



HAL
open science

Estimation of the Multivariate Conditional-Tail-Expectation for extreme risk levels: illustrations on environmental data-sets

Elena Di Bernardino, Clémentine Prieur

► **To cite this version:**

Elena Di Bernardino, Clémentine Prieur. Estimation of the Multivariate Conditional-Tail-Expectation for extreme risk levels: illustrations on environmental data-sets. 2018. hal-01524536v3

HAL Id: hal-01524536

<https://hal.science/hal-01524536v3>

Preprint submitted on 19 Mar 2018

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 Multivariate Conditional-Tail-Expectation for extreme risk levels: illustration on environmental data-sets

Elena Di Bernardino*and Clémentine Prieur†

19th March 2018

Abstract

This paper deals with the problem of estimating the Multivariate version of the Conditional-Tail-Expectation introduced by [Di Bernardino et al. \(2013\)](#) and [Cousin and Di Bernardino \(2014\)](#). We propose a new semi-parametric estimator for this risk measure, essentially based on statistical extrapolation techniques, well designed for extreme risk levels. We prove a central limit theorem for the obtained estimator. We illustrate the practical properties of our estimator on simulations. The performances of our new estimator are discussed and compared to the ones of the empirical Kendall's process based estimator, previously proposed in [Di Bernardino and Prieur \(2014\)](#). We conclude with two applications on real data-sets: rainfall measurements recorded at three stations located in the south of Paris (France) and the analysis of strong wind gusts in the north west of France.

Keywords: Multivariate extreme value theory, multivariate risk measures, central limit theorem, hydrological applications. 62H12; 62H05; 60G70.

Introduction

Multivariate risk-measures Modeling and quantifying uncertainties related to extreme events is of main interest in environmental sciences (hydrological extreme events, cyclonic intensity, storm surges, ...).

Most of the time, environmental risks involve several aleas which are often correlated. A flood, e.g., can be described by three main characteristics: the peak flow, the volume and the duration. As these three quantities are correlated, it is important to define and to estimate the risk in a multivariate setting. For the same reasons, the design of facilities installed alongside of rivers should be based on multivariate extreme value analysis. In the case where the installation lies downstream the confluence of two rivers, neglecting in the risk analysis

*Conservatoire National des Arts et Métiers, Paris, Laboratoire CEDRIC EA4629, 292 rue Saint-Martin, Paris Cedex 03, France; corresponding author. Email: elena.di_bernardino@cnam.fr

†Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France

the correlations between both rivers may lead to an over- or under-estimation of the risk, involving either unnecessary costs or the construction of unsafe dams, with potentially dramatic consequences.

The classical univariate frequency analysis in environmental sciences focuses on the estimation of the probably most popular risk measure: the *return level*. A return level with a return period (RP) of $T = 1/p$ years is a threshold z_p whose probability of exceedance is p . The return period is traditionally defined as “the average time elapsing between two successive realizations of a prescribed event” (Singh et al. (2007)). An alternative, that takes into account the intensity of an event above a given threshold, is the mean excess function, i.e., $\mathbb{E}(X | X \geq z_p)$, with X the variable of interest (flows, rainfall, temperature, ...). This measure is also known as the Conditional-Tail-Expectation (CTE).

As already mentioned, it is often insufficient to consider a single real measure to quantify risks. This is therefore challenging for practitioners to estimate a multivariate return period and to select a specific design event starting from multivariate dangerous hydrological situations. However, the notion of return period in the multivariate setting is not univalent (see for instance Vandenberghe et al. (2012)). From the years 2000, different multivariate risk measures have been indeed introduced. Recently, level-curves and level sets associated to the multivariate risk vector have been proposed as risk measures in multivariate hydrological models because of their many advantages: they are simple, intuitive, interpretable and probability-based (see Chebana and Ouarda (2011), de Haan and Huang (1995)).

The problem of consistent estimation of the univariate quantile-based risk-measures has received attention in literature essentially in the univariate case. There are less papers on the estimation of multivariate risk-measures, due to a number of theoretical and practical reasons. Recently, a *conditional return level* estimator was proposed in Di Bernardino and Palacios-Rodríguez (2016). An estimation of extreme *Component-wise Excess design realization* (δ_{CE}) was also proposed in Di Bernardino and Palacios-Rodríguez (2017). This measure was introduced and used for hydrological applications in Salvadori et al. (2011). Salvadori et al. (2014) provided practical guidelines for coastal and off-shore engineering by using the δ_{CE} risk measure. In Salvadori et al. (2013), multivariate RP were estimated by using a semi-parametric approximation of Kendall’s distribution function. A *directional multivariate quantile* was proposed in Torres et al. (2017) to detect multivariate extremes in environmental phenomena.

Multivariate CTE under study In the following we deal with a version of the multivariate CTE, previously proposed by Di Bernardino et al. (2013) (see also Cousin and Di Bernardino (2014), Di Bernardino and Prieur (2014)). Let $I = \{1, \dots, d\}$. It is constructed as the conditional expectation of a d -dimensional vector of risks $\mathbf{X} = (X^1, X^2, \dots, X^d)$ following the distribution function F , given that the associated multivariate probability integral transformation $Z := F(\mathbf{X})$ is large. More precisely, for $i \in I$, we will consider the multivariate Conditional-Tail-Expectation:

$$\mathbb{E}[X^i | Z > Q_Z(1 - p)], \quad \text{for } p \in (0, 1), \quad (1)$$

where Q_Z is the quantile function of Z and p is *small enough*. Then, we take the conditional expectation of \mathbf{X}

conditionally to the fact that it belongs to the joint risk scenario $\{\mathbf{x} \in \mathbb{R}^d : F(\mathbf{x}) \geq 1 - p\}$.

Remark that our measure is based on the Kendall's distribution function $K(t) = \mathbb{P}[Z \leq t]$, for $t \in [0, 1]$. For this reason, risk measure in (1) can be used for hydrological risk management in the same vein of the multivariate RP in Salvadori et al. (2013) and Salvadori et al. (2011). Furthermore, conversely to risk measures in Cai and Li (2005), Bargès et al. (2009) and Landsman and Valdez (2003), the multivariate risk measure proposed in (1) is a non-aggregated risk measure. Indeed, hydrological variables can be of different nature (e.g. precipitation, temperature, discharge, ...), prohibiting the aggregation of the various components.

Statistical inference for multivariate CTE in (1) Some consistent estimator of the multivariate risk measure $\mathbb{E}[X^i | Z > t]$, for fixed $t \in (0, 1)$, has been provided by Di Bernardino et al. (2013), who proposed a plug-in estimator based on the consistent estimation of the whole level sets associated to the vector of risks \mathbf{X} . As the level sets are not compact, their estimation procedure requires the choice of an increasing truncation sequence $(T_n)_{n \geq 1}$. Making the “best choice” for $(T_n)_{n \geq 1}$ is not trivial (see Di Bernardino et al. (2013)). Recently, Di Bernardino and Prieur (2014) proposed a non-parametric estimator for $\mathbb{E}[X^i | Z > t]$ based on the estimation of the Kendall's process. For this estimator they provide a functional central limit theorem without requiring the calibration of extra parameters or sequences. However, a global good performance of this estimator is illustrated only for moderate to high (but not extreme) fixed risk levels t (see Di Bernardino and Prieur (2014)).

Conversely, in the present paper, we will develop a consistent estimation procedure to estimate the multivariate CTE defined in (1) for extreme risk levels (that is for $p < 1/n$, where n is the sample size). Our estimator is based on the bivariate inferential procedure proposed in Cai et al. (2015) for the estimation of the Marginal Expected Shortfall. However, a main difference relies on the fact that our conditioning random variable Z in (1) is a latent variable, which is not observed and has to be estimated.

Organization of the paper In Section 1, we introduce some notation, tools and preliminary assumptions. In Section 2.1, we propose our estimation procedure for the multivariate CTE defined in (1), based on Extreme Value Theory. In Section 2.2 we establish the asymptotic normality for the proposed semi-parametric estimator. The practical properties of our estimator are further investigated, and compared to the ones of alternative empirical approaches on simulated data-sets in Section 3. Finally in Section 4, we consider two 3-dimensional real data-sets: first a rainfall data-set, then a wind gusts data-set. For the sake of clarity of presentation, proofs and auxiliary results are postponed to Appendix.

1 Preliminaries and notation

Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. Let $\mathbf{X} = (X^1, X^2, \dots, X^d)$ be a d -dimensional positive¹ random vector with distribution function F . Define $Z = F(\mathbf{X})$ and the associated multivariate Kendall distribution function $K(t) = \mathbb{P}[Z \leq t]$,

¹In the following, we restrict ourselves to \mathbb{R}_+^d . This choice is motivated essentially by our applications in environmental risk theory, where random variables consist in rainfall measurements (in *mm*), thus defined on a positive support. However the results in this paper can be adapted also in \mathbb{R}^d .

for $t \in [0, 1]$. For more details on the multivariate probability integral transformation the interested reader is referred to [Capéraà et al. \(1997\)](#), [Genest and Rivest \(2001\)](#), [Nelsen et al. \(2003\)](#), [Genest et al. \(2006\)](#) and [Belzunce et al. \(2007\)](#).

As a consequence of Sklar's Theorem, the Kendall distribution only depends on the dependence structure or the copula function C associated with \mathbf{X} (see [Sklar \(1959\)](#)). Thus, we also have $K(t) = \mathbb{P}[C(\mathbf{V}) \leq t]$, where $\mathbf{V} = (V_1, \dots, V_d)$ with uniform marginals $V_1 = F_{X^1}(X^1), \dots, V_d = F_{X^d}(X^d)$. The analytical formulation of the Kendall distribution is in general not available. However, for the particular case of multivariate Archimedean copulas, it can be derived explicitly (see [Section 3](#)).

Let $U_Z = (\frac{1}{1-K})^{\leftarrow}$ be the tail quantile function of Z , where \leftarrow denotes the left-continuous inverse. In this paper we aim to estimate the quantity

$$\theta_p^i := \mathbb{E}[X^i | Z > U_Z(1/p)] \quad \text{for } p \in (0, 1) \text{ and } i \in I,$$

on independent and identically distributed (*i.i.d.*) d -dimensional observations, $(\mathbf{X}_j)_{j=1, \dots, n}$ from F , for small values of p in a sense detailed below (see [Section 2.1](#), in particular [Equation \(7\)](#)).

To estimate θ_p^i we make assumptions both on the right-hand tail of X^i and on the right-hand upper tail dependence of X^i and Z . Then, we expect that high values of Z correspond to high values of X^i . These assumptions are classical in multivariate Extreme Value Theory (EVT). We begin with describing the right-hand upper tail dependence. In the whole paper we will suppose that, for all $(x, z) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}$, and for all $i \in I$, the following limits exist:

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left[1 - F_i(X^i) \leq \frac{x}{t}, 1 - K(Z) \leq \frac{z}{t} \right] =: R_{(X^i, Z)}(x, z). \quad (2)$$

Function $R_{(X^i, Z)}$ in [\(2\)](#) completely determines the so-called stable tail dependence function $l_{(X^i, Z)}$, as for all $x, z \geq 0$, $l_{(X^i, Z)}(x, z) = x + z - R_{(X^i, Z)}(x, z)$, (see, e.g., [Drees and Huang \(1998\)](#), [Beirlant et al. \(2004\)](#)). As analysed below, the case of asymptotic independence, i.e., $R_{(X^i, Z)} \equiv 0$, will not be included in our main central limit theorem result (see [Theorem 2.1](#) and [Remark 4](#)).

Let $i \in I$. For the marginal distribution F_i we assume that X^i follows a distribution with a heavy right tail, i.e., $\exists \gamma^i > 0$ such that $\forall x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U_i(tx)}{U_i(t)} = x^{\gamma^i}, \quad (3)$$

where $U_i = (\frac{1}{1-F_i})^{\leftarrow}$ and γ^i is the extreme tail index associated to F_i .

Suppose that [\(3\)](#) holds true with $\gamma^i \in (0, 1)$. In the following, we introduce different Gaussian processes that will be useful to state our main asymptotic normality result (see [Theorem 2.1](#)). Let W_R be a zero mean Gaussian process on $[0, \infty]^2 \setminus \{(\infty, \infty)\}$ with covariance structure

$$\mathbb{E}[W_{R_{(X^i, Z)}}(x_1, z_1) W_{R_{(X^i, Z)}}(x_2, z_2)] = R_{(X^i, Z)}(x_1 \wedge x_2, z_1 \wedge z_2).$$

Let $i \in I$ and $q \in [0, +\infty) \cup \{+\infty\}$. Let $(\Theta^i, \Gamma^i(q))^\top$ (with \mathbf{v}^\top the transpose of vector \mathbf{v}) denote the bivariate process described by:

$$\Theta^i = (\gamma^i - 1) W_{R_{(X^i, Z)}(\infty, 1)} + \left(\int_0^\infty R_{(X^i, Z)}(s, 1) ds^{-\gamma^i} \right)^{-1} \int_0^\infty W_{R_{(X^i, Z)}}(s, 1) ds^{-\gamma^i}, \quad (4)$$

$$\Gamma^i(q) = \begin{cases} \frac{\gamma^i}{\sqrt{q}} \left(-W_{R_{(X^i, Z)}}(q, \infty) + \int_0^q s^{-1} W_{R_{(X^i, Z)}}(s, \infty) ds \right) & \text{for } q \in (0, +\infty), \\ N(0, (\gamma^i)^2) \text{ random variable independent of } W_R & \text{for } q = 0 \text{ or } q = +\infty. \end{cases} \quad (5)$$

For $0 < q < \infty$, classical computations, mainly based on Fubini's theorem and on the covariance structure of $W_{R_{(X^i, Z)}}$, lead to $\text{Var}(\Theta^i) = (\gamma^i)^2 - 1 - b^2 \int_0^\infty R_{(X^i, Z)}(s, 1) ds^{-2\gamma^i}$, $\text{Var}(\Gamma^i(q)) = (\gamma^i)^2$ and $\text{Cov}(\Gamma^i(q), \Theta^i) = \frac{\gamma^i}{\sqrt{q}}(1 - \gamma^i + \frac{b}{q^{\gamma^i}}) R_{(X^i, Z)}(q, 1) - \frac{\gamma^i(1-\gamma^i)b}{\sqrt{q}} \int_0^q R_{(X^i, Z)}(s, 1) s^{-1-\gamma^i} ds - \frac{\gamma^i(1-\gamma^i)}{\sqrt{q}} \int_0^q R_{(X^i, Z)}(s, 1) s^{-1} ds - \frac{\gamma^i b}{\sqrt{q}} \int_0^q R_{(X^i, Z)}(s, 1) \ln(\frac{q}{s}) s^{-1-\gamma^i} ds$, with $b = \left(\int_0^\infty R_{(X^i, Z)}(s, 1) ds^{-\gamma^i} \right)^{-1}$.

Processes $\{(\Theta^i, \Gamma^i(q))^\top, i = 1, \dots, d\}$ appear as limiting processes in Theorem 2.1 of Section 2 below.

2 Main results

2.1 Estimation procedure

Let $i \in I$, $n_1, n_2 \in \mathbb{N}^*$ and the sample size $n := n_1 + n_2$. We consider $(\mathbf{X}_j)_{j=1, \dots, n}$ a d -dimensional *i.i.d.* sample of \mathbf{X} . For all $c \in \mathbb{R}_+^d$ we define the d -dimensional empirical distribution function of \mathbf{X} based on n_2 observations as $F_{n_2}(c) = \frac{1}{n_2} \sum_{j=n_1+1}^{n_1+n_2} 1_{\{\mathbf{X}_j \leq c\}}$. For all $j = 1, \dots, n_1$ we define $Z_j = F(\mathbf{X}_j)$ and $\tilde{Z}_j = F_{n_2}(\mathbf{X}_j)$.

Following classical Weissman-type extrapolation technique (see Weissman (1978)), we construct an estimator of θ_p^i by a two-stage approach. Let $k = k(n_1)$ be an intermediate sequence of integers which satisfies $k \rightarrow \infty$ and $k/n_1 \rightarrow 0$, as $n_1 \rightarrow \infty$. Firstly, we consider the estimation of $\theta_{\frac{k}{n_1}}^i$, i.e. the CTE at an intermediate (not extreme) probability level $\frac{k}{n_1}$. We can estimate non-parametrically $\theta_{\frac{k}{n_1}}^i$ by taking the empirical average of the X^i of those selected observations:

$$\hat{\theta}_{\frac{k}{n_1}, n_2}^i = \frac{1}{k} \sum_{j=1}^{n_1} X_j^i 1_{\{\tilde{Z}_j > \tilde{Z}_{n_1-k, n_1}\}}, \quad (6)$$

where \tilde{Z}_{n_1-k, n_1} is the $(n_1 - k)$ -th order statistic of $\tilde{Z}_1, \dots, \tilde{Z}_{n_1}$.

Secondly, using an extrapolation method based on Proposition 1 in Cai et al. (2015) applied to the bivariate vector (X^i, Z) and Equation (3), we have that, for $n_1 p = o(k)$ as $n_1 \rightarrow \infty$,

$$\theta_p^i \sim \frac{U_i(1/p)}{U_i(n_1/k)} \theta_{\frac{k}{n_1}}^i \sim \left(\frac{k}{n_1 p} \right)^{\gamma^i} \theta_{\frac{k}{n_1}}^i. \quad (7)$$

In order to apply the asymptotic approximation in Equation (7), we need to estimate the tail index γ^i . To this aim, we will consider the Hill estimator (see Hill (1975)), i.e.

$$\hat{\gamma}^i = \frac{1}{k_i} \sum_{j=1}^{k_i} \ln(X_{n_1-j+1, n_1}^i) - \ln(X_{n_1-k_i, n_1}^i), \quad (8)$$

where $k_i = k_i(n_1)$ is an intermediate sequence of integers and X_{j,n_1}^i , for $j = 1, \dots, n_1$, is the j -th order statistic of $X_1^i, \dots, X_{n_1}^i$. Finally, using Equations (6), (7) and (8), we estimate θ_p^i by

$$\widehat{\theta}_{p(n_1), n_2}^i = \left(\frac{k}{n_1 p} \right)^{\widehat{\gamma}^i} \widehat{\theta}_{\frac{k}{n_1}, n_2}^i. \quad (9)$$

The asymptotic normality of our estimator in (9) is stated in Theorem 2.1 below, for $n_1, n_2 \rightarrow \infty$. The limit process can be written as a combination of processes Θ^i and $\Gamma^i(q)$ in Equations (4) and (5) respectively. In particular, the process Θ^i plays a central role to describe the asymptotic behavior of $\widehat{\theta}_{\frac{k}{n_1}, n_2}^i$ (see Proposition 2.1). The process $\Gamma^i(q)$ is related to the asymptotic behavior of $\widehat{\gamma}^i$.

Remark 1 (On the split-up of the observation in two samples). To conclude this section, we notice that the split-up of the data in two samples of sizes n_1 and n_2 , respectively, seems artificial. Indeed, one would like to consider the estimator $\check{\theta}_{\frac{k}{n}}^i = \frac{1}{k} \sum_{j=1}^n X_j^i 1_{\{\widehat{Z}_j > \widehat{Z}_{n-k,n}\}}$, where $\widehat{Z}_{n-k,n}$ is the $(n-k)$ -th order statistic of $\widehat{Z}_1, \dots, \widehat{Z}_n$, with for $j = 1, \dots, n$ and $\widehat{Z}_j = F_n(\mathbf{X}_j)$. Note that the split-up of the data is a technical requirement in our paper for proving consistency results. Indeed, the split-up allows constructing multivariate pseudo-samples that are independent of the Hill estimator $\widehat{\gamma}^i$. To our knowledge, the asymptotic study of $\check{\theta}_{\frac{k}{n}}^i$ remains an open issue, despite recent studies, e.g., the one in van der Vaart and Wellner (2007).

2.2 Asymptotic normality

We now characterize the limit distribution of $\widehat{\theta}_{p(n_1), n_2}^i$ in Equation (9). The proof of our main result requires the following conditions.

Assumption 2.1. [Assumptions for the central limit Theorem 2.1]

(a.1) There exist $\beta > \max_{i \in I} \gamma^i$ and $\tau < 0$ such that, for any $i \in I$, as $t \rightarrow \infty$,

$$\sup_{\{0 < x < \infty, 1/2 \leq z \leq 2\}} \frac{|t \mathbb{P} [1 - F_i(X^i) \leq \frac{x}{t}, 1 - K(Z) \leq \frac{z}{t}] - R_{(X^i, Z)}(x, z)|}{x^\beta \wedge 1} = O(t^\tau).$$

(a.2) The Kendall distribution function $K(t)$, $t \in [0, 1]$ of $Z = F(\mathbf{X})$ admits a continuous density $K'(t)$ on $(0, 1]$.

(a.3) There exist $p_0 < \frac{1}{\max_{i \in I} \gamma^i}$, $1/p_0 + 1/q_0 = 1$ and $\epsilon > 0$ such that $\sup \left\{ \frac{n_1}{\sqrt{k}} n_2^{-\frac{-1+\epsilon}{2}}, \sqrt{n_1} \left(\frac{k}{n_1} \right)^{\frac{1}{p_0} - \frac{1}{2}} n_2^{-\frac{-1+\epsilon}{2q_0}} \right\} \rightarrow 0$, as $n_1, n_2 \rightarrow \infty$.

(b) For $i \in I$, there exist $\rho_i < 0$ and an eventually positive or negative function A_i such that as $t \rightarrow \infty$, $A_i(tx)/A_i(t) \rightarrow x^{\rho_i}$ for all $x > 0$ and $\sup_{x > 1} |x^{-\gamma^i} \frac{U_i(tx)}{U_i(t)} - 1| = O(A_i(t))$.

(c) For $i \in I$, as $n_1 \rightarrow \infty$, $\sqrt{k_i} A_i(n_1/k_i) \rightarrow 0$, where $k_i(n_1)$ is the intermediate sequence of integers in (8).

(d) For $i \in I$, as $n_1 \rightarrow \infty$, $k = O(n_1^\alpha)$ for some $\alpha < \min \left(\frac{-2\tau}{-2\tau+1}, \frac{2\gamma^i \rho_i}{2\gamma^i \rho_i + \rho_i - 1} \right)$, where $k(n_1)$ is the intermediate sequence of integers in (6).

Under assumptions presented above, we can state a central limit theorem for our estimator $\widehat{\theta}_{p(n_1), n_2}^i$.

Theorem 2.1. Let $i \in I$ and $p = p(n_1) \rightarrow 0$, for $n_1 \rightarrow \infty$. Assume that Assumptions (a.1)-(d) hold true and $\gamma^i \in (0, 1/2)$. Assume $d_{n_1} := \frac{k}{n_1 p} \geq 1$, $r := \lim_{n_1 \rightarrow \infty} \frac{\sqrt{k} \ln(d_{n_1})}{\sqrt{k_i}} \in [0, +\infty]$ and $q = \lim_{n \rightarrow \infty} \frac{k_i}{k} \in [0, +\infty]$. If $\lim_{n_1 \rightarrow \infty} \frac{\ln(d_{n_1})}{\sqrt{k_i}} = 0$, then for $n_1, n_2 \rightarrow \infty$,

$$v_{n_1} \left(\frac{\hat{\theta}_{\frac{p(n_1), n_2}{n_1}}^i}{\theta_{\frac{p(n_1)}{n_1}}^i} - 1 \right) \rightarrow \begin{cases} \Theta^i + r \Gamma^i(q), & \text{if } r \leq 1, \\ \frac{1}{r} \Theta^i + \Gamma^i(q), & \text{if } r > 1, \end{cases}$$

where $v_{n_1} = \min(\sqrt{k}, \frac{\sqrt{k_i}}{\ln(d_{n_1})})$, with $(\Theta^i, \Gamma^i(q))^\top$ defined in Section 1 (see Equations (4) and (5)).

Remark 2. Recall that we are interested in the multivariate CTE for very small risk levels p . It is thus not possible to use classical empirical estimators. We therefore turn to extrapolation techniques. The assumption $\lim_{n_1 \rightarrow \infty} \frac{\ln(d_{n_1})}{\sqrt{k_i}} = 0$ indicates that we should not extrapolate too far.

Remark 3. We now comment Items (a.1), (a.2), (a.3), (b), (c) and (d) in Assumption 2.1 above:

- Item (a.1) is a second order assumption, which allows to quantify the rate of convergence in (2) (see also Condition (7.2.8) in de Haan and Ferreira (2006) and Condition (2.1) in Draisma et al. (2004)). Note that the constants β and τ in (a.1) do not depend on $i \in I$.
- Item (a.2) is a regularity assumption on the Kendall density. It is satisfied for a large class of multivariate distributions, as the class of Archimedean copulas, bivariate extreme copulas, Farlie-Gumbel-Morgenstern copulas. It is required to describe the asymptotic behavior of the Kendall process (see, e.g., Barbe et al. (1996)).
- As already mentioned, the conditioning random variable Z in (1) is a latent variable, which is not observed and has to be estimated. Item (a.3) explains how large the fraction n_2 of the initial sample used to estimate Z has to be chosen not to affect the asymptotic behavior in Proposition A.1.
- Item (b) allows to quantify the rate of convergence in (3) (see also Condition (3.2.4) in de Haan and Ferreira (2006)).
- Item (c) ensures the convergence of the first and third terms (respectively $L_1^{n_1}$ and $L_3^{n_1}$) in the proof of Theorem 2.1.
- Practical application of the extreme value theory requires to select the tail sample fraction, i.e., the extreme values of the sample, that may contain most information on the tail behavior. Item (d) provides an asymptotic upper-bound on the fraction k .

The proof of Theorem 2.1 above is mainly based on arguments in de Haan and Ferreira (2006); Cai et al. (2015), as far as on Proposition 2.1 below. Both proofs of Proposition 2.1 and Theorem 2.1 are postponed to Appendix.

Proposition 2.1. Let $i \in I$. Under conditions of Theorem 2.1 for $n_1, n_2 \rightarrow \infty$, it holds that

$$\sqrt{k} \left(\frac{\hat{\theta}_{\frac{k}{n_1}, n_2}^i}{\theta_{\frac{k}{n_1}}^i} - 1 \right) \xrightarrow{\mathbb{P}} \Theta^i, \text{ with } \Theta^i \text{ defined in Equation (4).}$$

Remark 4 (Asymptotic independence and marginal tail behavior). We discuss here two main restrictions in the assumptions of Theorem 2.1.

1. Item (a.1) excludes the asymptotic independence, i.e., the case $R_{(X^i, Z)} \equiv 0$. Asymptotic dependence seems to be a necessary condition, as illustrated in Section 3.3, where we provide an example for which the assumption of asymptotic dependence is violated. The interested reader is also referred to Cai et al. (2015).
2. The assumption $\gamma^i \in (0, 1/2)$ is necessary for Theorem 2.1. A careful reading of the proof of auxiliary Proposition A.2 shows that the result does not hold true anymore when $\gamma^i = 1/2$. For the consistency of $\hat{\theta}_{p(n_1), n_2}^i$, this assumption can be relaxed to $\gamma^i \in (0, 1)$. Indeed, if we assume that (X^i, Z) satisfies (2), as far as Item (b) of Assumption 2.1, $R_{(X^i, \tilde{Z})}(1, 1) > 0$, $\lim_{n_1 \rightarrow \infty} \frac{\log(d_{n_1})}{\sqrt{k_i}} = 0$ and $\gamma^i \in (0, 1)$, then $\frac{\hat{\theta}_{p(n_1), n_2}^i}{\theta_{p(n_1)}^i} \xrightarrow{\mathbb{P}} 1$, for $n_1, n_2 \rightarrow +\infty$. In Section 3.3, we provide an example with $\gamma^i \notin (0, 1/2)$.

To conclude this section, note that the validation of the whole set of items in Assumption 2.1 on a real data-set is not an easy task. We describe in the introduction of Section 3 a graphical procedure to select the tail sample fractions k_i , for $i \in I$ and k (see also Cai et al. (2015)). The procedure is based on the following: a small tail sample fraction leads to a large variance in the estimation procedure. Conversely, for large values of the tail sample fraction, we observe a bias on the estimates. The graphical procedure helps in finding a bias-variance tradeoff. It is also possible to estimate $R_{(X^i, Z)}$ as far as the tail indices γ^i , for $i \in I$ to test the asymptotic dependence and the assumption $\gamma^i \in (0, 1/2)$. To the best of our knowledge, there does not exist any convincing empirical testing for second order assumptions. In the real-data section (Section 4), we propose an empirical procedure to choose the split-up sample sizes n_1 and n_2 such that $n_1 + n_2 = n$, with n the total sample size.

3 Simulation Study

In this section, a simulation and comparison study is implemented to investigate the finite sample performances of our estimator of the multivariate CTE defined in (9). We focus on dimension $d = 2$, considering different dependence structures and marginal distributions. Later in Section 4, we will provide an analysis of two real data-sets in dimension $d = 3$.

Graphical representation of the performances

We draw 500 samples of size n from each probability distribution under study. Based on each sample, we estimate $\theta_{p(n_1)}^1$ and $\theta_{p(n_1)}^2$ for different values of n_1 and $p(n_1)$. In Figures 1, 2 and 6, we displayed on the left side the boxplots for the ratio between the estimates and the true values, i.e., $\hat{\theta}_{p(n_1), n_2}^i / \theta_{p(n_1)}^i$ for different structures of dependence and different marginal distributions, which will be described hereafter (see details in Sections 3.1 and 3.2). On the right side of these figures, we plotted Q-Q plots. More precisely, for $r < \infty$, Theorem 2.1 can be expressed as $\sqrt{k} \left(\frac{\hat{\theta}_{p(n_1), n_2}^i}{\theta_{p(n_1)}^i} - 1 \right) \rightarrow \Theta^i + r \Gamma^i(q)$, for $n_1, n_2 \rightarrow +\infty$. Note that the limit distribution is a centered normal distribution. Let $(\sigma_p^i)^2 := \frac{1}{k} \text{Var}(\Theta^i + r \Gamma^i(q))$. We compare, through corresponding Q-Q plots, the distribution of $\frac{1}{\sigma_p^i} \ln \left(\frac{\hat{\theta}_{p(n_1), n_2}^i}{\theta_{p(n_1)}^i} \right)$ with the limit distribution $\mathcal{N}(0, 1)$.

Empirical procedure for the selection of parameters $k(n_1)$, $k_1(n_1)$ and $k_2(n_1)$

Specific values for $k(n_1)$ and $k_i(n_1)$ are chosen for each sample size n_1 , accordingly to the selection procedure

described by [Cai et al. \(2015\)](#). Indeed, a usual practice to choose the intermediate sequence k_i , is to select a range corresponding to the first stable region of the estimator $\widehat{\gamma}^i$. Then, to gain in stability, we average the estimations $\widehat{\gamma}^i$ corresponding to $k_i(n_1)$ in the selected range. Similarly, for the sequence $k(n_1)$, we select a range corresponding to the first stable region of the final estimator $\widehat{\theta}_{p(n_1),n_2}^i$. Then, we average the estimations $\widehat{\theta}_{p(n_1),n_2}^i$ corresponding to $k(n_1)$ in the stability window.

Comparison with the empirical estimator

In [Section 3.1](#), we compare the performance of our estimator with the empirical one, i.e.,

$$\widehat{\theta}_{p,emp}^i = \frac{1}{[np]} \sum_{j=1}^n X_j^i 1_{\{\widetilde{Z}_j > \widetilde{Z}_{n-[np],n}\}}, \quad (10)$$

with $\widetilde{Z}_j = F_n(\mathbf{X}_j)$ and $\widetilde{Z}_{n-k,n}$ the associated order statistic (see [Figure 3](#)). This estimator is the empirical counterpart of θ_p^i .

Sensitivity to the choice of the split-up parameters and robustness with respect to the hypotheses

In [Section 3.1](#), we study the sensitivity of the estimation to the choice of the split-up parameters n_1 and n_2 (see [Figures 4](#) and [5](#)). Furthermore, in [Section 3.3](#), we investigate the performance of our estimator when assumptions required in [Theorem 2.1](#) are partially violated, namely the assumptions related to the asymptotic dependence and to the marginal tails (see [Figures 7](#) and [8](#), see also [Remark 4](#) in [Section 2.2](#)).

3.1 Copula 4.2.2 in [Nelsen \(1999\)](#)

In this section, we assume that the copula associated to the bivariate vector (X^1, X^2) is Copula 4.2.2 in [Nelsen \(1999\)](#), i.e., $C_{(X^1, X^2)}(s, t) = 1 - ((1-s)^\theta + (1-t)^\theta)^{\frac{1}{\theta}}$, for $\theta \in [1, +\infty)$. For $\theta = 1$, we get $C_{(X^1, X^2)}(s, t) = s + t - 1$ (i.e., counter-monotonicity copula); for $\theta = +\infty$, we get $C_{(X^1, X^2)}(s, t) = \min(s, t)$ (i.e., comonotonicity copula). Furthermore, for $\theta > 1$, we have

$$R_{(X^1, Z)}(x, z) = \begin{cases} z, & \text{if } x \geq \frac{\theta z}{\theta-1}, \\ x - \frac{x^\theta}{\theta^\theta \left(\frac{z}{\theta-1}\right)^{\theta-1}}, & \text{if } x < \frac{\theta z}{\theta-1}. \end{cases}$$

From now on, we fix $\theta = 2$. In this case $R_{(X^1, Z)}(1, 1) = 0.75$. We remark that $\forall \theta \in (1, +\infty)$, [Item \(a.1\)](#) in [Assumption 2.1](#) is satisfied for any $\beta > \gamma^1$ and any $\tau < 0$. Let us now describe the marginal models we consider for the simulations.

Model (A)

- X^1 follows a Hall/Weiss distribution (i.e., $F_{X^1}(x) = 1 - \frac{1}{2}x^{-\alpha}(1 + x^\rho)$) with parameters $\alpha = 4$ and $\rho = -4$;
- X^2 follows a Pareto distribution (i.e., $F_{X^2}(x) = 1 - (\frac{1}{x+1})^\alpha$) with parameter $\alpha = 3$.

These marginal distributions satisfy [Item \(b\)](#) in [Assumption 2.1](#) with tail indices $\gamma^1 = 1/4$, $\rho^1 = -4$ and $\gamma^2 = 1/3$, $\rho^2 = -1$ (the interested reader is referred to [Example 1](#) in [Hua and Joe \(2011\)](#) and [Example 5.1](#) in [Mao and Hu \(2012\)](#)).

Model (B)

- X^1 follows a Burr distribution (i.e., $F_{X^1}(x) = 1 - (1 + x^b)^{-a}$) with parameters $a = 2$ and $b = 2$;
- X^2 follows a Burr distribution with parameters $a = 3$ and $b = 2$.

These marginal distributions satisfy Item (b) in Assumption 2.1 with tail indices $\gamma^1 = 1/4$, $\rho^1 = -2$ and $\gamma^2 = 1/6$, $\rho^2 = -2$ (the interested reader is referred to Example 2 in Hua and Joe (2011)).

In Figure 1 (*resp.* 2), we show the boxplots drawn from 500 realizations for Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (A) (*rep.* (B)) for $n_1 = n_2$ and $n = n_1 + n_2 \in \{100, 500\}$. We also show the corresponding QQ-plots. Such a simulation study was also conducted for Copula 4.2.15 in Nelsen (1999). Since the results looked very similar, they are not included here.

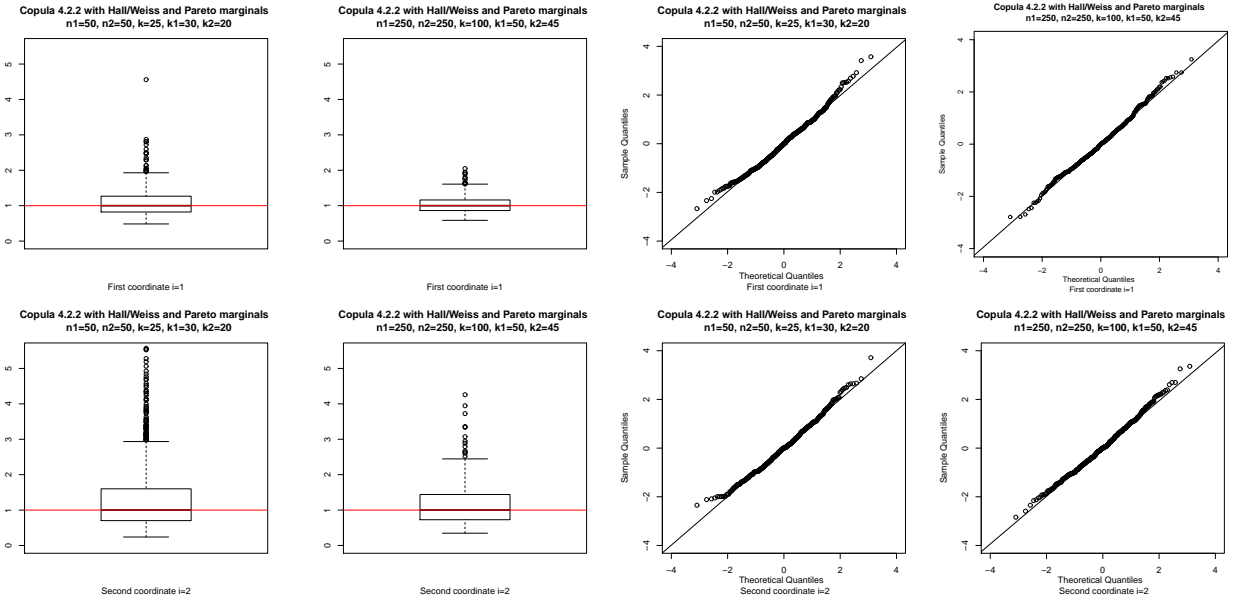


Figure 1: Boxplots and associated QQ-plots based on 500 realizations for Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (A), $p = 1/4 n_1$ for different sample sizes n_1 . **First row:** $\hat{\theta}_{p(n_1), n_2}^{i=1}$. **Second row:** $\hat{\theta}_{p(n_1), n_2}^{i=2}$.

The accuracy of the estimate of the marginal tail index γ^i is related to the parameter ρ^i . For a larger (in absolute value) ρ^i , we get a more accurate estimation (see, e.g., Section 4.3 in de Haan and Ferreira (2006)). In Figure 1, we observe much less variability for a similar bias for the estimation of $\theta_{p(n_1)}^1$ than for the estimation of $\theta_{p(n_1)}^2$. It is coherent with the values $\rho^1 = -4$ and $\rho^2 = -1$. It illustrates the influence of the marginal distributions on our estimator. In both Figures 1 and 2, we observe, as expected, that increasing the sample size n increases the accuracy of the estimation and the quality of the QQ-plots.

Comparison with the empirical estimator In the following, we compare the performance of our estimator with the empirical estimator $\hat{\theta}_{p, emp}^i$ defined in (10) for $p = 1/n$ and $n = \{200, 1500\}$. Clearly, $\hat{\theta}_{p, emp}^i$ is not designed to deal with small probabilities p such that $np < 1$. For that comparison, we consider Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (C) described just below.

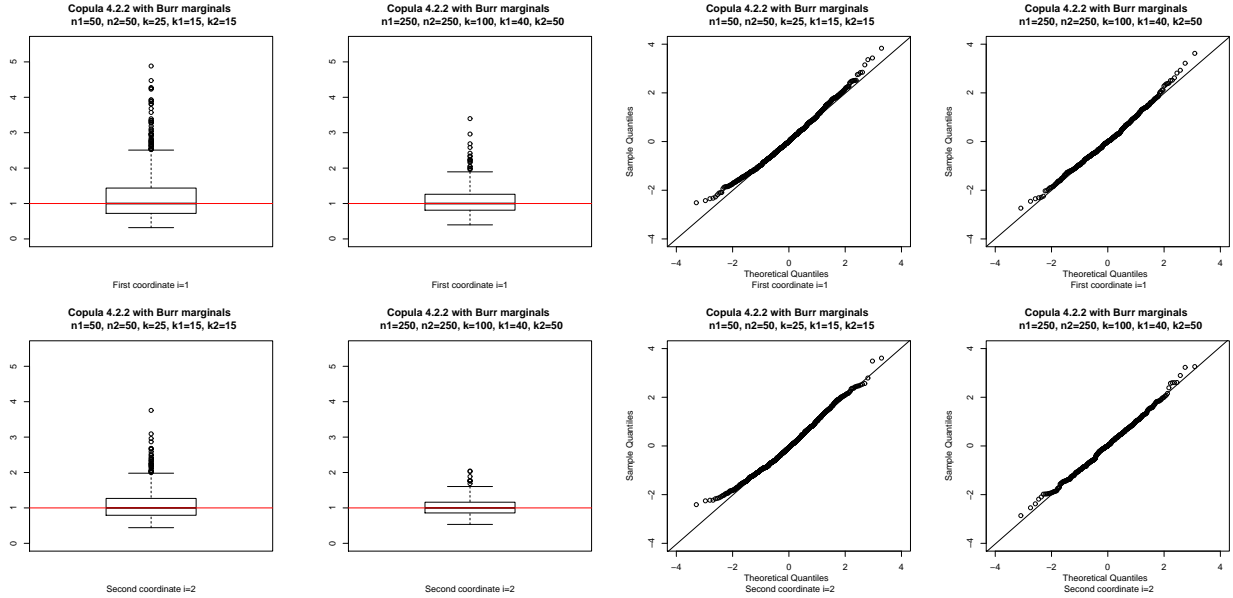


Figure 2: Boxplots and associated QQ-plots based on 500 realizations for Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (B), $p = 1/4 n_1$ for different sample sizes n_1 . **First row:** $\hat{\theta}_{p(n_1), n_2}^{i=1}$. **Second row:** $\hat{\theta}_{p(n_1), n_2}^{i=2}$.

Model (C)

- X^1 follows a Hall/Weiss distribution with parameters $\alpha = 4$ and $\rho = -4$ (i.e., $\gamma^1 = 1/4$, $\rho^1 = -4$);
- X^2 follows a Hall/Weiss distribution with parameters $\alpha = 6$ and $\rho = -6$ (i.e., $\gamma^2 = 1/6$, $\rho^2 = -6$).

Results are gathered in Figure 3. We remark that the empirical estimator $\hat{\theta}_{p, emp}^i$ underestimates the multivariate CTE and is consistently outperformed by our EVT estimator $\hat{\theta}_{p(n_1), n_2}^i$.

We also discuss the influence of the sample sizes n , n_1 and $n_2 = n - n_1$ on the quality of our estimator.

Sensitivity to the choice of the split-up parameters n_1 and n_2 We analyse the performance of our estimator for different choices of the total sample size n and of the splitting of the initial sample into two sub-samples of respective size n_1 and n_2 . The boxplots for the sensitivity analysis with respect to these parameters are presented in Figures 4 and 5. We consider Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and the marginal model (C). We observe on that example that for a total sample size of $n = 100$ (see Figure 4), our estimator of $(\theta_{p(n_1)}^1, \theta_{p(n_1)}^2)$ for $p(n_1) = 1/(2n_1)$ is biased if n_1 is not large enough. For both i coordinates (left and right), the bias becomes neglectable for $n_1 \geq 70$. For a total sample size of $n = 230$ (see Figure 5) and for both i coordinates (left and right), the bias becomes neglectable for $n_1 \geq 140$. If n_1 is too small, the Hill estimations $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are not accurate enough and introduce a bias in the estimation of $(\theta_{p(n_1)}^1, \theta_{p(n_1)}^2)$, with $p(n_1) = 1/(2n_1)$. We observe much less variability for a similar bias for the estimation of $\theta_{p(n_1)}^2$ than for the estimation of $\theta_{p(n_1)}^1$. It is coherent with the values $\rho^1 = -4$ and $\rho^2 = -6$.

Based on these simulated studies, in Section 4, an empirical procedure for the choice of the split-up parameters n_1 and n_2 will be proposed.

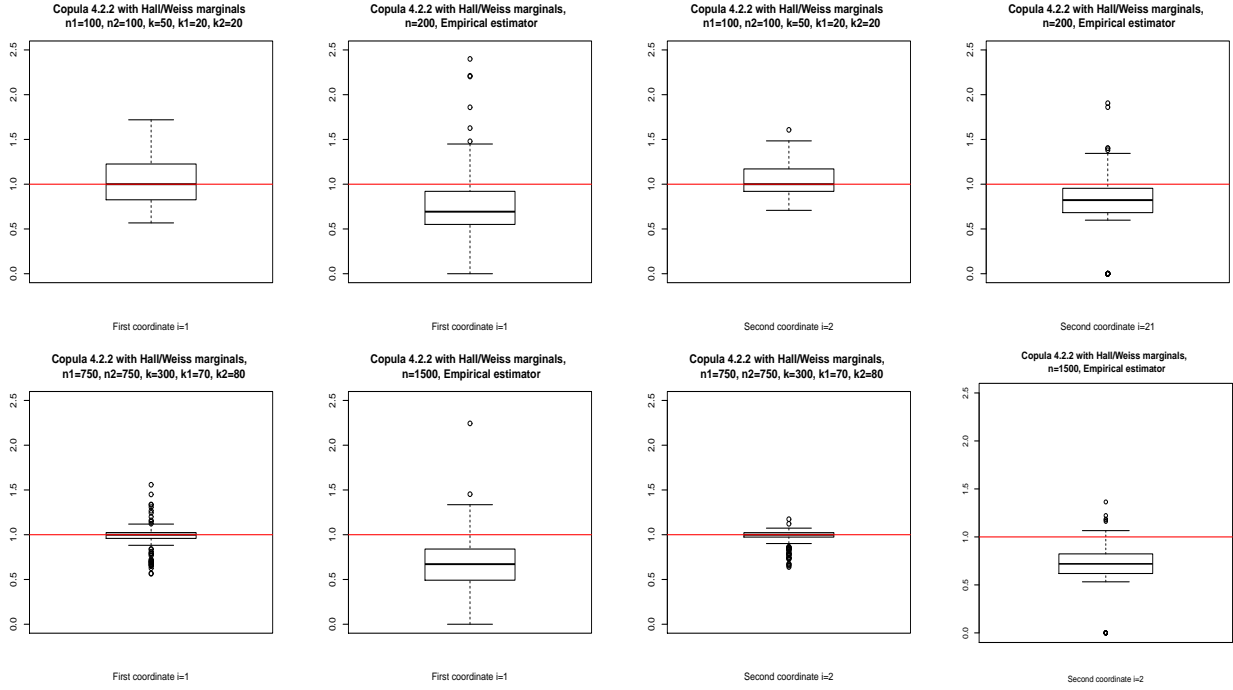


Figure 3: Boxplots on 500 Monte-Carlo simulations for Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (C), $p = 1/n$. **First row:** $n_1 = n_2 = 100$ ($n = 200$). **Second row:** $n_1 = n_2 = 750$ ($n = 1500$).

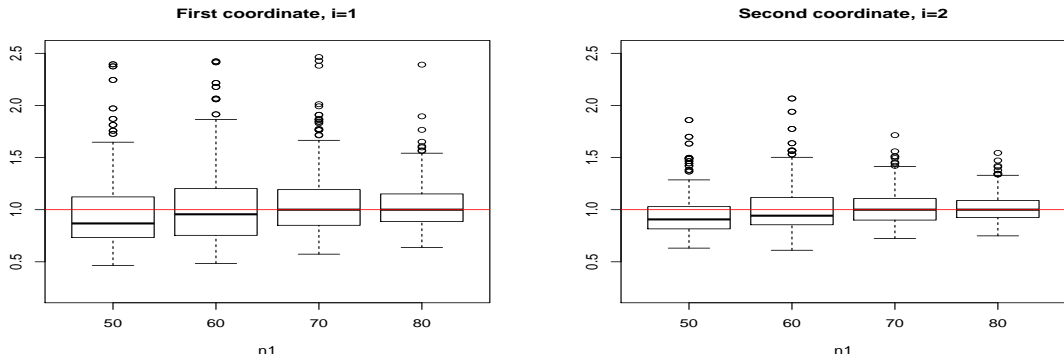


Figure 4: Boxplots based on 500 realizations for Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (C), $p = 1/2 n_1$ for $n_1 \in \{50, 60, 70, 80\}$. The sub-sample size $n_2 = n - n_1$ with $n = 100$.

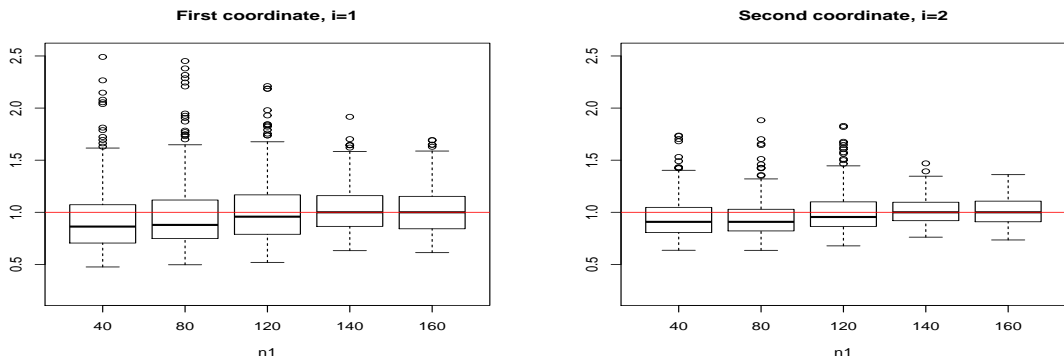


Figure 5: Boxplots based on 500 realizations for Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and marginal model (C), $p = 1/2 n_1$ for $n_1 \in \{40, 80, 120, 140, 160\}$. The sub-sample size $n_2 = n - n_1$ with $n = 230$.

3.2 Not-Archimedean example: HRT Copula

Let us consider the heavy right tail (HRT) copula (also called *Clayton survival copula* in the insurance and finance literature): $C(u, v) = u + v - 1 + ((1 - u)^{-1/\theta} + (1 - v)^{-1/\theta} - 1)^{-\theta}$, for $\theta > 0$.

This not-Archimedean copula has low correlation in the joint lower tail but high correlation in the joint upper tail. It was invented by Venter in 2001 to model the dependence on events of strong intensity (see, e.g., Section 3.8 in Gorge (2013)). Notice that in this case, we can not use the Archimedean generator in order to write the Kendall distribution. However, one can check the asymptotic dependence of the couple (X^i, \tilde{Z}) at least empirically, by estimating $\hat{R}_{(X^i, \tilde{Z})}(1, 1)$. We consider the previous marginal model (C). Results are gathered in Figure 6, and still show good performances for our estimator.

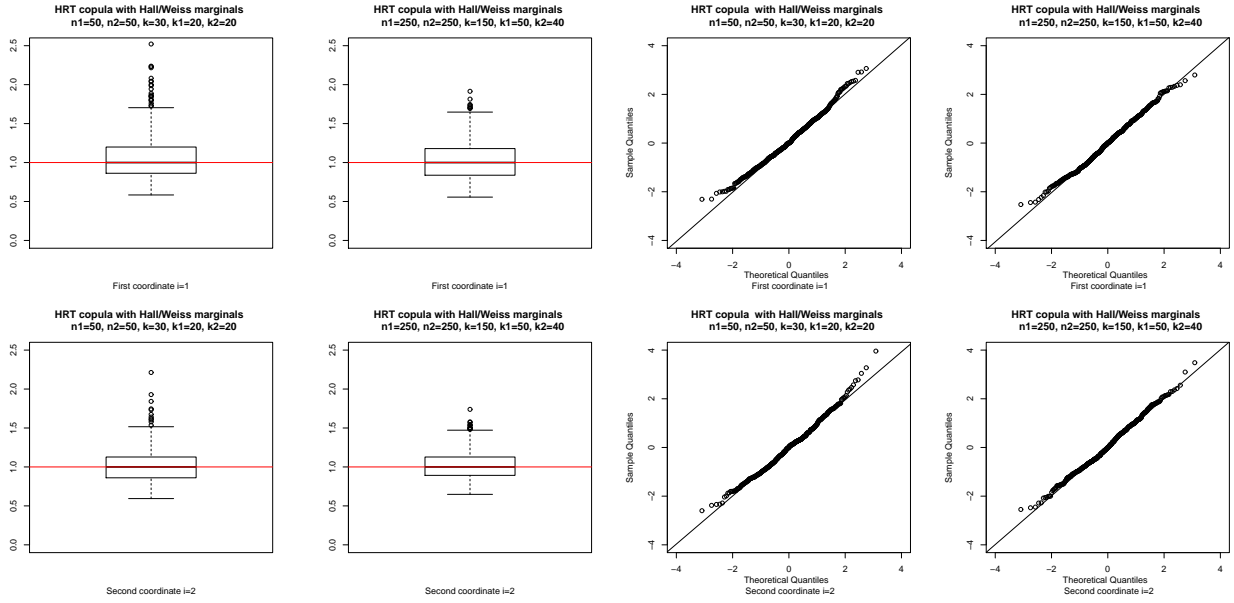


Figure 6: Boxplots and associated QQ-plots on 500 Monte-Carlo simulations for HRT copula with $\theta = 1$ and marginal model (C), $p = 1/4n_1$ for different sample sizes n_1 . **First row:** $\hat{\theta}_{p(n_1), n_2}^{i=1}$. **Second row:** $\hat{\theta}_{p(n_1), n_2}^{i=2}$.

3.3 Estimation when our assumptions are partially violated

In this section, we investigate the performance of our estimator when assumptions of Theorem 2.1 are partially violated.

Tail index $\gamma^i \notin (0, 1/2)$ We consider (X^1, X^2) with Copula 4.2.2 in Nelsen (1999) (see Section 3.1), with $\theta = 2$, and Pareto marginals $F(x) = 1 - (1/x)^{1/\gamma^i}$, with $\gamma^i = 1/2$, for $i = 1, 2$. Firstly, remark that the rate for the regular variation (3) is too fast then Item (b) of Assumption 2.1 does not hold true in this case. Furthermore, assumption $\gamma^i \in (0, 1/2)$ is violated for both components. However, our estimator $\hat{\theta}_{p(n_1), n_2}^i$ is still consistent (see Remark 4 in Section 2.2). The consistence is illustrated in Figure 7, where we present boxplots of the ratio of the estimates and the true values for $n_1 = n_2 = 750$ and different values of p .

Asymptotic independence We now consider the bivariate Independent copula with Pareto marginals

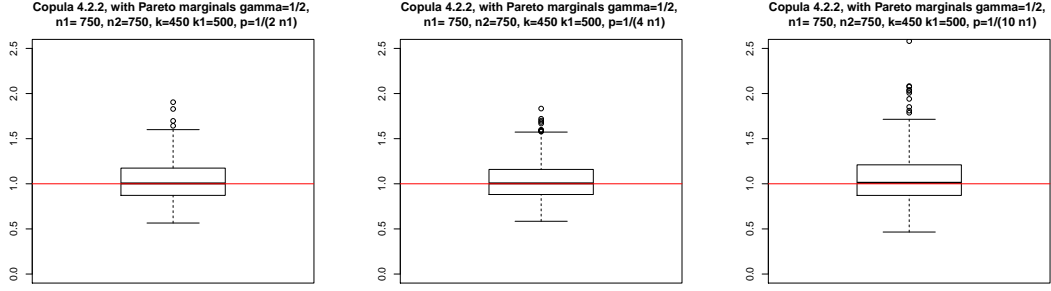


Figure 7: Copula 4.2.2 in Nelsen (1999) with $\theta = 2$ and Pareto marginals $F(x) = 1 - (1/x)^{\frac{1}{\gamma^i}}$ with $\gamma^1 = \gamma^2 = 1/2$. Here we take $n_1 = n_2 = 750$, different values of the risk level p and 500 Monte-Carlo simulations.

$F(x) = 1 - (1/x)^{\frac{1}{\gamma^i}}$ with $\gamma^1 = \gamma^2 = 1/4$. Note that $R_{(X^i, Z)} \equiv 0$, then this distribution does not satisfy Item (a.1) in Assumption 2.1. In this case, our EVT estimator overestimates the theoretical multivariate CTE (see Figure 8). For a similar behavior the reader is also referred to Figure 3 in Cai et al. (2015).

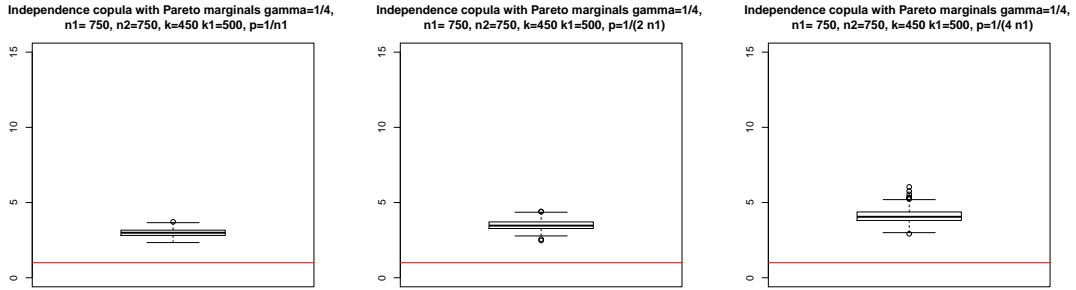


Figure 8: Independence copula and Pareto distributed marginals $F(x) = 1 - (1/x)^{\frac{1}{\gamma^i}}$ with $\gamma^1 = \gamma^2 = 1/4$. Here we take $n_1 = n_2 = 750$, different values of the risk level p and 500 Monte-Carlo simulations.

4 Applications to environmental real data

This section is devoted to the analysis of two real data-sets.

4.1 Analysis of rainfall measurements

In this first real application, we consider the monthly mean of the rainfall measurements recorded in 3 different stations of the region *Bièvre*, located in the south of Paris (France), from 2003 to 2013 (see also Di Bernardino and Prieur (2014)). This data-set was provided by the SIAVB². The unit of measurements is *mm*. The 3-dimensional data-set is represented in Figure 9 (left). The temporal series of monthly mean data are denoted by X^1 at Station 1, X^2 at Station 2 and X^3 at Station 3. Recall that our estimation of the risk measure θ_p^i is based on d -dimensional *i.i.d.* observations. This assumption was validated for the present data-set in Di Bernardino and Prieur (2014). The length of the data-set is $n = 125$. We aim at estimating the multivariate CTE, i.e., $\theta_p^i = \mathbb{E}[X^i | Z > U_Z(1 - p)]$, for $i = 1, 2, 3$ and $Z = F(X^1, X^2, X^3)$. As a preliminary step, we have

²Syndicat Intercommunal pour l'Assainissement de la Vallée de la Bièvre, <http://www.siavb.fr/>

to choose the split-up parameters n_1 and n_2 . We propose a two-steps empirical strategy.

Step 1: for different values of n_1 ($n_1 \in \{25, 45, \dots, 100, 105\}$ in the present case), we get $\hat{\gamma}^i$ as in (8). The procedure for selecting $k_i(n_1)$ has been described in the beginning of Section 3. We then select the smallest value of n_1^0 ($n_1^0 = 57$ in the present case) for which the estimation of γ^i seems to stabilize (see Table 1).

Step 2: we fix the split-up parameters to n_1^0 and to $n_2^0 = n - n_1^0$ ($n_2^0 = 125 - 57 = 68$ in the present case). The idea is to reduce at most the bias due to the estimation of Z_j by $\tilde{Z}_j = F_{n_2}(\mathbf{X}_j)$, for $j = 1, \dots, n_1$ (see Item (a.3) of Assumption 2.1).

In the following, we thus fix $n_1 = 57$ and $n_2 = 68$. To check the asymptotic dependence assumption between X^i and \tilde{Z} (see Remark 4), we estimate the tail dependence coefficient by $\hat{R}_{(X^i, \tilde{Z})}(1, 1) = \frac{1}{k} \sum_{j=1}^{n_1} \mathbf{1}_{\{X_j^i > X_{n_1 - [k], n_1}^i, \tilde{Z}_j > \tilde{Z}_{n_1 - [k], n_1}\}}$, with $\tilde{Z}_j = F_{n_2}(\mathbf{X}_j)$, for $j = 1, \dots, n_1$ and $[\cdot]$ denotes the integer part (see Figure 9, right).

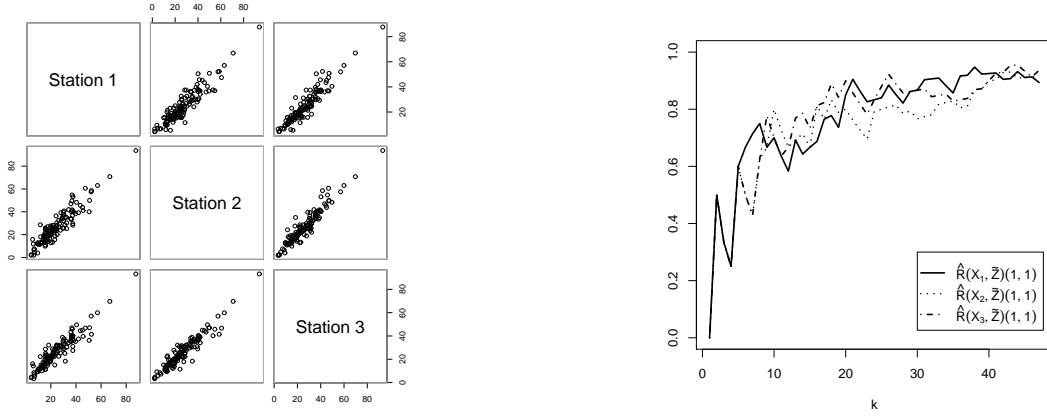


Figure 9: **Left:** Scatterplot of considered mean monthly rainfall data-set. **Right:** $\hat{R}_{(X^i, \tilde{Z})}(1, 1)$ in term of k .

The three estimators $\hat{R}_{(X^i, \tilde{Z})}(1, 1)$, for $i = 1, 2, 3$, are stable around the value 0.8. This strongly indicates that right-upper asymptotic dependence is present in this 3-dimensional real data-set.

For the assumption $\gamma^i < 1/2$, we plot the Hill estimation $\hat{\gamma}^i$ against various values of the intermediate sequence of integers k_i (see Figure 10). Accordingly to the procedure described at the beginning of Section 3, we select a stability window for k_i and we average the estimations $\hat{\gamma}^i$ on this window. On the present data-set, the window $k_i \in [20, 32]$ is selected for any $i = 1, 2, 3$.

We then estimate the Multivariate CTE, selecting a stability window for the intermediate sequence ($k \in [20, 35]$ on Figure 11). The results are gathered in Table 2 for each of the three stations and for $p(n_1) = 1/(2n_1)$, $1/(4n_1)$ and $1/(10n_1)$, corresponding to return periods of around 10, 20 and 50 years. On that data-set, the averaged monthly precipitations with return periods of around 10 years and 20 years are balanced between the three stations. However, for a return period of around 50 years, the contribution of the first station detaches from the other two contributions.

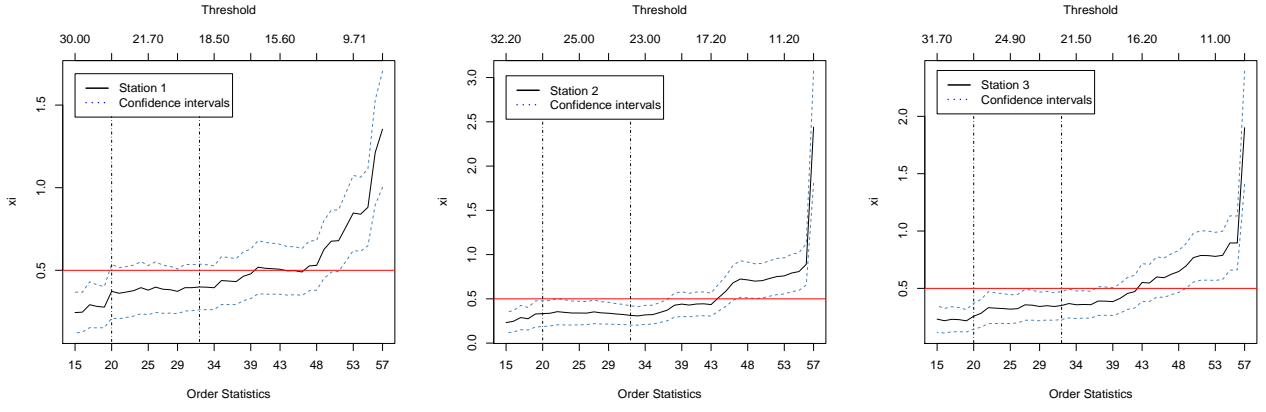


Figure 10: Mean monthly rainfall data-set. Hill estimates $\hat{\gamma}^i$ in Equation (8) for $i = 1, 2, 3$. Horizontal line represents the upper-bound of $1/2$. Asymptotic confidence intervals for level $\alpha = 0.95$ are also displayed.

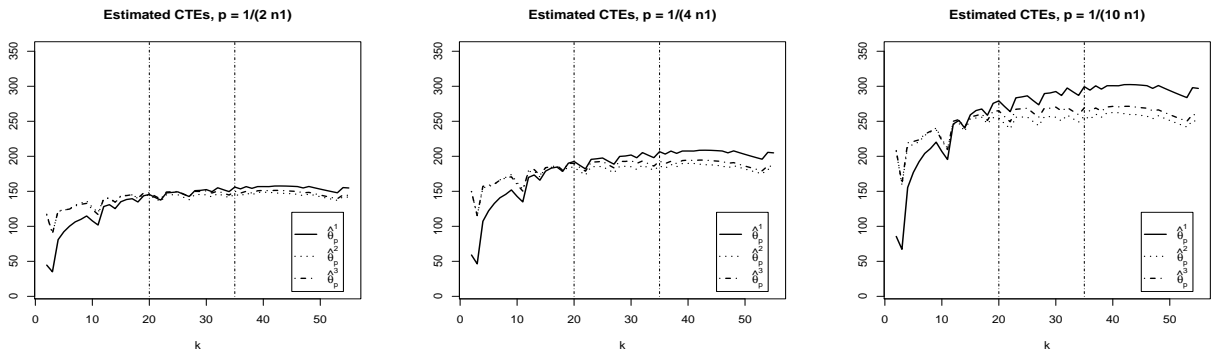


Figure 11: Mean monthly rainfall data-set. Estimated multivariate CTE against various values of the intermediate sequence $k(n_1)$, for $i = 1, 2, 3$, $n_1 = 57$ and different values of risk level p . Full line corresponds to Station 1, dotted to Station 2 and dashed-dotted to Station 3.

4.2 Analysis of high wind gusts

In this second application we focus on the study of strong wind-gusts. These data come from a larger data-set previously analyzed in Marcon et al. (2017). We consider the 2 weeks-max wind speed (WS) in meter per second (m/s), wind gust (WG) in m/s and positive increment air pressure (IP) at the sea level in millibar ($mbar$) recorded in Parcay-Meslay city in the north west of France, from July 2004 to July 2013 (see Figure 12, left).

Remark that this 3-dimensional data-set is composed by hydrological variables of different nature prohibiting the aggregation of the various components. In such a framework, our multivariate CTE measure defined in (1) is useful since it just considers as conditioning extreme event the behavior of the copula associated to the 3 different risk variables (see the Introduction section). The length of the data-set is $n = 232$. We select the two subsample sizes n_1 and n_2 with the two-steps empirical procedure proposed in Section 4.1. The results of **Step 1** are gathered in Table 3 and lead to the choice: $n_1 = 145$ and $n_2 = 87$.

On Figure 13, we have drawn the Hill estimates against different choices for the intermediate sequence of integers k_i . We see that the window $[15, 55]$ is a stability region for any $i = 1, 2, 3$. Then, to gain in stability, we average

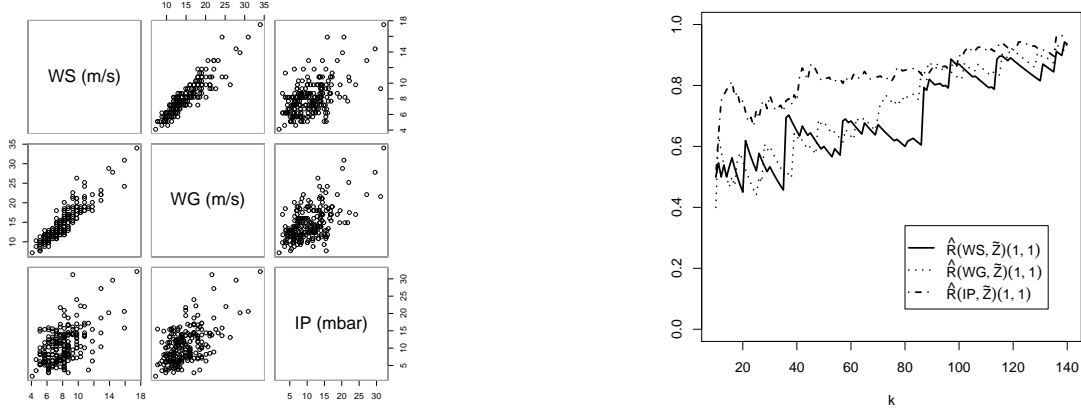


Figure 12: **Left:** Scatterplot of the 3–dimensional wind gusts data-set. **Right:** $\hat{R}_{(X^i, \tilde{Z})}(1, 1)$ against k .

the estimations $\hat{\gamma}^i$ corresponding to k_i in this stability region, as suggested in the beginning of Section 3. The results are reported in the first column of Table 4.

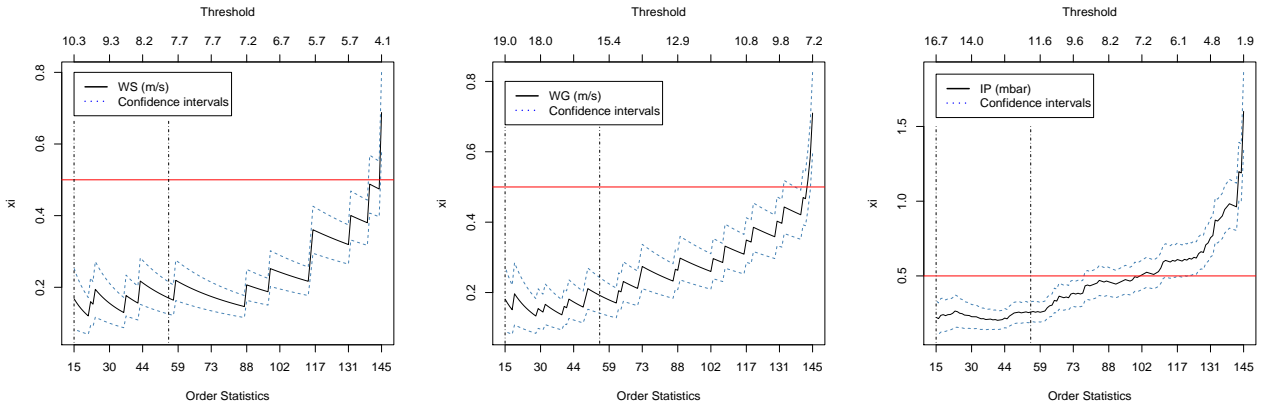


Figure 13: Wind gusts data-set. Hill estimates $\hat{\gamma}^i$ in Equation (8) against the intermediate sequence k_i , for $i = 1, 2, 3$. Horizontal line represents the upper-bound of $1/2$. Asymptotic confidence intervals for level $\alpha = 0.95$ are also displayed.

Then we analyse the asymptotic dependence of the couples (WS, \tilde{Z}) , (WG, \tilde{Z}) and (IP, \tilde{Z}) (see Remark 4), by using the estimated tail dependence coefficient $\hat{R}_{(X^i, \tilde{Z})}(1, 1)$. Figure 12 (right) shows that the estimations are stable around the value 0.8. This strongly indicates that right-upper asymptotic dependence is present in this wind gusts data-set. Using the values $\hat{\gamma}^i$ gathered in the first column of Table 4, we estimate the Multivariate CTE and plot the estimates against various values of the intermediate sequence k (see Figure 14). Following the same idea of balancing bias and variance, we choose $k \in [30, 65]$.

The final estimates based on averaging the estimates for $k \in [30, 65]$ are reported in Table 4 for different values of the risk level $p = p(n_1)$. We remark a lower contribution of WS to the multivariate stress scenario represented here by the event $\{Z > U_Z(1/p)\}$ for small values of risk level p . The contributions of WG and IP are similar, and about twice the one of WS for the three values of $p(n_1)$ (i.e., $p = 1/2 n_1$, $p = 1/4 n_1$ and $p = 1/10 n_1$).

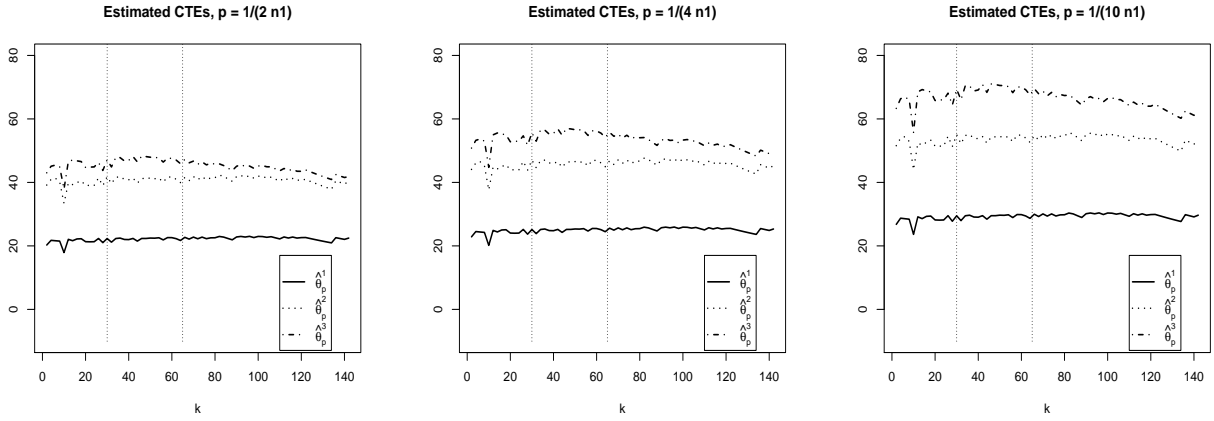


Figure 14: Wind gusts data-set. Estimated multivariate CTE against various values of the intermediate sequence k and for different values of risk level p . Full line corresponds to WS, dotted line to WG and dashed-dotted one to IP.

Acknowledgments: The authors acknowledge the anonymous reviewers and the associate editor for their numerous and very useful remarks and suggestions. The authors thank John H. J. Einmahl for valuable comments on that paper and Philippe Naveau and Simone Padoan for the high wind gusts data-set and the related discussion. This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d’avenir and by the project LEFE-MANU MULTIRISK. Within the CDP-Trajectories framework, this work is supported by the French National Research Agency in the framework of the ”Investissements d’avenir” program (ANR-15-IDEX-02).

A Proofs and auxiliary results

This appendix is devoted to the proofs of Theorem 2.1 and Proposition 2.1 stated in Section 2. We firstly focus on the central limit Theorem 2.1.

Proof of Theorem 2.1: Recall that $d_{n_1} := \frac{k}{n_1 p}$. Then, one can write

$$\frac{\widehat{\theta}_{p(n_1), n_2}^i}{\theta_{p(n_1)}^i} = \frac{\widehat{d}_{n_1}^i}{d_{n_1}^i} \times \frac{\widehat{\theta}_{\frac{k}{n_1}, n_2}^i}{\theta_{\frac{k}{n_1}}^i} \times \frac{d_{n_1}^i \theta_{\frac{k}{n_1}}^i}{\theta_{p(n_1)}^i} := L_1^{n_1} \times L_2^{n_1, n_2} \times L_3^{n_1}.$$

We now analyse these three factors separately. Under Assumptions (b) and (c), since $L_1^{n_1}$ does not depend on n_2 then, as in the proof of Theorem 4.3.8 in de Haan and Ferreira (2006), we get for $n_1 \rightarrow \infty$,

$$\frac{\sqrt{k_i}}{\ln(d_{n_1})} (L_1^{n_1} - 1) - \Gamma^i(q) \xrightarrow{\mathbb{P}} 0, \quad (11)$$

where $\Gamma^i(q)$ is defined in Equation (5).

The asymptotic behaviour of $L_2^{n_1, n_2}$, for $n_1, n_2 \rightarrow \infty$ is stated in Proposition 2.1 (see proof of Proposition 2.1 below). Finally for $L_3^{n_1}$, by using Equations (2)-(3) and their second order strengthening given by Assumptions

(b) and (d) respectively, we know from the proof of Theorem 1 in [Cai et al. \(2015\)](#) that

$$L_3^{n_1} = 1 + o\left(\frac{1}{\sqrt{k}}\right). \quad (12)$$

Now, combining (11)-(12) and Proposition 2.1 we obtain the convergence result of Theorem 2.1, where the covariance matrix of $(\Theta^i, \Gamma^i(q))$ is given in Section 1 and follows from straightforward computation. \square

The second part of this appendix is devoted to the proof of Proposition 2.1. This proof requires different preliminary results, introduced and proved below.

Proposition A.1. *Let $e_{n_1, n_2} := \frac{n_1}{k}(1 - K(\tilde{Z}_{n_1-k, n_1}))$, with $k = k(n_1)$, \tilde{Z}_{n_1-k, n_1} the $(n_1 - k)$ -th order statistic of $\tilde{Z}_1, \dots, \tilde{Z}_{n_1}$ and $\tilde{Z}_j = F_{n_2}(\mathbf{X}_j)$. Under Assumptions (a.2) and (a.3), $e_{n_1, n_2} \xrightarrow{\mathbb{P}} 1$, as $n_1, n_2 \rightarrow \infty$.*

Proof: Applying the d -dimensional extension of Kolmogorov-Smirnov Theorem and the properties of ordered statistics, we know that, for all $\epsilon > 0$

$$\left|Z_{n_1-k, n_1} - \tilde{Z}_{n_1-k, n_1}\right| = |(F(\mathbf{X}))_{n_1-k, n_1} - (F_{n_2}(\mathbf{X}))_{n_1-k, n_1}| = o(\sqrt{n_2}^{-1+\epsilon}), \quad (13)$$

where the ordered statistics are defined from the samples $(Z_j)_{j=1, \dots, n_1} = (F(\mathbf{X}_j))_{j=1, \dots, n_1}$ and $(\tilde{Z}_j)_{j=1, \dots, n_1} = (F_{n_2}(\mathbf{X}_j))_{j=1, \dots, n_1}$. Then we write

$$|e_{n_1, n_2} - 1| \leq \left|\frac{n_1}{k}(1 - K(Z_{n_1-k, n_1})) - 1\right| + \left|\frac{n_1}{k}(K(Z_{n_1-k, n_1}) - K(\tilde{Z}_{n_1-k, n_1}))\right|.$$

Since $1 - K(Z_{n_1-k, n_1})$ is the k -th order statistic of a random sample of size n_1 from the standard uniform distribution, we get $\left|\frac{n_1}{k}(1 - K(Z_{n_1-k, n_1})) - 1\right| \xrightarrow{\mathbb{P}} 0$.

We now study the second term. From Assumption (a.2), by applying a first-order Taylor approximation we get

$$\left|\frac{n_1}{k}(K(Z_{n_1-k, n_1}) - K(\tilde{Z}_{n_1-k, n_1}))\right| = \frac{n_1}{k} K'(Z_{n_1-k, n_1}) \left|(\tilde{Z}_{n_1-k, n_1} - Z_{n_1-k, n_1}) + o_{\mathbb{P}}(\tilde{Z}_{n_1-k, n_1} - Z_{n_1-k, n_1})\right|. \quad (14)$$

Since $K(Z_{n_1-k, n_1}) \sim 1 - \frac{k}{n_1}$ in probability for $n_1 \rightarrow \infty$, then, from Assumption (a.2), $K'(Z_{n_1-k, n_1}) = K'(K^{-1}(K(Z_{n_1-k, n_1})))$ is bounded for large values of n_1 . Then, by using Equations (13) and (14),

$$\left|\frac{n_1}{k}(K(Z_{n_1-k, n_1}) - K(\tilde{Z}_{n_1-k, n_1}))\right| = o\left(\frac{n_1}{k} n_2^{-\frac{1+\epsilon}{2}}\right),$$

which tends to zero as n_1 and n_2 tend to infinity from Assumption (a.3). Hence the result. \square

Lemma A.1 below is a variation of Lemma 1 in [Cai et al. \(2015\)](#) in our setting. The interested reader is also referred to Proposition 3.1 in [Einmahl et al. \(2006\)](#). The limit process is characterized by the aforementioned W_R -process (see Section 1). For convenient presentation, all the limit processes that are involved in Lemma A.1 are defined on the same probability space, via the Skorohod construction. However, they are only equal in distribution to the original processes.

Define, for $i \in I$,

$$R_{n_1}^i(x, z) := \frac{n_1}{k} \mathbb{P} \left[1 - F_i(X^i) < \frac{kx}{n_1}, 1 - K(Z) < \frac{kz}{n_1} \right]. \quad (15)$$

A non-parametric pseudo-estimator of $R_{n_1}^i$ (with unknown margins) is given by

$$T_{n_1, n_2}^i(x, z) := \frac{1}{k} \sum_{j=1}^{n_1} \mathbf{1}_{\{1-F_i(X_j^i) < \frac{kx}{n_1}, 1-K(\tilde{Z}_j) < \frac{kz}{n_1}\}}, \quad (16)$$

where $\tilde{Z}_j = F_{n_2}(\mathbf{X}_j)$. Its asymptotic behavior is stated in Lemma A.1 below.

Lemma A.1. *Let $i \in I$. Suppose that condition in (2) and Assumptions (a.2) and (a.3) hold true. Let $T > 0$ and $\eta \in (\max_{i \in I} \gamma^i, 1/2)$. Then, with probability 1, for $n_1, n_2 \rightarrow \infty$,*

$$\begin{aligned} \sup_{x, z \in (0, T]} \left| \frac{\sqrt{k} (T_{n_1, n_2}^i(x, z) - R_{n_1}^i(x, z)) - W_{R_{(X^i, Z)}}(x, z)}{x^\eta} \right| &\rightarrow 0, \\ \sup_{x \in (0, T]} \left| \frac{\sqrt{k} (T_{n_1, n_2}^i(x, \infty) - x) - W_{R_{(X^i, Z)}}(x, \infty)}{x^\eta} \right| &\rightarrow 0, \\ \sup_{z \in (0, T]} \left| \frac{\sqrt{k} (T_{n_1, n_2}^i(\infty, z) - z) - W_{R_{(X^i, Z)}}(\infty, z)}{z^\eta} \right| &\rightarrow 0, \end{aligned}$$

where $R_{n_1}^i(x, z)$ and $T_{n_1, n_2}^i(x, z)$ are defined by Equations (15) and (16) respectively.

Proof: Let us write

$$T_{n_1, n_2}^i(x, z) = T_{n_1}^i(x, z) + T_{n_1, n_2}^i(x, z) - T_{n_1}^i(x, z),$$

with $T_{n_1}^i(x, z) := \frac{1}{k} \sum_{j=1}^{n_1} \mathbf{1}_{\{1-F_i(X_j^i) < \frac{kx}{n_1}, 1-K(Z_j) < \frac{kz}{n_1}\}}$ and $Z_j = F(\mathbf{X}_j)$. Remark that, by using Lemma 1 in Cai et al. (2015), Lemma A.1 above holds true by replacing $T_{n_1, n_2}^i(x, z)$ by $T_{n_1}^i(x, z)$, $T_{n_1, n_2}^i(x, \infty)$ by $T_{n_1}^i(x, \infty)$ and $T_{n_1, n_2}^i(\infty, z)$ by $T_{n_1}^i(\infty, z)$. Let us thus study the term

$$\begin{aligned} D_{n_1, n_2}^i(x, z) &:= \sqrt{k} (T_{n_1, n_2}^i(x, z) - T_{n_1}^i(x, z)) \\ &= \frac{1}{\sqrt{k}} \sum_{j=1}^{n_1} \left(\mathbf{1}_{\{1-F_i(X_j^i) < \frac{kx}{n_1}, 1-K(\tilde{Z}_j) < \frac{kz}{n_1}\}} - \mathbf{1}_{\{1-F_i(X_j^i) < \frac{kx}{n_1}, 1-K(Z_j) < \frac{kz}{n_1}\}} \right) \\ &= \frac{1}{n_1} \sum_{j=1}^{n_1} L_{n_1, n_2, k, j}, \end{aligned}$$

with $L_{n_1, n_2, k, j} = \frac{n_1}{\sqrt{k}} \left(\mathbf{1}_{\{1-F_i(X_j^i) < \frac{kx}{n_1}, 1-K(\tilde{Z}_j) < \frac{kz}{n_1}\}} - \mathbf{1}_{\{1-F_i(X_j^i) < \frac{kx}{n_1}, 1-K(Z_j) < \frac{kz}{n_1}\}} \right)$.

One can deduce that, under Assumptions (a.2) and (a.3),

$$\frac{1}{n_1} \sum_{j=1}^{n_1} (L_{n_1, n_2, k, j} - \mathbb{E}(L_{n_1, n_2, k, j})) \xrightarrow[n_1, n_2]{\text{a.s.}} 0.$$

Thus, we now focus on $\frac{1}{n_1} \sum_{j=1}^{n_1} \mathbb{E}(L_{n_1, n_2, k, j})$. Using Hölder's Inequality, we get:

$$\begin{aligned} \left| \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbb{E}(L_{n_1, n_2, k, j}) \right| &\leq \frac{n_1}{\sqrt{k}} \left(\mathbb{P} \left[F_i(X_j^i) > 1 - \frac{kx}{n_1} \right] \right)^{1/p} \\ &\quad \cdot \left(\mathbb{E} \left[\left| \mathbf{1}_{1-K(\tilde{Z}_j) < \frac{kz}{n_1}, 1-K(Z_j) \geq \frac{kz}{n_1}} \right| + \left| \mathbf{1}_{1-K(\tilde{Z}_j) \geq \frac{kz}{n_1}, 1-K(Z_j) < \frac{kz}{n_1}} \right| \right]^q \right)^{1/q} \\ &= o \left(\frac{n_1}{\sqrt{k}} \left(\frac{kx}{n_1} \right)^{1/p} \left(n_2^{-\frac{-1+\epsilon}{2}} \right)^{1/q} \right) = o \left(\sqrt{n_1} \left(\frac{k}{n_1} \right)^{\frac{1}{p} - \frac{1}{2}} x^{\frac{1}{p}} n_2^{-\frac{-1+\epsilon}{2q}} \right). \end{aligned} \quad (17)$$

Then, from Assumption (a.3) it is possible to choose $p_0 < \frac{1}{\max_{i \in I} \gamma_i}$ and $\eta \in (\max_{i \in I} \gamma_i, 1/2)$ such that $\sup_{x, z \in (0, T]} \frac{|D_{n_1, n_2}^i(x, z)|}{x^\eta} = o\left(\sqrt{n_1} \left(\frac{k}{n_1}\right)^{\frac{1}{p_0} - \frac{1}{2}} n_2^{\frac{-1+\epsilon}{2q_0}} x^{1/p_0 - \eta}\right)$ tends to zero. \square

Finally, Proposition A.2 will be useful below to archive the proof of Proposition 2.1.

Proposition A.2. *Let $i \in I$. Define*

$$\bar{\theta}_{\frac{kz}{n_1}, n_2}^i := \frac{1}{kz} \sum_{j=1}^{n_1} X_j^i 1_{\{\tilde{Z}_j > U_Z(\frac{n_1}{kz})\}}. \quad (18)$$

Suppose that condition in (2) and Assumptions (a.2) and (a.3) hold with $\gamma^i \in (0, 1/2)$. Then, for $n_1, n_2 \rightarrow \infty$,

$$\sup_{1/2 \leq z \leq 2} \left| \left(\frac{\sqrt{k}}{U_i(\frac{n_1}{k})} \right) (\bar{\theta}_{\frac{kz}{n_1}, n_2}^i - \theta_{\frac{kz}{n_1}}^i) + \frac{1}{z} \int_0^\infty W_{R(x^i, z)}(s, z) ds^{-\gamma^i} \right| \xrightarrow{\mathbb{P}} 0,$$

with $\bar{\theta}_{\frac{kz}{n_1}, n_2}^i$ as in Equation (18) and $\theta_{\frac{kz}{n_1}}^i = \mathbb{E} \left[X^i \mid Z > Q_Z \left(1 - \frac{kz}{n_1} \right) \right]$.

Proof: Let $s_{n_1}(x) := \frac{n_1}{k} (1 - F_i(U_i(\frac{n_1}{k}) x^{-\gamma^i}))$, for $x > 0$. Remark that, from the regular variation condition in (3), $s_{n_1}(x) \rightarrow x$, as $n_1 \rightarrow \infty$. Furthermore Lemma 3 in Cai et al. (2015) states that, when handling proper integrals and by using the uniform convergence of $s_{n_1}(x)$ to x , as $n_1 \rightarrow \infty$, $s_{n_1}(x)$ can be substituted by x in the limit. We get

$$\begin{aligned} z \theta_{\frac{kz}{n_1}}^i &= \int_0^\infty \frac{n_1}{k} \mathbb{P} \left[X^i > s, Z > U_Z \left(\frac{n_1}{kz} \right) \right] ds = \int_0^\infty R_{n_1}^i \left(\frac{n_1}{k} (1 - F_i(s)), z \right) ds = \\ &= -U_i \left(\frac{n_1}{k} \right) \int_0^\infty R_{n_1}^i(s_{n_1}(x), z) dx^{-\gamma^i}, \quad (19) \end{aligned}$$

with $R_{n_1}^i$ as in (15). Similarly, $z \bar{\theta}_{\frac{kz}{n_1}, n_2}^i = -U_i(\frac{n_1}{k}) \int_0^\infty T_{n_1, n_2}^i(s_{n_1}(x), z) dx^{-\gamma^i}$, with T_{n_1, n_2}^i as in (16) and $\bar{\theta}_{\frac{kz}{n_1}, n_2}^i$ as in (18). For any $T > 0$

$$\begin{aligned} &\sup_{1/2 \leq z \leq 2} \left| \left(\frac{\sqrt{k}}{U_i(\frac{n_1}{k})} \right) \left(z \bar{\theta}_{\frac{kz}{n_1}, n_2}^i - z \theta_{\frac{kz}{n_1}}^i \right) + \int_0^\infty W_{R(x^i, z)}(x, z) dx^{-\gamma^i} \right| \\ &\leq \sup_{1/2 \leq z \leq 2} \left| \int_T^\infty W_{R(x^i, z)}(x, z) dx^{-\gamma^i} \right| + \sup_{1/2 \leq z \leq 2} \left| \int_T^\infty \sqrt{k} (T_{n_1, n_2}^i(s_{n_1}(x), z) - R_{n_1}^i(s_{n_1}(x), z)) dx^{-\gamma^i} \right| \\ &\quad + \sup_{1/2 \leq z \leq 2} \left| \int_0^T \sqrt{k} (T_{n_1, n_2}^i(s_{n_1}(x), z) - R_{n_1}^i(s_{n_1}(x), z)) - W_{R(x^i, z)}(x, z) dx^{-\gamma^i} \right| \\ &:= I_1(T) + I_2^{n_1, n_2}(T) + I_3^{n_1, n_2}(T). \end{aligned}$$

It is sufficient to prove that for any $\varepsilon > 0$, there exist $T_0 = T_0(\varepsilon)$ such that

$$\mathbb{P}[I_1(T_0) > \varepsilon] < \varepsilon \quad (20)$$

and $n_{1,0}, n_{2,0}$ such that for any $n_1 > n_{1,0}$ and $n_2 > n_{2,0}$,

$$\mathbb{P}[I_2^{n_1, n_2}(T_0) > \varepsilon] < \varepsilon, \quad (21)$$

$$\mathbb{P}[I_3^{n_1, n_2}(T_0) > \varepsilon] < \varepsilon. \quad (22)$$

Equation (20) holds true by application of Lemma 2 in Cai et al. (2015) with $\eta = 0$.

We now deal with (21). Once more we use the decomposition

$$\begin{aligned} T_{n_1, n_2}^i(s_{n_1}(x), z) &= T_{n_1}^i(s_{n_1}(x), z) + T_{n_1, n_2}^i(s_{n_1}(x), z) - T_{n_1}^i(s_{n_1}(x), z) \\ &= T_{n_1}^i(s_{n_1}(x), z) + \frac{1}{\sqrt{k}} D_{n_1, n_2}^i(s_{n_1}(x), z). \end{aligned}$$

Then, we deduce that $I_2^{n_1, n_2}(T_0)$ is bounded by

$$\sup_{1/2 \leq z \leq 2} \left| \int_{T_0}^{\infty} D_{n_1, n_2}^i(s_{n_1}(x), z) dx^{-\gamma^i} \right| + \sup_{1/2 \leq z \leq 2} \left| \int_{T_0}^{\infty} \sqrt{k} (T_{n_1}^i(s_{n_1}(x), z) - R_{n_1}^i(s_{n_1}(x), z)) dx^{-\gamma^i} \right|.$$

Then using the bound in (17) for $p = +\infty$ and $q = 1$, we get

$$\sup_{1/2 \leq z \leq 2} \left| \int_{T_0}^{\infty} D_{n_1, n_2}^i(s_{n_1}(x), z) dx^{-\gamma^i} \right| = O\left(\frac{1}{T_0^{\gamma^i}} \frac{n_1}{\sqrt{k}} n_2^{\frac{-1+\epsilon}{2}}\right),$$

which tends to zero from Assumption (a.3). We conclude for the term $I_2^{n_1, n_2}(T_0)$ by using the result of Proposition 2 in Cai et al. (2015) for $\sup_{1/2 \leq z \leq 2} \left| \int_{T_0}^{\infty} \sqrt{k} (T_{n_1}^i(s_{n_1}(x), z) - R_{n_1}^i(s_{n_1}(x), z)) dx^{-\gamma^i} \right|$.

It remains to handle the term (22). We get

$$\begin{aligned} \mathbb{P}[I_3^{n_1, n_2}(T) > \varepsilon] &= \mathbb{P}\left[\sup_{1/2 \leq z \leq 2} \left| \int_0^T \sqrt{k} (T_{n_1, n_2}^i(s_{n_1}(x), z) - R_{n_1}^i(s_{n_1}(x), z)) - W_{R_{(X^i, Z)}}(x, z) dx^{-\gamma^i} \right| > \varepsilon\right] \\ &\leq \mathbb{P}\left[\sup_{1/2 \leq z \leq 2} \left| \int_0^T \sqrt{k} (T_{n_1, n_2}^i(s_{n_1}(x), z) - R_{n_1}^i(s_{n_1}(x), z)) - W_{R_{(X^i, Z)}}(s_{n_1}(x), z) dx^{-\gamma^i} \right| > \varepsilon/2\right] \\ &\quad + \mathbb{P}\left[\sup_{1/2 \leq z \leq 2} \left| \int_0^T W_{R_{(X^i, Z)}}(s_{n_1}(x), z) - W_{R_{(X^i, Z)}}(x, z) dx^{-\gamma^i} \right| > \varepsilon/2\right] = p_{31}^{n_1, n_2} + p_{32}^{n_1}. \end{aligned}$$

Firstly we consider $p_{31}^{n_1, n_2}$. Notice that for any T , there exists $\tilde{n}_1 = n_1(T)$, such that for all $n_1 > \tilde{n}_1$, $s_{n_1}(T) < T + 1$. Hence for $n_1 > \tilde{n}_1$ and for $\eta_0 \in (\max_{i \in I} \gamma^i, 1/2)$,

$$p_{31}^{n_1, n_2} \leq \mathbb{P}\left[\sup_{\substack{1/2 \leq z \leq 2, \\ 0 < s \leq T+1}} \left| \frac{\sqrt{k} (T_{n_1, n_2}^i(s, z) - R_{n_1}^i(s, z)) - W_{R_{(X^i, Z)}}(s, z)}{s^{\eta_0}} \right| \left| \int_0^T (s_{n_1}(x))^{\eta_0} dx^{-\gamma^i} \right| > \varepsilon/2\right].$$

Notice that, by Lemma 3 in Cai et al. (2015), $\left| \int_0^T (s_{n_1}(x))^{\eta_0} dx^{-\gamma^i} \right| \rightarrow \frac{\gamma^i}{\eta_0 - \gamma^i} T^{\eta_0 - \gamma^i}$, as $n_1 \rightarrow \infty$. By application of Lemma A.1, we conclude the proof for $p_{31}^{n_1, n_2}$. Finally, since $p_{32}^{n_1}$ does not depend on n_2 we can conclude using Lemma 2 in Cai et al. (2015). \square

We now use the auxiliary results above to prove Proposition 2.1 stated in Section 2.

Proof of Proposition 2.1: From Proposition 1 in Cai et al. (2015) applied to the bivariate vector (X^i, Z) , we have that $\lim_{n_1 \rightarrow \infty} \frac{\theta_{\frac{k}{n_1}}^i}{U_i(n_1/k)} = \int_0^{\infty} R_{(X^i, Z)}(x^{-1/\gamma^i}, 1) dx$. Then, to prove Proposition 2.1 it is sufficient to prove that, for $n_1, n_2 \rightarrow \infty$

$$\frac{\sqrt{k}}{U_i(n_1/k)} \left(\hat{\theta}_{\frac{k}{n_1}, n_2}^i - \theta_{\frac{k}{n_1}}^i \right) \xrightarrow{\mathbb{P}} \Theta^i \int_0^{\infty} R_{(X^i, Z)}(x^{-1/\gamma^i}, 1) dx.$$

Note that $\hat{\theta}_{\frac{k}{n_1}, n_2}^i = e_{n_1, n_2} \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1}$. From Proposition A.1, we know that $e_{n_1, n_2} \xrightarrow{\mathbb{P}} 1$, as $n_1, n_2 \rightarrow \infty$. Using the analytical expression of process Θ^i in Equation (4), we can write:

$$\begin{aligned} & \frac{\sqrt{k}}{U_i(n_1/k)} \left(e_{n_1, n_2} \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1} - \theta_{\frac{k}{n_1}}^i \right) - \Theta^i \int_0^\infty R_{(X^i, Z)}(x^{-1/\gamma^i}, 1) dx \\ &= \left(\frac{\sqrt{k}}{U_i(n_1/k)} \left(e_{n_1, n_2} \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1} - e_{n_1, n_2} \theta_{\frac{k}{n_1}}^i \right) + \int_0^\infty W_{R_{(X^i, Z)}}(s, 1) ds^{-\gamma^i} \right) \\ &+ \left(\frac{\sqrt{k}}{U_i(n_1/k)} \left(e_{n_1, n_2} \theta_{\frac{k}{n_1}}^i - \theta_{\frac{k}{n_1}}^i \right) - W_{R_{(X^i, Z)}}(\infty, 1)(\gamma^i - 1) \int_0^\infty R_{(X^i, Z)}(s^{-1/\gamma^i}, 1) ds \right) \\ &=: J_1^{(n_1, n_2)} + J_2^{(n_1, n_2)}. \end{aligned}$$

We prove that both $J_1^{(n_1, n_2)}$ and $J_2^{(n_1, n_2)}$ converge to zero in probability as $n_1, n_2 \rightarrow \infty$. Using Lemma A.1, since $T_{n_1, n_2}^i(\infty, e_{n_1, n_2}) = 1$, we get, as $n_1, n_2 \rightarrow \infty$:

$$\sqrt{k}(e_{n_1, n_2} - 1) \xrightarrow{\mathbb{P}} -W_{R_{(X^i, Z)}}(\infty, 1). \quad (23)$$

This implies that $\lim_{n_1, n_2 \rightarrow +\infty} \mathbb{P}(|e_{n_1, n_2} - 1| > k^{-1/4}) = 0$. Thus

$$\begin{aligned} |J_1^{(n_1, n_2)}| &\leq \sup_{|z-1| < k^{-1/4}} \left| \frac{\sqrt{k}}{U_i(n_1/k)} \left(z \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1} - z \theta_{\frac{k}{n_1}}^i \right) + \int_0^\infty W_{R_{(X^i, Z)}}(s, z) ds^{-\gamma^i} \right| \\ &+ \sup_{|z-1| < k^{-1/4}} \left| \int_0^\infty W_{R_{(X^i, Z)}}(s, z) - W_{R_{(X^i, Z)}}(s, 1) ds^{-\gamma^i} \right|. \end{aligned}$$

The first term of the right hand term above converges to zero in probability by Proposition A.2. The second term can be handled as in the proof of Proposition 3 in Cai et al. (2015), using Lemma 2 in Cai et al. (2015).

We now focus on $J_2^{(n_1, n_2)}$. Firstly recall that, using Assumption (a.1), as $n_1 \rightarrow \infty$

$$\sup_{1/2 \leq z \leq 2} \sqrt{k} \left| \int_0^\infty R_{n_1}^i(s_{n_1}(x), z) - R_{(X^i, Z)}(x, z) dx^{-\gamma^i} \right| \rightarrow 0, \quad (24)$$

(see Equation (27) in Cai et al. (2015)). Combining (19) and (24), we get:

$$\frac{e_{n_1, n_2} \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1}}{U_i(n_1/k)} = - \int_0^\infty R_{n_1}^i(s_{n_1}(x), e_{n_1, n_2}) dx^{-\gamma^i} = - \int_0^\infty R_{(X^i, Z)}(x, e_{n_1, n_2}) dx^{-\gamma^i} + o_{\mathbb{P}}(1/\sqrt{k}),$$

where the last term $o_{\mathbb{P}}(1/\sqrt{k})$ does not depend on n_2 . Using the homogeneity of R function, we have:

$$e_{n_1, n_2} \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1} = e_{n_1, n_2}^{1-\gamma^i} \theta_{\frac{k}{n_1}}^i + o_{\mathbb{P}}\left(\frac{U_i(n_1/k)}{\sqrt{k}}\right),$$

still with the last term not depending on n_2 . By applying (23) and Proposition 1 in Cai et al. (2015) for the bivariate vector (X^i, Z) , as far as the Cramér's delta method, we get as $n_1, n_2 \rightarrow \infty$,

$$\begin{aligned} \frac{\sqrt{k}}{U_i(n_1/k)} \left(e_{n_1, n_2} \frac{\bar{\theta}_{\frac{k}{n_1}, n_2}^i}{n_1} - \theta_{\frac{k}{n_1}}^i \right) &= \sqrt{k} \left(e_{n_1, n_2}^{1-\gamma^i} - 1 \right) \frac{\theta_{\frac{k}{n_1}}^i}{U_i(n_1/k)} + o_{\mathbb{P}}(1) \\ &\xrightarrow{\mathbb{P}} (\gamma^i - 1) W_{R_{(X^i, Z)}}(\infty, 1) \int_0^\infty R_{(X^i, Z)}(s^{-1/\gamma^i}, 1) ds, \end{aligned}$$

uniformly in n_2 . Hence $J_2^{(n_1, n_2)}$ converges to zero in probability as $n_1, n_2 \rightarrow \infty$. Hence the result. \square

References

- Barbe, P., Genest, C., Ghoudi, K., and Rémillard, B. (1996). On Kendall's process. *J. Multivariate Anal.*, 58(2):197–229.
- Bargès, M., Cossette, H., and Marceau, É. (2009). TVaR-based capital allocation with copulas. *Insurance Math. Econom.*, 45(3):348–361.
- Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.
- Belzunce, F., Castaño, A., Olvera-Cervantes, A., and Suárez-Llorens, A. (2007). Quantile curves and dependence structure for bivariate distributions. *Comput. Statist. Data Anal.*, 51(10):5112–5129.
- Cai, J. and Li, H. (2005). Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, 42(3):810–825.
- Cai, J. J., Einmahl, J. H. J., de Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journals of the Royal Statistical Society, Series B*, 77(2):417–442.
- Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A stochastic ordering based on a decomposition of Kendall's tau. In *Distributions with given marginals and moment problems (Prague, 1996)*, pages 81–86. Kluwer Acad. Publ., Dordrecht.
- Chebana, F. and Ouarda, T. (2011). Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22(1):63–78.
- Cousin, A. and Di Bernardino, E. (2014). On multivariate extensions of Conditional-Tail-Expectation. *Insurance: Mathematics and Economics*, 55(0):272 – 282.
- de Haan, L. and Ferreira, A. (2006). *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York. An introduction.
- de Haan, L. and Huang, X. (1995). Large quantile estimation in a multivariate setting. *J. Multivariate Anal.*, 53(2):247–263.
- Di Bernardino, E., Laloë, T., Maume-Deschamps, V., and Prieur, C. (2013). Plug-in estimation of level sets in a non-compact setting with applications in multivariable risk theory. *ESAIM: Probability and Statistics*, (17):236–256.
- Di Bernardino, E. and Palacios-Rodríguez, F. (2016). Estimation of extreme quantiles conditioning on multivariate critical layers. *Environmetrics*, 27(3):158–168.
- Di Bernardino, E. and Palacios-Rodríguez, F. (2017). Estimation of extreme component-wise excess design realization: a hydrological application. *Stochastic Environmental Research and Risk Assessment*.
- Di Bernardino, E. and Prieur, C. (2014). Estimation of multivariate Conditional-Tail-Expectation using Kendall's process. *Journal of Nonparametric Statistics*, 26(2):241–267.

- Draisma, G., Drees, H., Ferreira, A., and De Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251–280.
- Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.*, 64(1):25–47.
- Einmahl, J. H. J., de Haan, L., and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.*, 34(4):1987–2014.
- Genest, C., Quesy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33:337–366.
- Genest, C. and Rivest, L.-P. (2001). On the multivariate probability integral transformation. *Statist. Probab. Lett.*, 53(4):391–399.
- Gorge, G. (2013). *Insurance Risk Management and Reinsurance*. Gorge.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174.
- Hua, L. and Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks. *Insurance Math. Econom.*, 49(3):537–546.
- Landsman, Z. M. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *N. Am. Actuar. J.*, 7(4):55–71.
- Mao, T. and Hu, T. (2012). Second-order properties of the Haezendonck-Goovaerts risk measure for extreme risks. *Insurance: Mathematics and Economics*, 51(2):333 – 343.
- Marcon, G., Naveau, P., and Padoan, S. (2017). A semi-parametric stochastic generator for bivariate extreme events. *STAT*, (in press).
- Nelsen, R. B. (1999). *An introduction to copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- Nelsen, R. B., Quesada-Molinab, J. J., Rodríguez-Lallenac, J. A., and Úbeda-Floresc, M. (2003). Kendall distribution functions. *Statistics and Probability Letters*, 65:263–268.
- Salvadori, G., De Michele, C., and Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15:3293–3305.
- Salvadori, G., Durante, F., and Perrone, E. (2013). Semi-parametric approximation of Kendall’s distribution function and multivariate Return Periods. *Journal de la Société Française de Statistique*, 154(1):151–173.
- Salvadori, G., Tomasicchio, G. R., and D’Alessandro, F. (2014). Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. *Coastal Engineering*, 88:1–14.
- Singh, V., Jain, S., and Tyagi, A. (2007). *Risk and Reliability Analysis*. ASCE Press, Reston, Virginia.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

- Torres, R., De Michele, C., Laniado, H., and Lillo, R. E. (2017). Directional multivariate extremes in environmental phenomena. *Environmetrics*, 28(2):e2428–n/a.
- van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, pages 234–252.
- Vandenbergh, S., van den Berg, M. J., Gräler, B., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C. (2012). Joint return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 9:6781–6828.
- Weissman, I. (1978). Estimation of parameters and larger quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.

| | | | | | | | | | | | | | |
|---------------------------|----------|----------|----------|----------|----------|--------------|--------------|--------------|----------|----------|----------|----------|--------|
| n_1 | 105 | 100 | 95 | 85 | 75 | 65 | 60 | 57 | 56 | 55 | 50 | 45 | 25 |
| n_2 | 20 | 25 | 30 | 40 | 50 | 60 | 65 | 68 | 69 | 70 | 75 | 80 | 100 |
| stable range k_i | [45, 60] | [40, 60] | [35, 60] | [33, 50] | [24, 40] | [22, 37] | [22, 35] | [20, 32] | [17, 30] | [17, 30] | [15, 27] | [14, 25] | [3, 7] |
| averaged $\hat{\gamma}_1$ | 0.404 | 0.399 | 0.399 | 0.403 | 0.413 | 0.407 | 0.405 | 0.405 | 0.362 | 0.344 | 0.353 | 0.365 | 0.209 |
| averaged $\hat{\gamma}_2$ | 0.355 | 0.357 | 0.381 | 0.367 | 0.363 | 0.355 | 0.352 | 0.353 | 0.315 | 0.298 | 0.292 | 0.305 | 0.213 |
| averaged $\hat{\gamma}_3$ | 0.367 | 0.363 | 0.367 | 0.366 | 0.362 | 0.368 | 0.363 | 0.363 | 0.303 | 0.297 | 0.288 | 0.302 | 0.226 |

Table 1: Mean monthly rainfall data-set. Estimation of γ_i for different values of n_1 and $n_2 = n - n_1$. Here $n = 125$.

| Station i | $\hat{\gamma}^i$ | $\hat{\theta}_{p=1/(2n_1),n_2}^i$ | $\hat{\theta}_{p=1/(4n_1),n_2}^i$ | $\hat{\theta}_{p=1/(10n_1),n_2}^i$ |
|-------------|------------------|-----------------------------------|-----------------------------------|------------------------------------|
| 1 | 0.405 | 148.089 | 196.083 | 284.188 |
| 2 | 0.353 | 142.781 | 182.363 | 252.005 |
| 3 | 0.363 | 147.084 | 189.164 | 263.811 |

Table 2: The estimates $\hat{\gamma}^i$ are computed by taking the average for $k_i \in [20, 32]$ for $i = 1, 2, 3$. The estimates of the multivariate CTE are based on these values of $\hat{\gamma}^i$. We report the average of $\hat{\theta}_{p(n_1),n_2}^i$ for $k(n_1) \in [20, 35]$ with $n_1 = 57$ and $p(n_1) = 1/(2n_1), 1/(4n_1), 1/(10n_1)$.

| | | | | | | | | | | |
|---------------------------|----------|----------|----------|----------|----------|--------------|--------------|--------------|----------|----------|
| n_1 | 210 | 220 | 205 | 190 | 175 | 160 | 150 | 145 | 140 | 135 |
| n_2 | 22 | 12 | 27 | 42 | 57 | 72 | 82 | 87 | 92 | 97 |
| stable range k_i | [30, 90] | [30, 90] | [20, 70] | [20, 60] | [20, 61] | [16, 60] | [15, 60] | [15, 55] | [15, 55] | [15, 45] |
| averaged $\hat{\gamma}_1$ | 0.172 | 0.169 | 0.169 | 0.179 | 0.172 | 0.173 | 0.173 | 0.173 | 0.165 | 0.164 |
| averaged $\hat{\gamma}_2$ | 0.181 | 0.179 | 0.178 | 0.178 | 0.174 | 0.171 | 0.171 | 0.171 | 0.161 | 0.154 |
| averaged $\hat{\gamma}_3$ | 0.235 | 0.231 | 0.236 | 0.243 | 0.245 | 0.241 | 0.240 | 0.240 | 0.237 | 0.227 |

Table 3: Wind gusts data-set. Estimation of γ_i for different values of n_1 and $n_2 = n - n_1$. Here $n = 232$.

| Variables | $\hat{\gamma}^i$ | $\hat{\theta}_{p=1/(2n_1),n_2}^i$ | $\hat{\theta}_{p=1/(4n_1),n_2}^i$ | $\hat{\theta}_{p=1/(10n_1),n_2}^i$ |
|-----------|------------------|-----------------------------------|-----------------------------------|------------------------------------|
| WS | 0.173 | 22.177 | 25.002 | 29.298 |
| WG | 0.171 | 40.981 | 46.137 | 53.964 |
| IP | 0.240 | 47.253 | 55.805 | 69.532 |

Table 4: The estimates $\hat{\gamma}^i$ are computed by taking the average for $k_i \in [15, 55]$ for $i = 1, 2, 3$. The estimates of the multivariate CTE are based on these values of $\hat{\gamma}^i$. We report the average of $\hat{\theta}_{p(n_1),n_2}^i$ for $k(n_1) \in [30, 65]$ with $n_1 = 145$ and $p(n_1) = 1/(2n_1), 1/(4n_1), 1/(10n_1)$.