_{n, k_n, \{i\}}} \\ &= \frac{\hat{L}_{n, u_n}(u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}})}{\hat{P}_{n, u_n}(u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}})}, \end{aligned}$$

cf. Eq. (28). Thus, we have to show that

$$(\tilde{H}_n(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}))_{1 \leq i \leq d} \rightarrow (\tilde{H}(\mathbf{1}_{\{i\}}))_{1 \leq i \leq d} =: \tilde{H}.$$

To this end, we use the consistency of the tail empirical measure to obtain

$$\frac{u_n}{\mathbf{X}_{k_n:n}^*} \sim \left( \frac{(F_i^*)^{-1}(1 - k_n/n)}{X_{k_n:n,i}^*} \right)_{i=1,\dots,d} \rightarrow_p \mathbf{1}, \quad n \rightarrow \infty, \quad (50)$$

cf. Equation (4.17) in Resnick [2007]. By Skorohod's representation theorem, without loss of generality, we may reformulate the statement in Prop. 21 as

$$\sup_{\mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}}} \left| \tilde{H}_n(\mathbf{s}) - \tilde{H}(\mathbf{s}) \right| \rightarrow 0 \quad \text{a.s.}, \quad n \rightarrow \infty. \quad (51)$$

Now, we estimate

$$\begin{aligned} & \left\| (\tilde{H}_n(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}))_{1 \leq i \leq d} - \tilde{H} \right\| \\ & \leq \max_{1 \leq i \leq d} \left| \tilde{H}_n(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}) - \tilde{H}(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}) \right| + \left\| (\tilde{H}(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}))_{1 \leq i \leq d} - \tilde{H} \right\|. \end{aligned} \quad (52)$$

Eq. (50) guarantees that, with probability tending to 1,  $\{u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}} \in A_{\delta, \{i\}}\}$ , and, in that case, we can bound the right-hand side of Eq. (52) by

$$\sup_{\mathbf{s} \in \bigcup_{i=1}^d A_{\delta, \{i\}}} \left| \tilde{H}_n(\mathbf{s}) - \tilde{H}(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{s}) \right| + \left\| (\tilde{H}(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}))_{1 \leq i \leq d} - \tilde{H} \right\|.$$

Here, the first term vanishes asymptotically by Eq. (51) and the second term tends to zero because of Eq. (50) and the uniform continuity of the sample paths of the tight process  $\tilde{H}$  (cf. Thm. 21). Thus, we obtain the desired convergence

$$\left\| (\tilde{H}_n(u_n/\mathbf{X}_{k_n:n}^* \circ \mathbf{1}_{\{i\}}))_{1 \leq i \leq d} - \tilde{H} \right\| \rightarrow_p 0.$$

□

## D Proof of Thm. 10

Before showing Thm. 10, we prove two auxiliary results. We start by establishing a first (but biased) version of the functional central limit theorem for the sequence  $\{\tilde{M}_{n,k_n,I}(\mathbf{v}, p)/\tilde{P}_{n,k_n,I}; \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K\}$  of processes.

**Proposition 22.** *Let  $X_l, l \in \mathbb{N}$ , be independent copies of a non-standard regularly varying  $[0, \infty)^d$ -valued random vector  $\mathbf{X}$  satisfying the assumptions of Cor. 6 and Thm. 8 for some  $\delta > 0$  and some  $K \subset \mathbb{N}$ . Then, for all non-empty  $I \subset \{1, \dots, d\}$ , we have that*

$$\begin{aligned} & \left\{ \sqrt{k_n} \left( \frac{\tilde{M}_{n,k_n,I}(\mathbf{v}, p)}{\tilde{P}_{n,k_n,I}} - c \left( \mathbf{v}, \frac{u_n}{\mathbf{X}_{k_n:n}^*} \circ \mathbf{1}_I, \frac{\boldsymbol{\alpha}}{\hat{\boldsymbol{\alpha}}_{n,k_n}}, p \right) \right); \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K \right\} \\ & \rightarrow \left\{ \tilde{G}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p); \mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K \right\}, \quad n \rightarrow \infty, \end{aligned}$$

weakly in  $\ell^\infty(\partial B_1^+(\mathbf{0}) \times K)$  where  $\mathbf{X}_{k_n:n}^*$  denotes the  $k_n$ -th order statistics of the transformed vector sample  $\mathbf{X}_l^*$ ,  $1 \leq l \leq n$ , and  $\tilde{G}$  is defined as in Cor. 6.

*Proof.* By Eq. (28), we have that

$$\sqrt{k_n} \left( \frac{\tilde{M}_{n,k_n,I}(\mathbf{v}, p)}{\tilde{P}_{n,k_n,I}} - c \left( \mathbf{v}, \frac{u_n}{\mathbf{X}_{k_n:n}^*} \circ \mathbf{1}_I, \frac{\boldsymbol{\alpha}}{\hat{\boldsymbol{\alpha}}_{n,k_n}}, p \right) \right) = \tilde{G}_n \left( \mathbf{v}, \frac{u_n}{\mathbf{X}_{k_n:n}^*} \circ \mathbf{1}_I, \frac{\boldsymbol{\alpha}}{\hat{\boldsymbol{\alpha}}_{n,k_n}}, p \right),$$

where  $\tilde{G}_N$  is defined as in Cor. 6. Furthermore, Eq. (50) and Thm. 8 provide

$$\left( \frac{u_n}{\mathbf{X}_{k_n:n}^*} \circ \mathbf{1}_I, \frac{\boldsymbol{\alpha}}{\hat{\boldsymbol{\alpha}}_{n,k_n}} \right) \rightarrow_p (\mathbf{1}_I, \mathbf{1}), \quad n \rightarrow \infty. \quad (53)$$

Again, Skohorod's representation theorem allows us to rewrite the result of Cor. 6 as

$$\sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K} \left| \tilde{G}_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) - \tilde{G}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) \right| \rightarrow 0 \quad \text{a.s.}, \quad n \rightarrow \infty. \quad (54)$$

Using Eq. (53), with probability tending to 1, we can bound the right-hand side of

$$\begin{aligned} & \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \tilde{G}_n(\mathbf{v}, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p) - \tilde{G}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p) \right| \\ \leq & \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \tilde{G}_n(\mathbf{v}, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p) - \tilde{G}(\mathbf{v}, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p) \right| \\ & + \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \tilde{G}(\mathbf{v}, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p) - \tilde{G}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p) \right|. \end{aligned}$$

by

$$\begin{aligned} & \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), (\mathbf{s}, \boldsymbol{\beta}) \in A'_\delta, p \in K} \left| \tilde{G}_n(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) - \tilde{G}(\mathbf{v}, \mathbf{s}, \boldsymbol{\beta}, p) \right| \\ & + \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \tilde{G}(\mathbf{v}, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p) - \tilde{G}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p) \right|. \end{aligned}$$

While the first term tends to zero by Eq. (54), the second term vanishes asymptotically because of Eq. (53) and the uniform continuity of  $\tilde{G}(\cdot, \cdot, p)$ , cf. the third part of Rem. 7. Consequently,

$$\sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \tilde{G}_n(\mathbf{v}, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p) - \tilde{G}(\mathbf{v}, \mathbf{1}_I, \mathbf{1}, p) \right| \rightarrow_p 0, \quad n \rightarrow \infty.$$

□

Note that the centering term in Prop. 22,  $c(\mathbf{v}_I, u_n / \mathbf{X}_{k_n:n}^* \circ \mathbf{1}_I, \boldsymbol{\alpha} / \hat{\boldsymbol{\alpha}}_{n,k_n}, p)$  is random and needs to be replaced by the desired limit  $c(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p)$ . It turns out that, in general, the difference is not negligible, i.e. the replacement of the centering term will influence the asymptotic distribution. The proof will rely on the following lemma.

**Lemma 23.** *Let the assumptions of Prop. 22 hold and assume that there exists some  $\delta > 0$  such that Eq. (32) holds for all  $i \in \{1, \dots, d\}$ . Then we have*

$$\left\{ \sqrt{k_n} \left( \frac{u_n}{X_{[sk_n]:n,i}^*} - s \right); (1 + \delta)^{-1} \leq s \leq 1 + \delta \right\} \rightarrow_w \{-G^0(s \mathbf{1}_{\{i\}})\}$$

in  $\ell^\infty([(1 + \delta)^{-1}, 1 + \delta])$ .

*Proof.* Applying Thm. 5 to the case  $p = 0$  and  $\mathbf{s} = s \mathbf{1}_{\{i\}}$  and making use of Eq. (32), we obtain that

$$\left\{ \sqrt{k_n} \left( \frac{1}{k_n} \sum_{l=1}^n \mathbf{1}\{s X_{li}^* > u_n\} - s \right); (1 + \delta)^{-1} \leq s \leq 1 + \delta \right\} \rightarrow_w \{G^0(s \mathbf{1}_{\{i\}})\}$$

in  $\ell^\infty([(1 + \delta)^{-1}, 1 + \delta])$  as  $n \rightarrow \infty$ . Using the notation  $\hat{F}_{n,i}$  for the empirical cumulative distribution function of the sample  $X_{1i}^*, \dots, X_{ni}^*$  and Skohorod's representation theorem, we may assume that

$$\sup_{(1+\delta)^{-1} \leq s \leq 1+\delta} \left| \sqrt{k_n} \left( \frac{n}{k_n} (1 - \hat{F}_{n,i}(u_n/s)) - s \right) - G^0(s \mathbf{1}_{\{i\}}) \right| \rightarrow 0 \quad \text{a.s.}, \quad n \rightarrow \infty.$$

Applying Vervaat's Lemma [cf. Lemma A.0.2 in De Haan and Ferreira, 2006, for instance] to the nondecreasing functions  $x_n(s) = n/k_n(1 - \hat{F}_{n,i}(u_n/s))$ , we obtain

$$\sup_{(1+\delta)^{-1} \leq s \leq 1+\delta} \left| \sqrt{k_n} \left( u_n / X_{[sk_n]:n,i}^* - s \right) + G^0(s \mathbf{1}_{\{i\}}) \right| \rightarrow 0 \quad \text{a.s.},$$

where we used that  $x_n^{\leftarrow}(s) = u_n / X_{[sk_n]:n,i}^*$ . □

Imposing some additional conditions on the function  $c$  and using our previous results, we can follow the same lines as the proof of Theorem 4.6 in Einmahl et al. [2012] in order to establish the desired convergence result.

*Proof of Thm. 10.* By Skohorod's representation theorem, we may assume that the weak convergences in Thm. 8, Prop. 22 and in Lemma 23 take place a.s. Then, the triangular inequality yields the bound

$$\begin{aligned}
& \sup_{\mathbf{v} \in B_1^+(\mathbf{0}), p \in K} \left| \sqrt{k} \left( \frac{\widetilde{M}_{n,k_n}(\mathbf{v}_I, p)}{\widetilde{P}_{n,k_n}} - c(\mathbf{v}_I, \mathbf{1}_I, p) \right) - \widetilde{G}(\mathbf{v}_I, \mathbf{1}_I, p) \right. \\
& \quad \left. + \sum_{i \in I} \left( c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, p) \cdot G^0(\mathbf{1}_{\{i\}}) - c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, p) \cdot \alpha_i \cdot \widetilde{H}_i \right) \right| \\
& \leq \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \widetilde{G}_n \left( \mathbf{v}_I, \frac{u_n}{\mathbf{X}_{k:n}^*} \circ \mathbf{1}_I, \frac{\alpha}{\widehat{\alpha}_{n,k_n}}, p \right) - \widetilde{G}(\mathbf{v}_I, \mathbf{1}_I, p) \right| \\
& \quad + \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| \sqrt{k} \left( c \left( \mathbf{v}_I, \frac{u_n}{\mathbf{X}_{k:n}^*} \circ \mathbf{1}_I, \frac{\alpha}{\widehat{\alpha}_{n,k_n}}, p \right) - c(\mathbf{v}_I, \mathbf{1}_I, p) \right) \right. \\
& \quad \left. + \sum_{i \in I} \left( c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, p) \cdot G^0(\mathbf{1}_{\{i\}}) - c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, p) \cdot \alpha_i \cdot \widetilde{H}_i \right) \right| \\
& =: D_1 + D_2.
\end{aligned}$$

We will show that  $D_1 \rightarrow_p 0$  and  $D_2 \rightarrow_p 0$ . By Prop. 22, we directly obtain that  $D_1 \rightarrow 0$  a.s. In order to show that  $D_2 \rightarrow_p 0$ , we first define

$$C_n(\mathbf{v}_I, p) := \sqrt{k} \left( c \left( \mathbf{v}_I, \frac{u_n}{\mathbf{X}_{k:n}^*} \circ \mathbf{1}_I, \frac{\alpha}{\widehat{\alpha}_{n,k_n}}, p \right) - c(\mathbf{v}_I, \mathbf{1}_I, p) \right), \quad \mathbf{v}_I \in \partial B_1^+(\mathbf{0}), p \in K.$$

The mean value theorem yields

$$\frac{1}{\sqrt{k}} C_n(\mathbf{v}_I, p) = \sum_{i \in I} \left( \left( \frac{u_n}{\mathbf{X}_{k:n,i}^*} - 1 \right) \cdot c_{s_i}(\mathbf{v}_I, \xi_n, \psi_n, p) + \left( \frac{\alpha_i}{\widehat{\alpha}_{n,k_n,i}} - 1 \right) \cdot c_{\beta_i}(\mathbf{v}_I, \xi_n, \psi_n, p) \right) \quad (55)$$

for some  $\xi_n = \xi_n(\mathbf{v}_I, p)$  between  $u_n/\mathbf{X}_{k:n}^* \circ \mathbf{1}_I$  and  $\mathbf{1}_I$  and  $\psi_n = \psi_n(\mathbf{v}_I, p)$  between  $\alpha/\widehat{\alpha}_{n,k_n}$  and  $\mathbf{1}$ . If  $(u_n/\mathbf{X}_{k:n}^* \circ \mathbf{1}_I, \alpha/\widehat{\alpha}_{n,k_n}) \in A'_{\delta,I}$ , which happens with probability tending to 1 by Eq. (53), we obtain from (55) that

$$\begin{aligned}
D_2 &= \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| C_n(\mathbf{v}_I, p) + \sum_{i \in I} c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, p) G^0(\mathbf{1}_{\{i\}}) - c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, p) \alpha_i \widetilde{H}_i \right| \\
&\leq \sum_{i \in I} \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| c_{s_i}(\mathbf{v}_I, \xi_n, \psi_n, p) \sqrt{k} \left( \frac{u_n}{\mathbf{X}_{k:n,i}^*} - 1 \right) + c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, p) G^0(\mathbf{1}_{\{i\}}) \right| \\
&\quad + \sum_{i \in I} \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| c_{\beta_i}(\mathbf{v}_I, \xi_n, \psi_n, p) \sqrt{k} \left( \frac{\alpha_i}{\widehat{\alpha}_{n,k_n,i}} - 1 \right) - c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, p) \alpha_i \widetilde{H}_i \right| \\
&\leq \sum_{i \in I} \left( \sup_{\mathbf{v} \in B_1^+(\mathbf{0}), (\mathbf{s}, \beta) \in A'_{\delta,I}, p \in K} |c_{s_i}(\mathbf{v}_I, \mathbf{s}, \beta, p)| \cdot \left| \sqrt{k} \left( \frac{u_n}{\mathbf{X}_{k:n,i}^*} - 1 \right) + G^0(\mathbf{1}_{\{i\}}) \right| \right. \\
&\quad \left. + \sup_{\mathbf{v} \in B_1^+(\mathbf{0}), p \in K} |c_{s_i}(\mathbf{v}_I, \xi_n, p) - c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, p)| \cdot |G^0(\mathbf{1}_{\{i\}})| \right. \\
&\quad \left. + \sup_{\mathbf{v} \in B_1^+(\mathbf{0}), (\mathbf{s}, \beta) \in A'_{\delta,I}, p \in K} |c_{\beta_i}(\mathbf{v}_I, \mathbf{s}, \beta, p)| \cdot \left| \sqrt{k} \left( \frac{\alpha_i}{\widehat{\alpha}_{n,k_n,i}} - 1 \right) - \alpha_i \widetilde{H}_i \right| \right. \\
&\quad \left. + \sup_{\mathbf{v} \in B_1^+(\mathbf{0}), p \in K} |c_{\beta_i}(\mathbf{v}_I, \xi_n, \psi_n, p) - c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, p)| \cdot |\alpha_i \widetilde{H}_i| \right) \\
&=: \sum_{i \in I} (D_{3,i} \cdot D_{4,i} + D_{5,i} \cdot D_{6,i} + D_{3,\beta,i} \cdot D_{4,\beta,i} + D_{5,\beta,i} \cdot D_{6,\beta,i}).
\end{aligned}$$

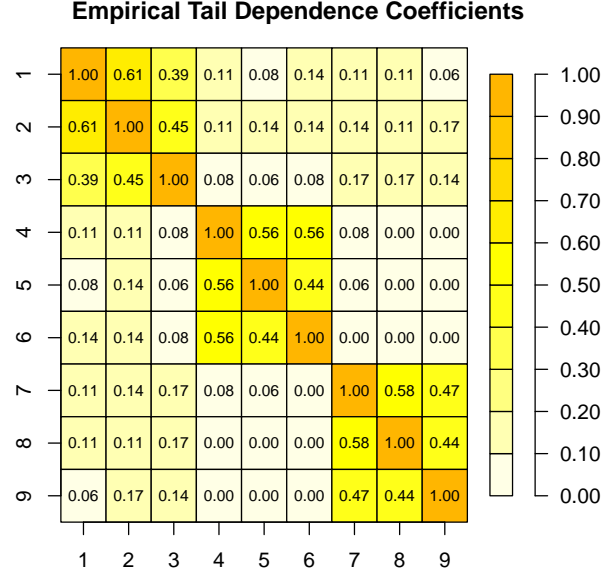


Figure 3: The empirical pairwise tail dependence coefficients  $\hat{\chi}_{ij}(u_i, u_j)$ ,  $1 \leq i, j \leq 9$  based on the empirical 95%-quantiles as thresholds. Here stations 1–3 belong to the region in the northwest, stations 4–6 to the region in the south and stations 7–9 to the region in the northeast of France.

Since  $c_{s_i}$  and  $c_{\beta_i}$ ,  $i \in I$ , are continuous on the compact set  $\partial B_1^*(\mathbf{0}) \times A'_{\delta, I} \times K$ , we have  $D_{3,i} < \infty$ ,  $i \in I$ , and  $D_{3,\beta,i} < \infty$  while  $D_{4,i} \rightarrow_p 0$ ,  $i \in I$  and  $D_{4,\beta,i} \rightarrow_p 0$  by Lemma 23 and Thm. 8, respectively. Furthermore, we have

$$\sup_{\mathbf{v} \in B_1^+(\mathbf{0}), p \in K} \left\| \begin{pmatrix} \xi_n \\ \psi_n \end{pmatrix} - \begin{pmatrix} \mathbf{1}_I \\ \mathbf{1}_I \end{pmatrix} \right\| \leq \left\| \begin{pmatrix} \frac{u_n}{\bar{\mathbf{X}}_{k_n, n}^* \circ \mathbf{1}_I} \\ \frac{\alpha}{\bar{\alpha}_{n, k_n}} \circ \mathbf{1}_I \end{pmatrix} - \begin{pmatrix} \mathbf{1}_I \\ \mathbf{1}_I \end{pmatrix} \right\| \rightarrow_p 0$$

by (53). Consequently, using the uniform continuity of  $c_{s_i}$  and  $c_{\beta_i}$ ,  $i \in I$ , we obtain that  $D_{5,i} \rightarrow_p 0$  and  $D_{5,\beta,i} \rightarrow_p 0$ . The facts that  $D_{6,i} < \infty$  a.s.,  $i \in I$ , and  $D_{6,\beta,i} < \infty$  a.s. finally yield

$$D_2 = \sup_{\mathbf{v} \in \partial B_1^+(\mathbf{0}), p \in K} \left| C_n(\mathbf{v}_I, p) + \sum_{i \in I} c_{s_i}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p) G^0(\mathbf{1}_{\{i\}}) - c_{\beta_i}(\mathbf{v}_I, \mathbf{1}_I, \mathbf{1}, p) \alpha_i \tilde{H}_i \right| \rightarrow_p 0.$$

□

## E Further Analysis of the Precipitation Data Considered in Subsection 5.3

In this section, we further analyze the precipitation data in the fall season in different regions in France as considered in Subsection 5.3. First, we estimate the pairwise upper tail dependence coefficients

$$\chi_{ij} = \lim_{u \rightarrow \infty} \mathbb{P}(X_i > u \mid X_j > u), \quad 1 \leq i, j \leq 9$$

empirically via

$$\hat{\chi}_{ij}(u_i, u_j) = \frac{\sum_{l=1}^n \mathbf{1}\{X_{li} > u_i, X_{lj} > u_j\}}{\sum_{l=1}^n \mathbf{1}\{X_{lj} > u_j\}}, \quad 1 \leq i, j \leq 9$$

for large thresholds  $u_1, \dots, u_9$ . Here, we choose the componentwise empirical 95%-quantiles as thresholds. The results are displayed in Figure 3. Here, the stations within the same region exhibit moderate dependence with empirical tail dependence coefficients approximately between 0.4 and 0.6, while all pairs of stations in different regions have empirical tail dependence less than 0.2, i.e., there is only weak extremal dependence across different regions. However, the three regions cannot be assumed to be asymptotically independent. In particular, there is still some weak dependence between the two regions in the northwest and the northeast.

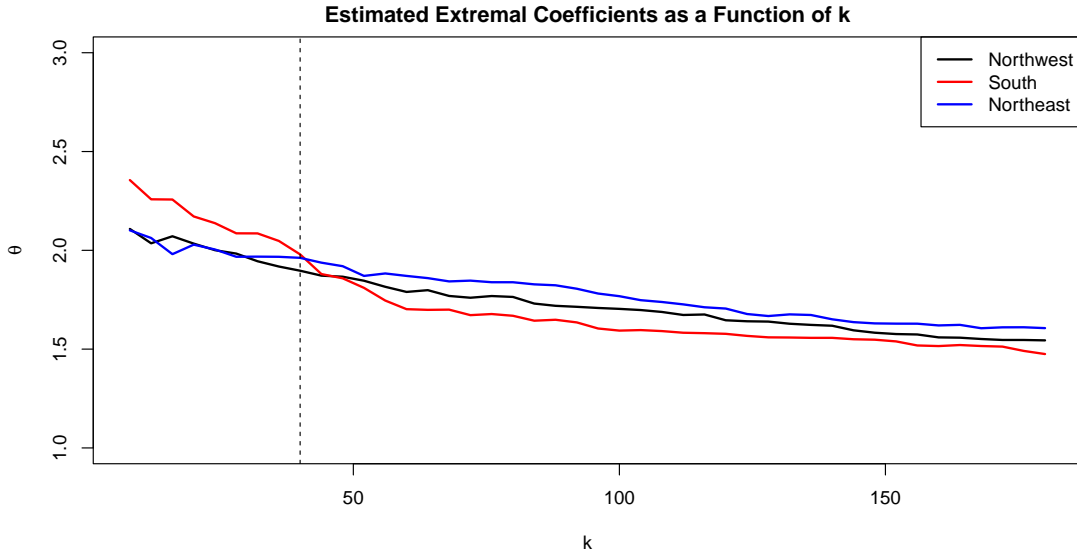


Figure 4: The estimated extremal coefficients  $1/\hat{\tau}_{I,n}^{MU}$  for the three regions in the northwest, the south and the northeast of France and different choices of  $k$ .

Furthermore, as detailed in Subsection 5.3, we apply our estimator for unknown margins  $\hat{\tau}_{I,n}^{MU}$  separately to each of the three regions. While the results in Table 6 are obtained for fixed  $k = 40$ , we also analyzed the effect of the choice of  $k$ . For different  $k$  from  $k = 8$  to  $k = 180$  (corresponding marginal to 1%–25% of the observations), the results are shown in Figure 4. In all cases, the estimated extremal coefficients grow slightly, i.e. dependence weakens, as the threshold grows (which means that  $k$  gets smaller), a phenomenon that is often observed in environmental data. This effect is even a bit more pronounced for the region in the south of France.