Estimating covariate functions associated to multivariate risks: a level sets approach

Elena Di Bernardino, Thomas Laloë, Rémi Servien

Noname manuscript No.
(will be inserted by the editor)

Estimating covariate functions associated to multivariate risks: a level set approach

Elena Di Bernardino · Thomas Laloë · Rémi Servien

the date of receipt and acceptance should be inserted later

Abstract The aim of this paper is to study the behavior of a covariate function in a multivariate risks scenario. The first part of this paper deals with the problem of estimating the c-upper level sets $L(c) = \{F(x) \geq c\}$, with $c \in (0, 1)$, of an unknown distribution function F on $\mathbb{R}^d_+$. A plug-in approach is followed. We state consistency results with respect to the volume of the symmetric difference. In the second part, we obtain the L_p-consistency, with a convergence rate, for the regression function estimate on these level sets $L(c)$. We also consider a new multivariate risk measure: the Covariate-Conditional-Tail-Expectation. We provide a consistent estimator for this measure with a convergence rate. We propose a consistent estimate when the regression cannot be estimated on the whole data set. Then, we investigate the effects of scaling data on our consistency results. All these results are proven in a non-compact setting. A complete simulation study is detailed and a comparison with parametric and semi-parametric approaches is provided. Finally, a real environmental application of our risk measure is provided.

Keywords Multidimensional distribution function, plug-in estimation, regression function.

Mathematics Subject Classification (2000) 62G05, 62G20, 60E05, 91B30

E. Di Bernardino
CNAM, 292 rue Saint-Martin, 75141 Paris Cedex 03, France.

T. Laloë
Université de Nice Sophia-Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.
Tel. : +33492076020.
E-mail: laloe@unice.fr

R. Servien
INRA, Universite de Toulouse, UMR1331 Toxalim, Research Centre in Food Toxicology, F-31027 Toulouse.

1 Introduction

Traditionally, risk measures are thought of as mappings from a set of real-valued random variables to the real numbers. However, it is often insufficient to consider a single real measure to quantify risks, especially when the risk-problem is affected by other external risk factors. Note that the evaluation of an individual risk may strongly be affected by the degree of dependence amongst all risks. Modeling the dependency structure of multivariate data is helpful to obtain meaningful and accurate inference and prediction results in risk analysis. Several important challenges exist in this sense: the ability to incorporate important known covariates in the model; the dependence between variables is neither linear nor constant; they exhibit tail dependence, and so on. For instance, several hydrological phenomena are described by two or more correlated characteristics. These dependent characteristics should be considered jointly to be more representative of the multivariate nature of the phenomenon. Consequently, probabilities of occurrence of risks cannot be estimated on the basis of univariate analysis. The multivariate hydrological risks literature mainly treated one or more of the following three elements: (1) showing the importance and explaining the usefulness of the multivariate framework, (2) fitting the appropriate multivariate distribution in order to model risks and (3) defining and studying multivariate return periods (see Chebana and Ouarda (2011)). One of the most popular measures in hydrology and climate is undoubtedly the return period. It is closely related to the notion of quantile which has therefore been extensively studied in dimension one. For a random variable X that represents the magnitude of an event that occurs at a given time and at a given location, the quantile of order $1-\frac{1}{T}$ expresses the magnitude of the event which is exceeded with a probability equal to $\frac{1}{T}$. T is then called the return period. In univariate risk theory the quantile is known as the *Value-at-Risk* (VaR) and is defined by

$$Q_X(c) = \inf\{x \in \mathbb{R} : F_X(x) \geq c\}, \quad \text{for } c \in (0,1),$$

with F_X the univariate distribution of random variable X. A second important univariate risk measure, based on the quantile notion, is the *Conditional-Tail-Expectation* (CTE) defined by

$$\mathrm{CTE}_c(X) = \mathbb{E}[\, X \mid X > Q_X(c)\,], \quad \text{for } c \in (0,1).$$

From the year 2000 onward, much research has been devoted to risk measures, and many extensions to multidimensional settings have been suggested (see, e.g., Jouini et al (2004); Bentahar (2006); Embrechts and Puccetti (2006); Nappo and Spizzichino (2009); Ekeland et al (2012)). As a starting point, in the following, we consider the multivariate version of the CTE measure, proposed by Di Bernardino et al (2013) and Cousin and Di Bernardino (2013). It is constructed as the conditional expectation of a multivariate random vector given that the latter is located in the c-upper level set of the associated multivariate distribution function. In this sense this measure is essentially based on a "multivariate distributional approach". More precisely they define, for $i = 1, \dots, d$ and for $c \in (0, 1)$,

$$\mathrm{CTE}^i_c(\mathbf{X}) = \mathbb{E}[\, X_i \,|\, \mathbf{X} \in L(c)\,], \tag{1.1}$$

where $\mathbf{X} = (X_1, \dots, X_d)$ is a non-negative multivariate risk portfolio with distribution function F. In particular, Cousin and Di Bernardino (2013) proved that properties of the multivariate Conditional-Tail-Expectation in (1.1) turn to be consistent with existing properties on univariate risk measures (positive homogeneity, translation invariance, increasing in risk-level c, ...). In the financial econometrics literature, we are often interested in analyzing the behavior of a univariate return measure (average return, skewness, ...) with respect to a set of d risk factors $\mathbf{X}$ (volatility or variance, kurtosis, ...). In other words, we consider a dependent multivariate vector of risk-factors $\mathbf{X}$ and a univariate covariate Y (i.e. a dependent variable on $\mathbf{X}$). Furthermore, in climatology, one may be interested in how climate change over years might affect high temperatures. Multivariate examples include the study of rainfall as a covariate function represented by the geographical location. In this sense Daouia et al (2010) deal with the problem of estimating quantiles when covariate information is available.

So, the goal of this paper is the study of the behavior of a covariate Y on the level sets of a d-dimensional vector of risk-factors $\mathbf{X}$. More precisely, adapting the multivariate risk measure in (1.1), we deal with the multivariate Covariate-Conditional-Tail-Expectation (CCTE) defined by:

$$\mathrm{CCTE}_c(\mathbf{X}, Y) := \mathbb{E}[Y \,|\, \mathbf{X} \in L(c)], \tag{1.2}$$

where $c \in (0, 1)$. In order to estimate this risk measure, we first need to estimate the level sets $L(c)$ associated to the d-dimensional distribution function F of $\mathbf{X}$. For a non-negative d-dimensional risk portfolio with distribution function F, the c-upper level set of F (i.e., $L(c) = \{x \in \mathbb{R}^d_+ : F(x) \geq c\}$) and its associated c-level curve (i.e., $\partial L(c) = \{x \in \mathbb{R}^d_+ : F(x) = c\}$) have recently been proposed as risk measures in multivariate hydrological models. Among their many advantages, it appears that they are simple, intuitive, interpretable and probability-based (see Chebana and Ouarda (2011)). A risk-problem of flood in the bivariate setting using an estimator of level curves $\partial L(c)$ of the bivariate distribution function is proposed by de Haan and Huang (1995). Furthermore, as noticed by Embrechts and Puccetti (2006), $\partial L(c)$ can be viewed as a natural multivariate version of the univariate quantile. The interested reader is also referred to Tibiletti (1993), Belzunce et al (2007), Nappo and Spizzichino (2009).

The problem of estimating level sets of an unknown function (for instance of a density function and more recently a regression function) has received attention. However, most of the existing literature has focused on the density or regression function (Baíllo et al (2001); Cavalier (1997); Cuevas et al (2006);

Laloë and Servien (2013); Rigollet and Vert (2009)). Mason and Polonik (2009) obtained the asymptotic normality of plug-in level set estimates in the density case. As we consider the level sets of a multivariate distribution function, i.e. $L(c)$ for some fixed $c \in (0, 1)$, the commonly assumed property of compactness for these sets (required both in the density and in the regression cases) is no more reasonable. Then, differently from the literature cited above, in the present work, a special attention is given to this non-compact setting.

Considering a consistent estimator F_n of the distribution function F, we propose a *plug-in* approach to estimate the c-upper level set $L(c)$ by

$$L_n(c) = \{x \in \mathbb{R}^d_+ : F_n(x) \geq c\}, \text{ for } c \in (0, 1).$$

The regularity properties of F and F_n as well as the consistency properties of F_n will be specified in the statements of our theorems. Our consistency result for $L(c)$ is stated with respect to a criterion of "physical proximity" between sets: the volume of the symmetric difference. Obviously, the convergence rate suffers from the well-known *curse of dimensionality* (see Theorem 1). Using Theorem 1, we state L_p-consistency with a convergence rate for the estimation of the regression function

$$r(x) = \mathbb{E}[Y \mid \mathbf{X} = x],$$

on these level sets $L(c)$, i.e., for $x \in L(c)$, and $c \in (0, 1)$ (see Theorems 2 and 3). The motivation behind the point-wise estimation of $r(x)$ for $x \in L(c)$ is an interesting problem, for different practical problems. Indeed, this represents the expected value of a covariate Y given that a dependent multivariate vector of risk-factors $\mathbf{X}$ takes value in a specific risk area $L(c)$ (for instance $L(c)$ can represent a risk-scenario or critical-layer and so on). For the importance of $L(c)$ in the risk-management in the environmental or hydrological fields see for Chebana and Ouarda (2011), Salvadori et al (2011). Finally, we provide a consistency result for the estimation of the CCTE risk measure, if Y is completely available (see Theorem 4) or not (see Theorem 5). Furthermore, we investigate the impact of a change in the scale of data on our results. In particular this property is related to the suitable positive homogeneity property of risk measures (e.g. see Artzner et al (1999)).

The paper is organized as follows. We introduce some notation, tools and technical assumptions in Section 2. Consistency and asymptotic properties of our estimator of $L(c)$ are given in Section 3. Section 4 is devoted to the L_p-consistency of the estimation of the regression function r on the level set $L(c)$ and Section 5 to the consistency of the CCTE's estimation. The effects of scaling data are analyzed in Section 6. Illustrations with simulated data are presented in Section 7, and a comparison with parametric and semi-parametric approaches is detailed. A real example is studied in Section 8. Section 9 summarizes and briefly mentions directions for future research. Finally, proofs are postponed to Section 10.

2 Notation and preliminaries

In this section we introduce some notations and tools which will be useful later.

Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$ and ${\mathbb{R}_+^d}^* = \mathbb{R}_+^d \setminus \{0\}$. Let $\mathcal{F}$ be the set of continuous distribution functions $\mathbb{R}_+^d \to [0,1]$ and $\mathbf{X} := (X_1, X_2, \ldots, X_d)$ a random vector with distribution function $F \in \mathcal{F}$. Given an i.i.d. sample $\{\mathbf{X}_i\}_{i=1}^n$ in $\mathbb{R}_+^d$ with distribution function $F \in \mathcal{F}$, we denote by F_n an estimator of F based on this finite sample. Let Y be a random variable with values in $J \subset \mathbb{R}_+$, where J is supposed to be bounded. We denote by $\{(\mathbf{X}_i, Y_i)\}_{i=1,\ldots,n}$ the associated i.i.d. sample.
Define, for $c \in (0,1)$, the *c-upper level set* of $F \in \mathcal{F}$ and its plug-in estimator

$$L(c) = \{x \in \mathbb{R}_+^d : F(x) \geq c\}, \quad L_n(c) = \{x \in \mathbb{R}_+^d : F_n(x) \geq c\},$$

and

$$\partial L(c) = \{x \in \mathbb{R}_+^d : F(x) = c\}.$$

In addition, given $T > 0$, we set

$$L(c)^T = \{x \in [0,T]^d : F(x) \geq c\}, \;\; L_n(c)^T = \{x \in [0,T]^d : F_n(x) \geq c\},$$
$$\partial L(c)^T = \{x \in [0,T]^d : F(x) = c\}.$$

Note that, in the presence of a plateau at level c, $\partial L(c)$ can be a portion of quadrant $\mathbb{R}_+^d$ instead of a set of Lebesgue measure null in $\mathbb{R}_+^d$. In the statement of our results we will require suitable conditions in order to avoid this situation.

We denote by $B(x, \rho)$ the closed ball centered on $x \in \mathbb{R}_+^d$ and with positive radius ρ. Let $B(S, \rho) = \bigcup_{x \in S} B(x, \rho)$, with S a closed set of $\mathbb{R}_+^d$. For $\kappa > 0$ and $\zeta > 0$, define

$$E = B(\{x \in \mathbb{R}_+^d : \mid F(x) - c \mid \leq \kappa\}, \zeta),$$

and, for a twice differentiable function F,

$$m^{\nabla} = \inf_{x \in E} \|(\nabla F)_x\|, \qquad M_H = \sup_{x \in E} \|(HF)_x\|_{\mathcal{M}},$$

where $(\nabla F)_x$ is the gradient vector of F evaluated at x and $\|(\nabla F)_x\|$ its Euclidean norm, $(HF)_x$ the Hessian matrix evaluated in x and $\|(HF)_x\|_{\mathcal{M}}$ its matrix norm induced by the Euclidean norm.

We define:

$$\|g\|_p := \left(\int_{\mathbb{R}^d} \mid g(x) \mid^p \; f(x) \, \mathrm{d}x \right)^{1/p},$$

where f denotes the density function associated to the probability measure μ,

$$\|g\|_{p,\lambda} := \left(\int_{\mathbb{R}^d} \mid g(x) \mid^p \; \mathrm{d}x \right)^{1/p} \qquad \text{and} \qquad \|g\|_{\infty} := \sup \left\{ \mid g(x) \mid : \; x \, \in \, \mathbb{R}^d \right\}.$$

For the sake of simplicity, we use from now on the notation $\|.\|$ instead of $\|.\|_2$.

Let $r(x)$ be the regression function such that

$$r(x) := \mathbb{E}[Y \mid \mathbf{X} = x]$$

and let r_n be a consistent estimate of r.

Finally, given two functions f and g, $f(n) = o(g(n))$ as $n \to \infty$ means that for every positive constant ϵ there exists a constant N such that $|f(n)| \leq \epsilon |g(n)|$, for all $n \geq N$. The $o(\cdot)$-notation is called in the literature the Landau symbol.

3 Estimating level sets of a multidimensional distribution function using a plug-in method

We consider the consistency in terms of the volume (in the Lebesgue measure sense) of the symmetric difference between $L(c)^{T_n}$ and $L_n(c)^{T_n}$. This means that we define the distance between two subsets A_1 and A_2 of $\mathbb{R}^d_+$ by

$$d_\lambda(A_1, A_2) = \lambda(A_1 \triangle A_2),$$

where λ stands for the Lebesgue measure on $\mathbb{R}^d$ and $\triangle$ for the symmetric difference.

Let us introduce the following assumption:

A1 There exist positive increasing sequences $(v_n)_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ such that $v_n \to \infty$, $T_n \to \infty$, and

$$v_n \int_{[0,T_n]^d} \mid F(x) - F_n(x) \mid^p \lambda(\mathrm{d}x) \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0, \text{ for some } 1 \leq p < \infty.$$

We now establish our consistency result with convergence rate.

Theorem 1 *Let c be in $(0,1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on $\mathbb{R}^{d*}_+$. Assume that there exist $\kappa > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : \mid t - c \mid \leq \kappa$, $\partial L(t)^{T_1} \neq \emptyset$. Assume that for each n, F_n is measurable. Let $(v_n)_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ positive increasing sequences such that Assumption* **A1** *is satisfied. Then, it holds that*

$$p_n \, d_\lambda(L(c)^{T_n}, L_n(c)^{T_n}) \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0,$$

with p_n an increasing positive sequence such that $p_n = o\left(v_n^{\frac{1}{p+1}} / T_n^{\frac{(d-1)\,p}{p+1}} \right)$.

Remark 1 Note that in the univariate case ($d = 1$) the convergence rate of Theorem 1 does not depend on the truncation sequence T_n. The interested reader is referred to the proof of this result in Section 10 for further details.

Theorem 1 provides a convergence rate, which is closely related to the choice of the sequence T_n, for $d > 1$. Note that, as in Theorem 3 in Cuevas et al (2006), Theorem 1 above does not require any continuity assumption on F_n. Furthermore, we remark that a sequence T_n whose divergence rate is large, implies a convergence rate p_n quite slow. Moreover, this phenomenon is emphasized by the dimension d of the data, and we face here the well-known *curse of dimensionality*. In the following we will illustrate this aspect by giving convergence rate in the case of the empirical distribution function (see Example 1). Firstly, from Theorem 1 we can derive the following result.

Corollary 1 *Let $c \in (0,1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on $\mathbb{R}_+^{d*}$. Assume that there exist $\kappa > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : \mid t - c \mid \leq \kappa$, $\partial L(t)^{T_1} \neq \emptyset$. Assume that for each n, F_n is measurable. Assume that there exists a positive increasing sequence $(w_n)_{n \in \mathbb{N}^*}$ such that $w_n \, \| F - F_n \|_\infty \underset{n \to \infty}{\overset{\mathbb{P}}{\to}} 0$. Then, it holds that*

$$p_n \, d_\lambda(L(c)^{T_n}, L_n(c)^{T_n}) \underset{n \to \infty}{\overset{\mathbb{P}}{\to}} 0,$$

with p_n an increasing positive sequence such that $p_n = o\left(w_n^{\frac{p}{p+1}} / T_n^{\frac{d+(d-1)\,p}{p+1}} \right)$.

This result comes trivially from Theorem 1 and the fact that $w_n \, \|F - F_n\|_\infty \underset{n \to \infty}{\overset{\mathbb{P}}{\to}} 0$ implies

$$\forall \; p \geq 1, \quad v_n \int_{[0,T_n]^d} \mid F - F_n \mid^p \; \lambda(\mathrm{d}x) \underset{n \to \infty}{\overset{\mathbb{P}}{\to}} 0, \quad \text{with} \quad v_n = \frac{w_n^p}{T_n^d}.$$

Let us now present a more practical example in the case of a d-variate empirical distribution function.

Example 1 Let F_n the d-variate empirical distribution function. Then, it holds that $w_n \, \|F - F_n\|_\infty \underset{n \to \infty}{\overset{\mathbb{P}}{\to}} 0$, with $w_n = o(\sqrt{n})$. From Corollary 1, with $p = 2$, we obtain for instance:

$$p_n = o\left(\frac{n^{1/3}}{T_n^{\,7/3}} \right), \quad \text{for} \quad d = 3; \qquad p_n = o\left(\frac{n^{1/3}}{T_n^{\,10/3}} \right), \quad \text{for} \quad d = 4.$$

4 L_p-consistency of r_n on the level sets

In this section we study the L_p-consistency of an estimator r_n of the regression function $r(x)$, for $x \in L(c)$. More precisely, in the first part we provide a consistency result in terms of the L_p-distance of the absolute error between $r_n \, \mathbf{1}_{L_n(c)^{T_n}}$ and $r \, \mathbf{1}_{L(c)}$ (see Theorem 2). In the second part we analyze the problem of a convergence rate (see Theorem 3). Remind that the level set $L(c)$ is not compact. The point-wise estimation of $r(x)$ for $x \in L(c)$ is an interesting aim, for different practical problem, as previously discussed in the introduction of this paper. Then, since the final goal is to study the behavior of the regression function $r(x)$, knowing that x is in the "risk area" $L(c)$, it is natural to check if $r_n(x)$, for x in the estimated level set $L_n(c)$ is globally close to $r(x)$, for x in the theoretical level set $L(c)$.

Let us introduce the following assumptions:

A2 $\|F_n - F\|_{p,\lambda} \xrightarrow{\mathbb{P}} 0$ and $\|r_n - r\|_p \xrightarrow{\mathbb{P}} 0,$ for some $1 \leq p < \infty$.

A3 The density function f of $\mathbf{X}$ is such that $\|f\|_{1+\epsilon,\lambda} < \infty$, with $\epsilon > 0$.

Theorem 2 *Let c be in $(0,1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on $\mathbb{R}_+^{d*}$ with an associated density f such that Assumption* **A3** *is satisfied. Assume that there exist $\kappa > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Assume that for each n, F_n is measurable. Assume that condition* **A2** *is satisfied. Let $(T_n)_{n\in\mathbb{N}^*}$ be an increasing sequence of positive values with $T_1 > 0$ such that for all $t : |\, t - c \,| \leq \kappa,\ \partial L(t)^{T_1} \neq \emptyset$. Then it holds that*

$$\|r_n\, \mathbf{1}_{L_n(c)^{T_n}} - r\, \mathbf{1}_{L(c)}\|_p \xrightarrow{\mathbb{P}} 0, \qquad \textit{for } n \to \infty.$$

Remark 2 If $\|r_n - r\|_\infty \to 0$, *a.s.* we have

$$\|r_n\, \mathbf{1}_{L_n(c)} - r\, \mathbf{1}_{L(c)}\|_\infty = \sup_{\mathbb{R}^d} |r_n\, \mathbf{1}_{L_n(c)} - r\, \mathbf{1}_{L(c)}| \nrightarrow 0, \ \ a.s.$$

Indeed, we have $\sup_{\mathbb{R}^d} |r_n\, \mathbf{1}_{L_n(c)} - r\, \mathbf{1}_{L(c)}| = \max_{x\in L_n(c)\triangle L(c)}(r(x),\, r_n(x))$ which is greater than zero if $L_n(c)\triangle L(c) \neq \emptyset$.

Let us introduce the following assumption:

A4 There exist positive increasing sequences $(v_{1,n})_{n\in\mathbb{N}^*}$, $(v_{2,n})_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ such that $v_{1,n} \to \infty,\ v_{2,n} \to \infty,\ T_n \to \infty$, and

$$v_{1,n}\, \|r_n - r\|_p \xrightarrow{\mathbb{P}} 0, \qquad v_{2,n} \int_{[0,T_n]^d} |\, F(x) - F_n(x) \,|^p \ \lambda(\mathrm{d}x) \xrightarrow{\mathbb{P}} 0,$$

for some $1 \leq p < \infty$.

Note that the control of the convergence rate of the distribution function estimate which was previously handled by Assumption **A1** is now included in Assumption **A4**.

Theorem 3 *Let c be in $(0,1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on $\mathbb{R}_+^{d*}$ with an associated density f such that Assumption* **A3** *is satisfied. Assume that there exist $\kappa > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : |\, t - c \,| \leq \kappa,\ \partial L(t)^{T_1} \neq \emptyset$. Assume that for each n, F_n is measurable. Let $(v_{1,n})_{n\in\mathbb{N}^*}$, $(v_{2,n})_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ positive increasing sequences such that Assumption* **A4** *is satisfied. Then it holds that*

$$w_n \|r_n\, \mathbf{1}_{L_n(c)^{T_n}} - r\, \mathbf{1}_{L(c)^{T_n}}\|_p \xrightarrow{\mathbb{P}} 0, \qquad \textit{for } n \to \infty.$$

where $w_n = \min\{v_{1,n}, a_n\}$ with $a_n = o\left(v_{2,n}^{\frac{\epsilon}{(\epsilon+1)p(p+1)}} / T_n^{\frac{\epsilon(d-1)}{(\epsilon+1)(p+1)}} \right)$.

Remark 3 Note that we do not provide a convergence rate for $\|r_n \mathbf{1}_{L_n(c)^{T_n}} - r \mathbf{1}_{L(c)}\|_p$ because it implies the knowledge of the rate of convergence of $\|r \mathbf{1}_{L(c)^{T_n}} - r \mathbf{1}_{L(c)}\|_p = \left(\int_{L(c)\setminus[0,T_n]^d} |r(x)|^p f(x) dx\right)^{1/p}$. Indeed the rate of convergence of $\mathbb{P}[X_1 > T_n, \ldots, X_d > T_n] \to 0$, for $n \to \infty$, is unknown. In order to overcome this problem we should have to assume that the vector $\mathbf{X}$ belongs to a particular distribution class in order to know the rate of decay of its multivariate tails.

5 Covariate-Conditional-Tail-Expectation consistency

A risk measure has recently received growing attention in risk theory literature: the CTE measure. According to Artzner et al (1999), Dedu and Ciumara (2010), Denuit et al (2005), for a continuous univariate loss distribution function F_X the CTE at level $c \in (0,1)$ is defined by

$$\mathrm{CTE}_c(X) = E[\, X \mid X \geq F_X^{-1}(c)].$$

Several multivariate generalizations of the classical univariate CTE have been proposed (see for instance Cai and Li (2005)). Using the same approach as Di Bernardino et al (2013), we define the multivariate CTE in such a way as to preserve both the complete information about dependence structure between $\mathbf{X}$ and Y, and the marginal behavior of each component.

Definition 1 Consider a random vector $\mathbf{X}$ with distribution function F and a random variable Y. For $c \in (0,1)$, we define the theoretical multivariate c-Covariate-Conditional-Tail-Expectation as

$$\mathrm{CCTE}_c(\mathbf{X}, Y) = \mathbb{E}\left[Y \mid \mathbf{X} \in L(c)\right].$$

Using the truncated version of the c-upper level set defined in Section 2 we also define

$$\mathrm{CCTE}_c^{T_n}(\mathbf{X}, Y) = \mathbb{E}\left[Y \mid \mathbf{X} \in L(c)^{T_n}\right].$$

In the following we define consistent estimates for the CCTE in Definition 1.

5.1 Covariable Y is measured

In this subsection, we assume to have an i.i.d. sample $\{(\mathbf{X}_i, Y_i)\}_{i=1,\ldots,n}$ (see Section 2) and we introduce the following estimate for the CCTE.

Definition 2 Consider a random vector $\mathbf{X}$ with distribution function F and a random variable Y. For $c \in (0,1)$, using the truncated theoretical multivariate c-Covariate-Conditional-Tail-Expectation introduced in Definition 1, we define the associated estimate as

$$\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X},Y) = \mathbb{E}_n\left[Y|\mathbf{X} \in L_n(c)^{T_n}\right],$$

where $\mathbb{E}_n$ denotes the empirical version of the expected value.

Using Definition 2, we now establish the consistency of this estimate with a convergence rate.

Theorem 4 *Let c be in $(0,1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on $\mathbb{R}_+^{d*}$ with an associated density f such that Assumption* **A3** *is satisfied. Assume that there exist $\kappa > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : \mid t - c \mid \leq \kappa$, $\partial L(t)^{T_1} \neq \emptyset$. Assume that for each n, F_n is measurable. Let $(v_{1,n})_{n\in\mathbb{N}^*}$, $(v_{2,n})_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ positive increasing sequences such that Assumption* **A4** *is satisfied. We have*

$$\beta_n \left| \widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X},Y) - \mathrm{CCTE}_c^{T_n}(\mathbf{X},Y) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

with $\beta_n = \min\left(p_n^{\frac{\epsilon}{2(1+\epsilon)}}, d_n\right)$, where $p_n = o\left(v_{2,n}^{\frac{p}{p+1}} / T_n^{\frac{d+(d-1)p}{p+1}}\right)$ and $d_n = o\left(\sqrt{n}\right)$.

Remark 4 Obviously, it could also be very interesting to consider the convergence $\left|\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X},Y) - \mathrm{CCTE}_c(\mathbf{X},Y)\right|$. We remark that in this case the speed of convergence will also depend on the convergence rate to zero of $\left|\mathrm{CCTE}_c^{T_n}(\mathbf{X},Y) - \mathrm{CCTE}_c(\mathbf{X},Y)\right|$, then, in particular of $\mathbb{P}[\mathbf{X} \in L(c) \setminus L(c)^{T_n}]$, for $n \to \infty$. This could not be done without adding strong assumptions and careful developments and, by consequence, remains out of the scope of this paper.

Example 2 Let F_n the d-variate empirical distribution function. Then, it holds that $v_{2,n} = o\left(\sqrt{n}\right)$. From Theorem 4, with $p = 2$, we obtain for instance:

$$p_n = o\left(\frac{n^{1/3}}{T_n^{7/3}}\right), \text{ for } d = 3.$$

This gives us $\beta_n = \min\left(o\left(\frac{n^{1/3}}{T_n^{7/3}}\right)^{\frac{\epsilon}{2(1+\epsilon)}}, o\left(n^{1/2}\right)\right)$. In the case of a bounded density function f we obtain $\beta_n = \min\left(o\left(\frac{n^{1/3}}{T_n^{7/3}}\right)^{\frac{1}{2}}, o\left(n^{1/2}\right)\right) = o\left(\frac{n^{1/6}}{T_n^{7/6}}\right)$.

5.2 Covariable Y is partially unknown

In this subsection, we deal with a more difficult case. We suppose that the covariable Y cannot be measured for all the individuals. It could happen if a measure of Y is very expensive or invasive (in some medical treatment, for example). So we have two different i.i.d. samples: $S_N^1 = \{(\mathbf{X}_i, Y_i)\}_{i=1}^N$ and $S_n^2 = \{\mathbf{X}_j\}_{j=1}^n$, with n potentially much bigger than N.

In this case we use S_N^1 to get an estimate r_N of the regression function r. Then, we apply this estimate on the sample S_n^2 in order to estimate the CCTE measure. To this aim we define:

Definition 3 Consider a random vector $\mathbf{X}$ with distribution function F, a random variable Y and two i.i.d. samples S_N^1 and S_n^2. Using the truncated version of the c-upper level set defined in Section 2, for $c \in (0,1)$, we define

$$\widehat{\mathrm{CCTE}}^{\star\,T_n}_{c,n,N}(\mathbf{X},Y) = \mathbb{E}_n\left[r_N(\mathbf{X})|\mathbf{X} \in L_n(c)^{T_n}\right].$$

The following result proves the consistency of the estimate introduced in Definition 3.

Theorem 5 *Let c be in $(0,1)$. Let $F \in \mathcal{F}$ be a twice differentiable distribution function on $\mathbb{R}_+^{d*}$ with an associated density f such that Assumption* **A3** *is satisfied. Assume that there exist $\kappa > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : \mid t - c \mid \leq \kappa$, $\partial L(t)^{T_1} \neq \emptyset$. Assume that for each n, F_n is measurable. Let $(v_{1,N})_{N\in\mathbb{N}^*}$, $(v_{2,n})_{n\in\mathbb{N}^*}$ and $(T_n)_{n\in\mathbb{N}^*}$ positive increasing sequences such that Assumption* **A4** *is satisfied. Let r be a continuous positive regression function such that $E\left[r(\mathbf{X})^2\right] < \infty$. We have*

$$\beta_{n,N}\left|\widehat{\mathrm{CCTE}}^{\star\,T_n}_{c,n,N}(\mathbf{X},Y) - \mathrm{CCTE}_c^{T_n}(\mathbf{X},Y)\right| \xrightarrow{\mathbb{P}} 0, \qquad \textit{for } N,n \to \infty,$$

with $\beta_{n,N} = \min\left(p_n^{\frac{\epsilon}{2(1+\epsilon)}}, c_N, d_n\right)$, *where* $p_n = o\left(v_{2,n}^{\frac{p}{p+1}} / T_n^{\frac{d+(d-1)p}{p+1}}\right)$, $c_N = o\left(\mathbb{E}\left|r_N(\mathbf{X}) - r(\mathbf{X})\right|\right)$ *and* $d_n = o\left(\sqrt{n}\right)$.

Note that we have a supplementary term (c_N) comparing to Theorem 4. This term controls the rate of convergence of r_N toward r.

Example 3 Using the same context as in Example 2, taking r_N the kernel regression estimator of Kohler et al (2009), assuming that r is a $(1+k,C)$-smooth with $0 < k \leq 1$ function[1] and that there exist $\psi > 2(1+k)$ such that $\mathbb{E}\|\mathbf{X}\|^\psi < \infty$, we have $c_N = o\left(N^{2/5}\right)$. Then we obtain

$$\beta_{n,N} = \min\left(o\left(\frac{n^{1/3}}{T_n^{7/3}}\right)^{\frac{\epsilon}{2(1+\epsilon)}}, o\left(N^{2/5}\right), o\left(n^{1/2}\right)\right).$$

In the case of a bounded density function f and as long as $N >> \frac{n^{5/12}}{T_n^{35/12}}$, we obtain

$$\beta_{N,n} = \min\left(o\left(\frac{n^{1/3}}{T_n^{7/3}}\right)^{\frac{1}{2}}, o\left(N^{2/5}\right), o\left(n^{1/2}\right)\right) = o\left(\frac{n^{1/6}}{T_n^{7/6}}\right),$$

which is the same rate as in Example 2.

[1] This requires that the partial derivatives of the regression function r are k-Hölderian with a constant C (for further details see Definition 1 in Kohler et al (2009)).

6 About the effects of scaling data

In this section we study the impact of a change in the scale of data with respect to the estimate $\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X}, Y)$. In particular, this property is related to the suitable positive homogeneity property of risk measures (e.g. see Artzner et al (1999)). In the literature, the homogeneity property of a risk measure is often motivated by a change of currency argument: the amount of required capital in order to manage risks should be independent of the currency in which it is expressed. Suppose that we scale our data using a scale parameter $a \in \mathbb{R}_+^*$. In our case, the scaled random vector will be $(a\, X_1, a\, X_2, \ldots, a\, X_d) =: a\, \mathbf{X}$. From now on we denote $F_{a\,\mathbf{X}}$ (resp. $F_{\mathbf{X}}$) the distribution function associated to $a\,\mathbf{X}$ (resp. to $\mathbf{X}$). Using notation of Section 2, let

$$L_a(c) = \{x \in \mathbb{R}_+^d : F_{a\,\mathbf{X}}(x) \geq c\}.$$

We can now consider the effects of scaling data with respect to the volume of the symmetric difference. Corollary 2 is a straightforward consequence of Theorem 1 using the fact that $L_a^{T_n}(c) \triangle L_{n,a}^{T_n}(c) = a\left(L^{T_n}(c) \triangle L_n^{T_n}(c)\right)$, for$a \in \mathbb{R}_+^*$.

Corollary 2 *Assume that $F_{n,a\mathbf{X}}(x) = F_{n,\mathbf{X}}(x/a)$.*
Under same notations and assumptions of Theorem 1 it holds that

$$p_{n,\,a}\, d_\lambda(L_a(c)^{T_n}, L_{n,\,a}(c)^{T_n}) \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0,$$

with $p_{n,\,a}$ an increasing positive sequence such that $p_{n,\,a} = o\left(v_n^{\frac{1}{p+1}} / \left(a^d\, T_n^{\frac{(d-1)\,p}{p+1}}\right)\right)$.

Corollary 2 states that a change of scale of the data implies a convergence rate as

$$o\left(v_n^{\frac{1}{p+1}} / \left(a^d\, T_n^{\frac{(d-1)\,p}{p+1}}\right)\right) \text{ instead of } o\left(v_n^{\frac{1}{p+1}} / T_n^{\frac{(d-1)p}{p+1}}\right).$$

Logically, it means that the scale factor $a \in \mathbb{R}_+^*$ impacts the volume in $\mathbb{R}^d$ with an exponent d.

Let us now investigate the effects of scaling data on Theorem 3 and Theorem 4. Let denote $r_{a\mathbf{X},bY}(x) = E[b\,Y \,|\, a\,\mathbf{X} = x]$ ($r_{n,a\mathbf{X},bY}$ its estimate version) and $\mathrm{CCTE}_c^{T_n}(a\mathbf{X}, bY) = \mathbb{E}\left[b\,Y \,|\, a\,\mathbf{X} \in L_a(c)^{T_n}\right]$ ($\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(a\mathbf{X}, bY)$ its estimate version). We obtain the following result.

Theorem 6 *Let a and b in $\mathbb{R}_+^*$ and assume that $r_{n,a\mathbf{X},bY}(x) = b\, r_{n,\mathbf{X},Y}\left(\frac{x}{a}\right)$.*

1. *Under same hypotheses and notations of Theorem 3 we have*

$$b\, \| r_{n,\mathbf{X},Y}\, \mathbf{1}_{L_n(c)^{T_n}} - r_{\mathbf{X},Y}\, \mathbf{1}_{L(c)^{T_n}} \|_p = \\ \| r_{n,a\mathbf{X},bY}\, \mathbf{1}_{L_{a,n}(c)^{T_n}} - r_{a\mathbf{X},bY}\, \mathbf{1}_{L_a(c)^{T_n}} \|_p.$$

2. Under same hypotheses and notations of Theorem 4 we have

$$b\left|\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X},Y)\;-\;\mathrm{CCTE}_c^{T_n}(\mathbf{X},Y)\right| =$$
$$\left|\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(a\mathbf{X},bY)-\mathrm{CCTE}_c^{T_n}(a\mathbf{X},bY)\right|.$$

Using Theorem 6 we obtain that the rates of convergence of Theorem 3 and Theorem 4 are not affected by any scaling on the data.

Remark 5 Note that if r_n is the classical kernel estimator, assumption $r_{n,a\mathbf{X}}(x) = r_{n,\mathbf{X}}\left(\frac{x}{a}\right)$ is not automatically satisfied. However, it can be satisfied if the scaling is also applied to the bandwidth.

7 Illustrations

In the following we consider some different simulated cases for which we illustrate the finite sample properties of our estimation of r and $\mathrm{CCTE}_c(\mathbf{X},Y)$. In particular, we will consider an independent copula (Section 7.1), and Ali-Mikhail-Haq copula (Section 7.2). To compare the estimated results with the theoretical ones we consider cases for which we can calculate (using Maple) the explicit value of the theoretical $\mathrm{CCTE}_c(\mathbf{X},Y)$. However, our estimator can be applied to much more general cases.

In this section we consider the kernel regression estimate proposed by Kohler et al (2009) in order to estimate r. Furthermore, the plug-in estimation of level sets, i.e., $L_n(c)$, is constructed using the empirical estimator F_n of the distribution function. Remark that these two estimates both satisfy Assumptions **A1**, **A2** and **A4**, and that the considered random vectors $\mathbf{X}$ in this section satisfy the conditions required for the kernel regression estimate (see Example 3 page 11).

7.1 Independent Copula

We consider here a bivariate independent copula with exponentially distributed marginals with parameter 1. Furthermore we choose the bounded regression function $y = r(x_1,x_2) = 1/(\mathrm{e}^{x_1}\mathrm{e}^{x_2})$, for all $x_1,x_2 \geq 0$ (Figure 1).

7.1.1 L_p-consistency of r_n

In this section we provide an illustration of Theorem 3. Remark that the assumptions of Theorem 3 are satisfied. In particular, we denote

$$E_{L_2} := \|r_n\,\mathbf{1}_{L_n(c)^{T_n}} - r\,\mathbf{1}_{L(c)^{T_n}}\|_2.$$

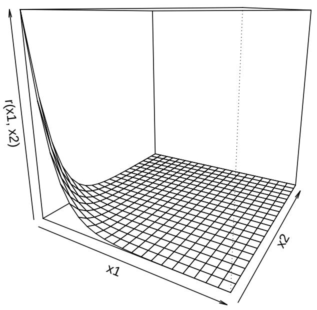

Figure 1 Regression function: $y = r(x_1, x_2) = 1/(\mathrm{e}^{x_1}\mathrm{e}^{x_2})$ for all $x_1, x_2 \geq 0$.

According to Remark 3, we take $T_n = \ln(n)$ (the interested reader is also referred to Di Bernardino et al (2013)). In Table 1 we illustrate the mean of E_{L_2} (i.e., L_2-consistency for the estimator r_n) on $M = 50$ simulated samples, for different level of risk c and different sample size n.

n	$c = 0.1$	$c = 0.25$	$c = 0.5$	$c = 0.7$	$c = 0.9$
$n = 1000$	0.0567	0.0377	0.0254	0.0197	0.0131
$n = 5000$	0.0323	0.0176	0.0072	0.0051	0.0024
$n = 10000$	0.0256	0.0134	0.0051	0.0025	0.0016

Table 1 E_{L_2} for independent copula and exponentially distributed marginals with parameter 1. The bounded regression function is $y = r(x_1, x_2) = 1/(\mathrm{e}^{x_1}\mathrm{e}^{x_2})$ for all $x_1, x_2 \geq 0$.

As expected, the performance of the estimation increases with the size of the sample. However, the fact that the quality for the E_{L_2} error seems better for large risk level c is quite surprising. This behavior of the estimation comes from the fact that the regression r has a flat plateau for $x_1, x_2 \to \infty$ (see Figure 1).

7.1.2 $\mathrm{CCTE}_c(\mathbf{X}, Y)$ *estimation*

In the following, we compare $\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X}, Y)$ with the theoretical $\mathrm{CCTE}_c(\mathbf{X}, Y)$. Following Remark 4, we obtain that $\left|\,\mathrm{CCTE}_c(\mathbf{X}, Y) - \widehat{\mathrm{CCTE}}_c^{T_n}(\mathbf{X}, Y)\,\right|$ decays to zero at least with a convergence rate $o\left(\frac{n^{1/6}}{\ln(n)^{4/6}}\right)$, with a choice of sequence $T_n = \ln(n)$. This kind of compromise provides an illustration on how to choose T_n, apart from satisfying the assumptions of Theorem 4.

In the following tables we denote Mean= $\overline{\widehat{\mathrm{CCTE}}_c^{T_n}}(\mathbf{X},Y)$, i.e., the mean of $\widehat{\mathrm{CCTE}}_c^{T_n}(\mathbf{X},Y)$ on $M=50$ simulated samples. We denote $\widehat{\sigma}$ the empirical standard deviation

$$\widehat{\sigma}=\sqrt{\frac{1}{M-1}\sum_{j=1}^{M}\left(\widehat{\mathrm{CCTE}}_c^{T_n}(\mathbf{X},Y)_j-\overline{\widehat{\mathrm{CCTE}}_c^{T_n}}(\mathbf{X},Y)\right)^2}.$$

Finally, we denote RMAE the Relative Mean Absolute Error, i.e.,

$$\mathrm{RMAE}=\frac{1}{M}\sum_{j=1}^{M}\frac{\left|\widehat{\mathrm{CCTE}}_c^{T_n}(\mathbf{X},Y)_j-\mathrm{CCTE}_c(\mathbf{X},Y)\right|}{\mathrm{CCTE}_c(\mathbf{X},Y)}.$$

The results are gathered in Table 2. Furthermore, in Table 3, we provide an illustration of the convergence rate of Theorem 4. In this case, we remark that $\beta_n=\frac{n^{1/6}}{\ln(n)^{4/6}}$ is at least the convergence rate of this CCTE estimation.

n		$c=0.1$ CCTE=0.142	$c=0.25$ CCTE=0.0792	$c=0.5$ CCTE=0.0279	$c=0.7$ CCTE=0.0087	$c=0.9$ CCTE=0.0009
1000	RMAE	0.0357	0.0388	0.0725	0.1187	0.2898
	Mean	0.1432	0.0798	0.0281	0.0090	0.0009
	$\widehat{\sigma}$	0.0062	0.0038	0.0026	0.0012	0.0003
5000	RMAE	0.0151	0.0194	0.0317	0.0484	0.1623
	Mean	0.1425	0.0795	0.0278	0.0087	0.0008
	$\widehat{\sigma}$	0.0027	0.0021	0.0011	0.0004	0.0002
10000	RMAE	0.0110	0.0137	0.0211	0.0453	0.0875
	Mean	0.1428	0.0792	0.0279	0.0088	0.0008
	$\widehat{\sigma}$	0.0019	0.0013	0.0007	0.0004	$9.43\,10^{-5}$

Table 2 Estimation of $\mathrm{CCTE}_c(\mathbf{X},Y)$ in the case of independent copula and exponentially distributed marginals with parameter 1.

	$c=0.1$	$c=0.25$	$c=0.5$	$c=0.7$	$c=0.9$
$n=1000$	0.0311	0.0339	0.0632	0.1035	0.2526
$n=5000$	0.0151	0.0191	0.0314	0.0479	0.1609
$n=10000$	0.0116	0.0145	0.0223	0.0478	0.0925

Table 3 Approximated β_n ·RMAE, with $\beta_n=\frac{n^{1/6}}{\ln(n)^{4/6}}$, in the case of independent copula and exponentially distributed marginals with parameter 1.

As expected, the greater n is, the better the estimations are. Furthermore, results in Table 3 set out how $\beta_n=\frac{n^{1/6}}{\ln(n)^{4/6}}$ is at least the convergence rate of $\left|\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X},Y)-\mathrm{CCTE}_c(\mathbf{X},Y)\right|$, in this particular case.

7.2 Ali-Mikhail-Haq Copula

We consider here a bivariate Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1 (see Nelsen (1999)). We take the bounded regression function $y = r(x_1, x_2) = -(-2 + \mathrm{e}^{-x_1 - x_2})^3$, for all $x_1, x_2 \geq 0$. Conversely to the previous section we now deal with dependent variables $X_1, \ldots, X_n$ (i.e., only the coordinates of the X_i are dependent) and an increasing regression function r (see Figure 2).

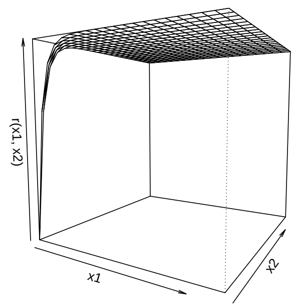

Figure 2 Regression function $y = r(x_1, x_2) = -(-2 + \mathrm{e}^{-x_1 - x_2})^3$ for all $x_1, x_2 \geq 0$.

7.2.1 L_p-consistency of r_n

In this section we provide an illustration of Theorem 3. We can verify that the assumptions of Theorem 3 are satisfied. As explained in Remark 4, we choose $T_n = n^{0.2}$. In Table 4 we illustrate E_{L_2} (i.e. the L_2-consistency for the estimator r_n) for different level of risk c and different sample size n.

n	$c = 0.1$	$c = 0.25$	$c = 0.5$	$c = 0.7$	$c = 0.9$
$n = 1000$	0.7736	0.7980	0.7599	0.5817	0.2928
$n = 5000$	0.4846	0.5801	0.5047	0.3975	0.1781
$n = 10000$	0.4167	0.4883	0.4459	0.3612	0.1679

Table 4 E_{L_2} for Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1. The bounded regression function is $y = r(x_1, x_2) = -(-2 + \mathrm{e}^{-x_1 - x_2})^3$, for all $x_1, x_2 \geq 0$.

As before, we see that the quality of the estimation is better when the size of the sample growths. Again, the presence of a plateau for $r(x_1, x_2)$ when $x_1, x_2 \to \infty$ explains the good results for large risk levels c (see Figure 2).

7.2.2 $\mathrm{CCTE}_c(\mathbf{X}, Y)$ *estimation*

In the following, we compare $\widehat{\mathrm{CCTE}}_{c,n}^{T_n}(\mathbf{X}, Y)$ with the theoretical $\mathrm{CCTE}_c(\mathbf{X}, Y)$ in the case of Ali-Mikhail-Haq copula with parameter 0.5 (see Table 5). Furthermore, following Remark 4, we obtain that $\left| \mathrm{CCTE}_c(\mathbf{X}, Y) - \widehat{\mathrm{CCTE}}_c^{T_n}(\mathbf{X}, Y) \right|$ decays to zero at least with a convergence rate $o\left(n^{1/10}\right)$, with a choice of sequence $T_n = n^{1/5}$. Again, our theoretical results are confirmed by these simulations (see Table 6).

n		$c = 0.1$ CCTE=6.3752	$c = 0.25$ CCTE=7.0393	$c = 0.5$ CCTE=7.6513	$c = 0.7$ CCTE=7.8909	$c = 0.9$ CCTE= 7.9893
500	RMAE	0.0127	0.0091	0.0037	0.0023	0.0005
	Mean	6.3660	7.0283	7.6429	7.8860	7.9899
	$\widehat{\sigma}$	0.0968	0.0830	0.0380	0.0240	0.0051
1000	RMAE	0.0081	0.0061	0.0038	0.0015	0.0004
	Mean	6.3885	7.0421	7.6544	7.8917	7.9887
	$\widehat{\sigma}$	0.0639	0.0526	0.0345	0.0147	0.0041
5000	RMAE	0.0032	0.0028	0.0015	0.0006	0.0001
	Mean	6.3761	7.0458	7.6461	7.8906	7.9891
	$\widehat{\sigma}$	0.0259	0.0249	0.0138	0.0071	0.0013
10000	RMAE	0.0021	0.0023	0.0008	0.0005	0.0001
	Mean	6.3725	7.0302	7.6527	7.8911	7.9893
	$\widehat{\sigma}$	0.0171	0.0173	0.0081	0.0047	0.0012

Table 5 Estimation of $\mathrm{CCTE}_c(\mathbf{X}, Y)$ in the case of Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1.

	$c = 0.1$	$c = 0.25$	$c = 0.5$	$c = 0.7$	$c = 0.9$
$n = 500$	0.0236	0.0169	0.0068	0.0043	0.0009
$n = 1000$	0.0161	0.0121	0.0076	0.0030	0.0008
$n = 5000$	0.0075	0.0066	0.0035	0.0014	0.0002
$n = 10000$	0.0053	0.0057	0.0020	0.0013	0.0002

Table 6 Approximated $\beta_n \cdot \mathrm{RMAE}$, with $\beta_n = n^{1/10}$ in the case of Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1.

A comparative study

In this section we provide a detailed comparative study to illustrate the advantages of our non-parametric estimation procedure. In particular, we compare the performances of the estimation of CCTE_c using a parametric (*resp.*

semi-parametric) estimation procedure based on Maximum Likelihood Estimation (MLE), by fitting a parametric copula model with parametric (*resp.* non-parametric) marginals (see Tables 7, 8 and 9).

We consider the following parametric estimator of the Covariate-Conditional-Tail-Expectation, i.e.,

$$\mathrm{CCTE}_c^P = \mathbb{E}_n[Y|X \in \widehat{L}_{\overrightarrow{\theta},c}], \text{ for } c \in (0,1),$$

where

$$\widehat{L}_{\overrightarrow{\theta},c} = \{x \in \mathbb{R}_+^d : \widehat{F}_{\overrightarrow{\theta},c}(x) \geq c\}, \text{ and } \widehat{F}_{\overrightarrow{\theta},c}(x) = C_{\widehat{\theta}}(\widehat{F}_{\widehat{\theta}_{X_1}}(x_1), \ldots, \widehat{F}_{\widehat{\theta}_{X_d}}(x_d)),$$

with $C_{\widehat{\theta}}$ a copula with parameter $\widehat{\theta}$ fitted by MLE, and $\widehat{F}_{\widehat{\theta}_{X_i}}$ the parametric marginal distribution of X_i, for $i = 1, \ldots, d$ with parameter (or eventually a vector of parameters) $\widehat{\theta}_{X_i}$ fitted by MLE.

We also introduce the semi-parametric estimator of the Covariate-Conditional-Tail-Expectation, i.e.,

$$\mathrm{CCTE}_c^{SP} = \mathbb{E}_n[Y|X \in \widehat{L}_{\theta,n,c}], \text{ for } c \in (0,1),$$

where

$$\widehat{L}_{\theta,n,c} = \{x \in \mathbb{R}_+^d : \widehat{F}_{\theta,n,c}(x) \geq c\}, \; \widehat{F}_{\theta,n,c}(x) = C_{\widehat{\theta}}(F_{X_1,n}(x_1), \ldots, F_{X_d,n}(x_d)),$$

with $C_{\widehat{\theta}}$ a copula with parameter $\widehat{\theta}$ fitted by MLE, and $F_{X_i,n}$ the empirical marginal distribution of X_i, for $i = 1, \ldots, d$.

The idea is to compare the performance of these three different estimators (non-parametric, semi-parametric and parametric). We aim to consider two different situations: when the copula model used in the parametric and semi-parametric estimator is the good one or not. For the second case we consider the Clayton and Gumbel copulas. The Clayton copula is closer to the true one (Ali-Mikhail-Haq) since they are both non-heavy tail copulas (upper-tail dependence coefficient $\lambda_U = 0$, see Nelsen (1999)). Conversely the Gumbel copula presents an upper-tail dependence structure. In this sense the Gumbel copula is substantially different from the others two. The results are gathered in Tables 7, 8 and 9.

Not surprisingly, we see that our estimator performs better than the parametric and the semi-parametric one in the case when the chosen class of copulas does not contain the true model (i.e., here Clayton and Gumbel copulas). And as expected the worst results are for the Gumbel copula as it is the farthest from the true model. We also note that in this case the semi-parametric approach gives no better results than the parametric one. Remark also that our non-parametric estimation is comparable to the parametric and semi-parametric ones in case the chosen class of copulas contains the true model.

	$c = 0.1$	$c = 0.25$	$c = 0.7$	$c = 0.9$
$\widehat{\text{CCTE}}_{c,n}^{T_n}$	0.008171	0.005974	0.001528	0.000348
AMH Copula, CCTE_c^{P}	0.007726	0.005920	0.001637	0.000344
AMH Copula, CCTE_c^{SP}	0.008356	0.006091	0.001617	0.000361
Clayton Copula, CCTE_c^{P}	0.007958	0.007207	0.001566	0.000343
Clayton Copula, CCTE_c^{SP}	0.008696	0.007570	0.001402	0.000377
Gumbel Copula, CCTE_c^{P}	0.011499	0.007324	0.001811	0.000392
Gumbel Copula, CCTE_c^{SP}	0.011699	0.006584	0.001765	0.000483

Table 7 RMAE of the estimations of the CCTE using our estimator, CCTE_c^{SP} (semi-parametric), CCTE_c^{P} (parametric) for Ali-Mikhail-Haq, Clayton and Gumbel copulas for different levels of c. (X_1, X_2) is a vector with Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1. Sample size $\underline{n = 1000}$, Monte-Carlo iterations $M = 300$.

	$c = 0.1$	$c = 0.25$	$c = 0.7$	$c = 0.9$
$\widehat{\text{CCTE}}_{c,n}^{T_n}$	0.003745	0.002851	0.000705	0.000175
AMH Copula, CCTE_c^{P}	0.003586	0.002746	0.000703	0.000151
AMH Copula, CCTE_c^{SP}	0.003912	0.002773	0.000636	0.000162
Clayton Copula, CCTE_c^{P}	0.004497	0.005322	0.000758	0.000158
Clayton Copula, CCTE_c^{SP}	0.004349	0.005113	0.000732	0.000157
Gumbel Copula, CCTE_c^{P}	0.010739	0.005604	0.000892	0.000191
Gumbel Copula, CCTE_c^{SP}	0.010616	0.005554	0.000867	0.000213

Table 8 RMAE of the estimations of the CCTE using our estimator, CCTE_c^{SP} (semi-parametric), CCTE_c^{P} (parametric) for Ali-Mikhail-Haq, Clayton and Gumbel copulas for different levels of c. (X_1, X_2) is a vector with Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1. Sample size $\underline{n = 5000}$, Monte-Carlo iterations $M = 300$.

	$c = 0.1$	$c = 0.25$	$c = 0.7$	$c = 0.9$
$\widehat{\text{CCTE}}_{c,n}^{T_n}$	0.002556	0.002029	0.000479	0.000105
AMH Copula, CCTE_c^{P}	0.002529	0.001949	0.000476	0.000102
AMH Copula, CCTE_c^{SP}	0.002555	0.001992	0.000476	0.000117
Clayton Copula, CCTE_c^{P}	0.003691	0.005542	0.000569	0.000118
Clayton Copula, CCTE_c^{SP}	0.003597	0.005364	0.000620	0.000125
Gumbel Copula, CCTE_c^{P}	0.010701	0.005546	0.000752	0.000170
Gumbel Copula, CCTE_c^{SP}	0.010609	0.005527	0.000768	0.000169

Table 9 RMAE of the estimations of the CCTE using our estimator, CCTE_c^{SP} (semi-parametric), CCTE_c^{P} (parametric) for Ali-Mikhail-Haq, Clayton and Gumbel copulas for different levels of c. (X_1, X_2) is a vector with Ali-Mikhail-Haq copula with parameter 0.5 and exponentially distributed marginals with parameter 1. Sample size $\underline{n = 10000}$, Monte-Carlo iterations $M = 300$.

8 Real data study: waves and water levels in coastal engineering design

On coasts with high tidal ranges, or subject to high surges, both still water levels and waves can be important in assessing flood risk; their relative importance depends on location and on the type of sea defence. The simultaneous occurrence of large waves and a high still water level is, therefore, important in estimating their combined effect on sea defences. In design of a sea defence, a key step is the estimation of the probability of failure to protect against sea conditions. It is important in engineering design to identify the combinations of sea condition variables which cause each failure. The interested reader is referred for instance to Hawkes et al (2002).

For any particular mode of failure (structural failure, excessive overtopping, ...), the regression function r is dependent on the sea condition variables. Then, in particular, at any particular time t, the overtopping covariate Y will be related to the sea condition vector ($\mathbf{X}$). In the literature the sea condition variables are often represented by the significant Wave height Hm0 (X_1), the Still Water level SWL(X_2), and the Wave period Tpb(X_3), then $\mathbf{X} = (X_1, X_2, X_3)$ (see Figure 3).

The regression function $r(x) := \mathbb{E}[Y \,|\, \mathbf{X} = x]$ represents the relationship between the sea conditions and the overtopping at a given time t. This relationship could be complex and in some real analysis, can be represented by equations. The most advantage of the use of this regression function is to reduce a joint probability risk problem to a single covariate problem.

In this section we analyze the Wave height (Hm0), Still Water level (SWL), Wave period (Tpb) data, recorded during 828 storm events spread over 13 years in front of the Dutch coast near the town of Petten (Figure 4).

These data has been recently studied in the literature (for details see for instance Draisma et al (2004)). Following Tau and Dam (2011), at a given time t, the principal equation used for overtopping discharge (l/m/s) Y is given by:

$$Y = a\,\mathrm{e}^{\frac{-b\,(h-\mathrm{SWL})}{\mathrm{Hm0}}}\sqrt{g\,(\mathrm{Hm0})^3}, \tag{8.1}$$

with

$$a = 0.04\sqrt{\frac{\tan(\beta)\,L}{\mathrm{Hm0}}}, \qquad b = \frac{7.05\sqrt{\frac{\mathrm{Hm0}}{L}}}{\tan(\beta)}, \qquad L = \frac{g\,(\mathrm{Tpb})^2}{2\,\pi},$$

where

- Hm0 (m) is the wave height at the toe of the structure at time t;
- Tpb(s) is the wave period at time t. In particular the number of waves in a storm (N) can easily be computed from information about the wave period and the duration of the storm, i.e., $N = (\text{duration(h)}/\text{Tpb(s)}) \cdot 3600$;
- SWL(m) the level of the sea if it is flat, without any waves at time t;
- h (m) is the height of the costal design above SWL= 0 (see Figure 3);
- g (m/s^2) the gravitational acceleration (i.e., 9.8 m/s^2);
- β (rad) is the seaward slope steepness. In the following we consider the case $\tan(\beta) = 0.3$ and $\tan(\beta) = 0.6$ (see Figure 3).

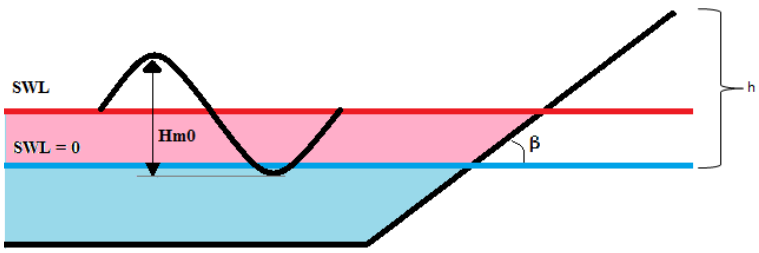

Figure 3 Definition of some parameters for the calculation of overtopping.

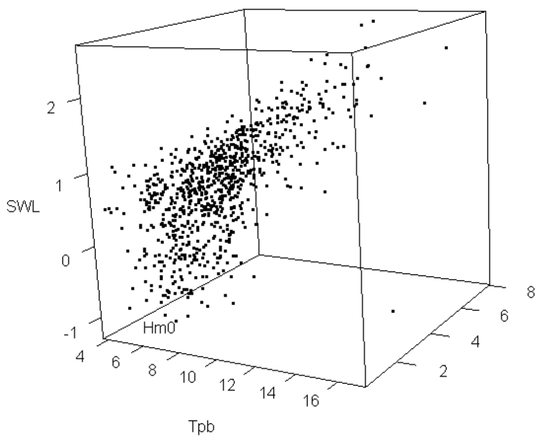

Figure 4 Representation of the data in three dimensions.

Note that Theorem 6 allows us to make any changes on the units before or after the estimation. We now estimate the mean overtopping rate (i.e., $\text{CCTE}_c(\mathbf{X}, Y)$) using Equation (8.1) for Y and conditionally to the fact that the sea variable conditions $\mathbf{X}$ belong to the joint risk area $L(c)$. In particular, we consider the jointly large realisations of sea variables Hm0, SWLand Tpbfor different values of risk level c. Obviously, the dangerous effects of the sea conditions will be increasing with respect to c. The results are gathered in Figure 5.

As expected, we obtain a decreasing behavior of mean overtopping discharge according to the height of the costal design h for a fixed risk level c. Furthermore, for a fixed height of the costal design, the higher the risk c, the higher the mean overtopping discharge. In our study we consider both the case $\tan(\beta) = 0.3$ (Figure 5, left) and $\tan(\beta) = 0.6$ (Figure 5, right). As expected, we remark that for fixed risk level c and height of costal design h, a lower seaward slope steepness β generates a smaller mean overtopping discharge.

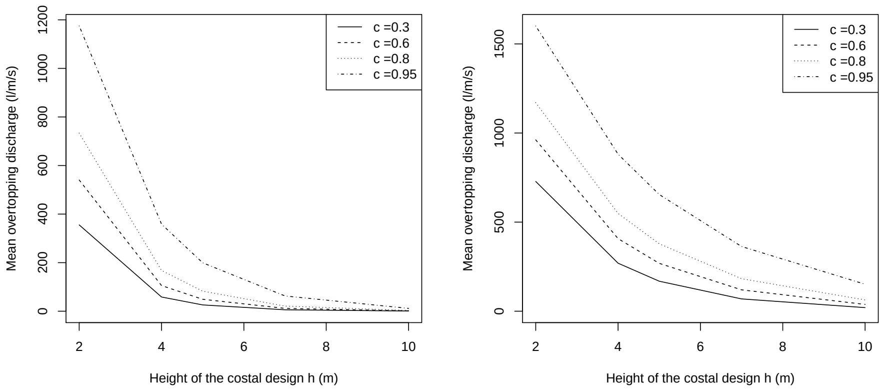

Figure 5 Mean overtopping discharge $\mathrm{CCTE}_c(\mathbf{X}, Y)$ in function of the height of the costal design h for different condition of risk-sea variables. We take here $\tan(\beta) = 0.3$ (left) and $\tan(\beta) = 0.6$ (right).

9 Conclusion

We propose in this paper a generalization to the estimation of the level sets of a d-variate distribution function. The non-compactness of the level sets requires special attention in the statement of the problem. The consistency results with a convergence rate are stated in terms of the volume of the symmetric difference. In a second part, we analyze the problem of the estimation of a regression function on the level sets of a d-variate distribution function and we obtain the consistency with a convergence rate in terms of the L_p-distance. Then, we study a new multivariate risk measure: the Covariate-Conditional-Tail-Expectation, i.e. the Conditional-Tail-Expectation of the regression function. A consistent estimator and a rate of convergence are provided. Moreover, we analyze the impact of scaling data on our results. Our theoretical results are illustrated on a complete simulation study. We discuss a real application in the evaluation of the mean overtopping discharge conditionally to the fact that the sea variable conditions belong to some joint risk area. It highlights the importance of the parameter T_n (which solved the problem of the non-compactness of the level sets) as well as the curse of dimensionality. An interesting future work could be a deep investigation about these points, with a focus on the optimal choice for this parameter. Furthermore, the proposed methods are based on an *i.i.d.* samples framework. We remark that in real applications such as seasonal pattern in the temperature and water level rise series, data can have different types of serial correlations like nonlinear or non-stationary correlations (Fan and Yao 2003). It would be interesting in a future work to include other more interesting and complex types of serial correlation structures and to analyze how this affects the performance of the proposed procedure. Finally, a study on the lower bounds of our estimation problem could be an interesting development of the present work.

10 Proofs

Proof of Theorem 1: Under assumptions of Theorem 1, we can always take $T_1 > 0$ such that for all $t : |\, t - c\,| \leq \kappa,\ \partial L(t)^{T_1} \neq \emptyset$. Then for each n, for all $t : |\, t - c\,| \leq \kappa,\ \partial L(t)^{T_n}$ is a non-empty (and compact) set on $\mathbb{R}^d_+$.

We consider a positive sequence ε_n such that $\varepsilon_n \underset{n\to\infty}{\to} 0$. For each $n \geq 1$ the random sets $L(c)^{T_n} \triangle L_n(c)^{T_n}$, $Q_{\varepsilon_n} = \{x \in [0, T_n]^d : \mid F - F_n \mid \leq \varepsilon_n\}$ and $\widetilde{Q}_{\varepsilon_n} = \{x \in [0, T_n]^d : \mid F - F_n \mid > \varepsilon_n\}$ are measurable and

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) = \lambda(L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n}) + \lambda(L(c)^{T_n} \triangle L_n(c)^{T_n} \cap \widetilde{Q}_{\varepsilon_n}).$$

Since $L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n} \subset \{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}$ we obtain

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) \leq \lambda(\{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}) + \lambda(\widetilde{Q}_{\varepsilon_n}).$$

From assumptions of Theorem 1 and Proposition 2.1 in Di Bernardino et al (2013), it follows that there exists a $\gamma > 0$ such that, if $2\,\varepsilon_n \leq \gamma$ then

$$d_H(\partial L(c + \varepsilon_n)^{T_n}, \partial L(c - \varepsilon_n)^{T_n}) \leq 2\,\varepsilon_n\,A$$

where $A = \frac{2}{m^{\nabla}}$ and d_H is the Hausdorff distance. From assumptions on first derivatives of F and Property 1 in Imlahi et al (1999), we can write

$$\lambda(\{x \in [0, T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}) \leq\ (2\,\varepsilon_n\,A)\,d\,T_n^{d-1}.$$

Interestingly we remark that in the univariate case ($d = 1$) the Hausdorff distance between the two points $\partial L(c - \varepsilon_n)^{T_n}$ and $\partial L(c + \varepsilon_n)^{T_n}$ is also the Lebesgue measure (in dimension 1) for this interval. Then $\lambda(\{x \in [0, T_n] : c - \varepsilon_n \leq F < c + \varepsilon_n\}) \leq\ 2\,\varepsilon_n\,A$. This means that in this case, the result does not depend on the truncation sequence T_n.

If we now choose

$$\varepsilon_n = o\left(\frac{1}{p_n\,T_n^{d-1}}\right), \tag{10.1}$$

we obtain that, for n large enough, $2\,\varepsilon_n \leq \gamma$ and

$$p_n\,\lambda(\{x \in [0,T_n]^d : c - \varepsilon_n \leq F < c + \varepsilon_n\}) \underset{n\to\infty}{\to} 0.$$

Let us now prove that $p_n\,\lambda(\widetilde{Q}_{\varepsilon_n}) \overset{\mathbb{P}}{\underset{n\to\infty}{\to}} 0$. To this end, we write

$$p_n\,\lambda(\widetilde{Q}_{\varepsilon_n}) \;=\; p_n \int 1_{\{[0,T_n]^d :\, |F-F_n| > \varepsilon_n\}}\,\lambda(\mathrm{d}x) \leq \frac{p_n}{\varepsilon_n^p} \int_{[0,T_n]^d} |\; F - F_n \;|^p \;\; \lambda(\mathrm{d}x).$$

Take ε_n such that

$$\varepsilon_n = \left(\frac{p_n}{v_n}\right)^{\frac{1}{p}}. \tag{10.2}$$

So, from Assumption **A1**, we obtain $p_n\,\lambda(\widetilde{Q}_{\varepsilon_n}) \overset{\mathbb{P}}{\underset{n\to\infty}{\to}} 0$. As $p_n = o\left(v_n^{\frac{1}{p+1}} / T_n^{\frac{(d-1)\,p}{p+1}}\right)$ we can choose ε_n that satisfies (10.1) and (10.2). Hence the result. □

In the following proofs, K denotes a constant whose value may change from line to line.

Proof of Theorem 2: We have

$$\begin{aligned}
&\|r\,\mathbf{1}_{L_n(c)^{T_n}} - r\,\mathbf{1}_{L(c)^{T_n}}\|_p \\
&\qquad = \left(\int_{\mathbb{R}^d} \left|r(x)\,\mathbf{1}_{L_n(c)^{T_n}} - r(x)\,\mathbf{1}_{L(c)^{T_n}}\right|^p f(x)\,\mathrm{d}x\right)^{1/p} \\
&\qquad \leq \left(\int_{L(c)^{T_n}\triangle L_n(c)^{T_n}} |r(x)|^p\,f(x)\,\mathrm{d}x\right)^{1/p} \\
&\qquad \leq K\,\left(\mu\left(L(c)^{T_n}\triangle L_n(c)^{T_n}\right)\right)^{1/p} \\
&\qquad \leq K\left(\|f\|_{1+\epsilon,\lambda}\,\left(\lambda\left(L(c)^{T_n}\triangle L_n(c)^{T_n}\right)\right)^{\frac{\epsilon}{1+\epsilon}}\right)^{1/p}
\end{aligned}$$

which gives us

$$\begin{aligned}
\|r_n\,\mathbf{1}_{L_n(c)^{T_n}} - r\,\mathbf{1}_{L(c)^{T_n}}\|_p &\leq \|r_n - r\|_p \\
&\quad + \;K\,\left(\|f\|_{1+\epsilon,\lambda}\,\left(\lambda\left(L(c)^{T_n}\triangle L_n(c)^{T_n}\right)\right)^{\frac{\epsilon}{1+\epsilon}}\right)^{1/p}.
\end{aligned}$$

Since

$$\|r\,\mathbf{1}_{L(c)^{T_n}} - r\,\mathbf{1}_{L(c)}\|_p = \left(\int_{L(c)\setminus[0,T_n]^d} |r(x)|^p f(x) dx\right)^{1/p},$$

we finally get

$$
\begin{aligned}
& \| r_n \, \mathbf{1}_{L_n(c)^{T_n}} - r \, \mathbf{1}_{L(c)} \|_p \\
= & \| r_n \, \mathbf{1}_{L_n(c)^{T_n}} - r \, \mathbf{1}_{L(c)^{T_n}} + r \, \mathbf{1}_{L(c)^{T_n}} - r \, \mathbf{1}_{L(c)} \|_p \\
\leq & \| r_n \, \mathbf{1}_{L_n(c)^{T_n}} - r \, \mathbf{1}_{L(c)^{T_n}} \|_p + \| r \, \mathbf{1}_{L(c)^{T_n}} - r \, \mathbf{1}_{L(c)} \|_p \\
\leq & \| r_n - r \|_p + K \, \left(\| f \|_{1+\epsilon, \lambda} \, \left(\lambda \left(L(c)^{T_n} \triangle L_n(c)^{T_n} \right) \right)^{\frac{\epsilon}{1+\epsilon}} \right)^{1/p} \\
+ & \left(\int_{L(c) \setminus [0, T_n]^d} |r(x)|^p f(x) dx \right)^{1/p} .
\end{aligned}
$$

From Theorem 1 and assumptions of Theorem 2, the last inequality concludes the proof. □

Proof of Theorem 3: Note that the proofs of Theorem 3 and Theorem 1 are strongly related. We have

$$
\begin{aligned}
& w_n \ \| r_n \, \mathbf{1}_{L_n(c)^{T_n}} - r \, \mathbf{1}_{L(c)^{T_n}} \|_p \\
\leq \ & w_n \| r_n \, \mathbf{1}_{L_n(c)^{T_n}} - r \, \mathbf{1}_{L_n(c)^{T_n}} \|_p + w_n \, \| r \, \mathbf{1}_{L_n(c)^{T_n}} - r \, \mathbf{1}_{L(c)^{T_n}} \|_p \\
\leq \ & w_n \, \| r_n - r \|_p \, + \, w_n \, K \, \left(\| f \|_{1+\epsilon, \lambda} \, \left(\lambda \left(L(c)^{T_n} \triangle L_n(c)^{T_n} \right) \right)^{\frac{\epsilon}{1+\epsilon}} \right)^{1/p}
\end{aligned}
$$

and Theorem 1 concludes the proof. □

Proof of Theorem 4: The proof is a straightforward application of Lemma 1 and Lemma 2. □

Lemma 1 *Under assumptions of Theorem 4, we have*

$$
p_n^{\frac{\epsilon}{2(1+\epsilon)}} | \, \mathbb{E} \left[Y \, | \, \mathbf{X} \in L(c)^{T_n} \right] - \mathbb{E} \left[Y \, | \, \mathbf{X} \in L_n(c)^{T_n} \right] | \xrightarrow[n \to \infty]{\mathbb{P}} 0,
$$

with $\epsilon > 0$ such that f is $1 + \epsilon$ integrable.

Proof of Lemma 1: Using Theorem 1, we obtain

$$
p_n^{\frac{\epsilon}{2(1+\epsilon)}} | P \left[\mathbf{X} \in L(c)^{T_n} \triangle L_n(c)^{T_n} \right] | \leq p_n^{\frac{\epsilon}{2(1+\epsilon)}} d_\lambda \left(L(c)^{T_n}, L_n(c)^{T_n} \right)^{\frac{\epsilon}{1+\epsilon}} \| f \|_{1+\epsilon} \xrightarrow[n \to \infty]{\mathbb{P}} 0. \tag{10.3}
$$

Then we get

$$
p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| P \left[\mathbf{X} \in L(c)^{T_n} \right] - P \left[\mathbf{X} \in L_n(c)^{T_n} \right] \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0.
$$

Denote g the density of the pair $(\mathbf{X}, Y)$. Under assumptions, and since Y is bounded, we also obtain

$$\begin{aligned}
& p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \int y \mathbf{1}_{L(c)^{T_n}} g(x,y)\lambda(dxdy) - \int y \mathbf{1}_{L_n(c)^{T_n}} g(x,y)\lambda(dxdy) \right| \\
& \leq p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \int y \mathbf{1}_{L(c)^{T_n} \triangle L_n(c)^{T_n}} g(x,y)\lambda(dxdy) \right| \\
& \leq p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left(\int y^2 g(x,y)\lambda(dxdy) \right)^{1/2} \left(\int g(x,y) \mathbf{1}_{L(c)^{T_n} \triangle L_n(c)^{T_n}} \lambda(dxdy) \right)^{1/2} \\
& \leq p_n^{\frac{\epsilon}{2(1+\epsilon)}} E\left[Y^2\right]^{\frac{1}{2}} \left(\int f(x) \mathbf{1}_{L(c)^{T_n} \triangle L_n(c)^{T_n}} \lambda(dx) \right)^{1/2} \\
& \leq p_n^{\frac{\epsilon}{2(1+\epsilon)}} E\left[Y^2\right]^{\frac{1}{2}} d_\lambda\left(L(c)^{T_n}, L_n(c)^{T_n}\right)^{\frac{\epsilon}{2(1+\epsilon)}} \|f\|_{1+\epsilon}^{1/2} \xrightarrow[n\to\infty]{\mathbb{P}} 0. \qquad (10.4)
\end{aligned}$$

Then

$$\begin{aligned}
& p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \mathbb{E}[Y|\mathbf{X} \in L(c)^{T_n}] - \mathbb{E}[Y|\mathbf{X} \in L_n(c)^{T_n}] \right| \\
& = p_n^{\frac{\epsilon}{2(1+\epsilon)}} \left| \int y\, \mathbf{1}_{L(c)^{T_n}} g(x,y)\lambda(\mathrm{d}x\mathrm{d}y)\ \mathbb{P}[\mathbf{X} \in L(c)^{T_n}]^{-1} \right. \\
& \quad \left. - \int y\, \mathbf{1}_{L_n(c)^{T_n}} g(x,y)\lambda(\mathrm{d}x\mathrm{d}y)\ \mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}]^{-1} \right| \\
& \leq \frac{p_n^{\frac{\epsilon}{2(1+\epsilon)}}}{\mathbb{P}[\mathbf{X} \in L(c)^{T_n}]\, \mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}]} \left(\mathbb{P}[\mathbf{X} \in L(c)^{T_n}] \left| \int y\, \mathbf{1}_{L(c)^{T_n}} g(x,y)\lambda(\mathrm{d}x\mathrm{d}y) \right. \right. \\
& \quad \left. - \int y\, \mathbf{1}_{L_n(c)^{T_n}} g(x,y)\lambda(\mathrm{d}x\mathrm{d}y) \right| \\
& \quad \left. + \int y\, \mathbf{1}_{L_n(c)^{T_n}} g(x,y)\lambda(\mathrm{d}x\mathrm{d}y) \cdot \left| \mathbb{P}[\mathbf{X} \in L(c)^{T_n}] - \mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}] \right| \right).
\end{aligned}$$

Using (10.3) and (10.4) we obtain the result. □

Lemma 2 *Under assumptions of Theorem 4 and with* $d_n = o\left(\sqrt{n}\right)$*, we have*

$$d_n \left| \mathbb{E}\left[Y|\mathbf{X} \in L_n(c)^{T_n}\right] - \mathbb{E}_n\left[Y|\mathbf{X} \in L_n(c)^{T_n}\right] \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Proof of Lemma 2: We have

$$\begin{aligned}
& d_n \left| \mathbb{E}\left[Y|\mathbf{X} \in L_n(c)^{T_n}\right] - \mathbb{E}_n\left[Y|\mathbf{X} \in L_n(c)^{T_n}\right] \right| \\
& = d_n \left| \frac{\int y \mathbf{1}_{L_n(c)^{T_n}} g(x,y)\lambda(dxdy)}{\mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}]} - \frac{\sum_{i=1}^n Y_i\ \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}}}{\sum_{i=1}^n \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}}} \right|.
\end{aligned}$$

Under assumptions of the Lemma 2 and using Theorem 27.2 in Billingsley (1995), we obtain that

$$d_n \left| \mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}] - \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}} \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

and

$$d_n \left| \int y \mathbf{1}_{L_n(c)^{T_n}} g(x,y)\lambda(dxdy) - \frac{1}{n}\sum_{i=1}^{n} Y_i \, \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}} \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Hence the result. □

Proof of Theorem 5: The proof is a straightforward application of Lemmas 1, 3 and 4. □

Lemma 3 *Under assumptions of Theorem 5, we have*

$$c_N \left| \mathbb{E}[Y \mid \mathbf{X} \in L_n(c)^{T_n}] - \mathbb{E}[r_N(\mathbf{X}) \mid \mathbf{X} \in L_n(c)^{T_n}] \right| \xrightarrow[N\to\infty]{\mathbb{P}} 0,$$

with $c_N = o(\mathbb{E}\,|r_N(\mathbf{X}) - r(\mathbf{X})|)$.

Proof of Lemma 3: We have

$$\begin{aligned}
\left| \mathbb{E}[Y \mid \mathbf{X} \in L_n(c)^{T_n}] \right. & - \left. \mathbb{E}[r_N(\mathbf{X}) \mid \mathbf{X} \in L_n(c)^{T_n}] \right| \\
&= \frac{\left| \int_{L_n(c)^{T_n}} (r(x) - r_N(x)) f(x) d\lambda(x) \right|}{\mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}]} \\
&\leq K \int_{L_n(c)^{T_n}} |r(x) - r_N(x)| \, f(x) d\lambda(x) \\
&\leq \int_{\mathbb{R}^d} |r(x) - r_N(x)| \, f(x) d\lambda(x) = \mathbb{E}\,|r_N(\mathbf{X}) - r(\mathbf{X})|.
\end{aligned}$$

Hence the result. □

Lemma 4 *Under assumptions of Theorem 5, we have*

$$d_n \left| \mathbb{E}\left[r_N(\mathbf{X}) | \mathbf{X} \in L_n(c)^{T_n} \right] - \mathbb{E}_n \left[r_N(\mathbf{X}) | \mathbf{X} \in L_n(c)^{T_n} \right] \right| \xrightarrow{\mathbb{P}}_{n\to\infty} 0$$

where $d_n = o\left(\sqrt{n}\right)$.

Proof of Lemma 4: We have

$$\begin{aligned}
& d_n \left| \mathbb{E}\left[r_N(\mathbf{X}) | L_n(c)^{T_n} \right] - \mathbb{E}_n \left[r_N(\mathbf{X}) | L_n(c)^{T_n} \right] \right| \\
&= d_n \left| \frac{\int_{L_n(c)^{T_n}} r_N(x) f(x) \lambda(dx)}{\mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}]} - \frac{\sum_{i=1}^n r_N(X_i) \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}}}{\sum_{i=1}^n \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}}} \right|.
\end{aligned}$$

Under the assumptions of the lemma and using Theorem 27.2 in Billingsley (1995), we obtain that

$$d_n \left| \mathbb{P}[\mathbf{X} \in L_n(c)^{T_n}] - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}} \right| \xrightarrow{\mathbb{P}}_{n\to\infty} 0,$$

and

$$d_n \left| \int_{L_n(c)^{T_n}} r_N(x) f(x) \lambda(dx) - \frac{1}{n} \sum_{i=1}^{n} r_N(\mathbf{X}_i) \mathbf{1}_{\{\mathbf{X}_i \in L_n(c)^{T_n}\}} \right| \overset{\mathbb{P}}{\to}_{n \to \infty} 0$$

which gives us the Lemma. □

Proof of Theorem 6: We have

$$F_{a\mathbf{X}}(x) = F_{\mathbf{X}}\left(\frac{x}{a}\right), \; f_{a\mathbf{X}}(x) = \frac{1}{a} f_{\mathbf{X}}\left(\frac{x}{a}\right) \text{ and } r_{a\mathbf{X}, bY}(x) = b\, r_{\mathbf{X},Y}\left(\frac{x}{a}\right).$$

Proof of 1. in Theorem 6: We have

$$\begin{aligned}
&\| r_{n,a\mathbf{X},bY}\, \mathbf{1}_{L_{a,n}(c)^{T_n}} - r_{a\mathbf{X},bY}\, \mathbf{1}_{L_a(c)^{T_n}} \|_p \\
&= \left(\int_{\mathbb{R}^d} \left| r_{n,a\mathbf{X},bY}(x) \mathbf{1}_{L_{a,n}(c)^{T_n}} - r_{a\mathbf{X},bY}(x) \mathbf{1}_{L_a(c)^{T_n}} \right|^p f_{a\mathbf{X}}(x) dx \right)^{1/p} \\
&= \left(\int_{\mathbb{R}^d} b^p \left| r_{n,\mathbf{X},Y}\left(\frac{x}{a}\right) \mathbf{1}_{\{x/a \in L_n(c)^{T_n}\}} - r_{\mathbf{X},Y}\left(\frac{x}{a}\right) \mathbf{1}_{\{x/a \in L(c)^{T_n}\}} \right|^p \frac{1}{a} f_{\mathbf{X}}\left(\frac{x}{a}\right) dx \right)^{1/p}
\end{aligned}$$

and taking $t = x/a$ we obtain

$$\begin{aligned}
&\| r_{n,a\mathbf{X},bY}\, \mathbf{1}_{L_{a,n}(c)^{T_n}} - r_{a\mathbf{X},bY}\, \mathbf{1}_{L_a(c)^{T_n}} \|_p \\
&= \left(\int_{\mathbb{R}^d} b^p \left| r_{n,\mathbf{X},Y}(t)\, \mathbf{1}_{L_n(c)^{T_n}} - r_{\mathbf{X},Y}(t)\, \mathbf{1}_{L(c)^{T_n}} \right|^p f_{\mathbf{X}}(t)\, dt \right)^{1/p} \\
&= b \| r_{n,\mathbf{X},Y}\, \mathbf{1}_{L_n(c)^{T_n}} - r_{\mathbf{X},Y}\, \mathbf{1}_{L(c)^{T_n}} \|_p
\end{aligned}$$

Hence the result. □

Proof of 2. in Theorem 6: We have

$$\begin{aligned}
&\left| \widehat{\mathrm{CCTE}}_{c,n}^{T_n}(a\mathbf{X}, bY) \; - \; \mathrm{CCTE}_c^{T_n}(a\mathbf{X}, bY) \right| \\
&\qquad = \left| \mathbb{E}_n \left[bY | a\mathbf{X} \in L_{n,a}(c)^{T_n} \right] - \mathbb{E}\left[bY | a\mathbf{X} \in L_a(c)^{T_n} \right] \right|.
\end{aligned}$$

Using $L_a(c)^{T_n} = aL(c)^{T_n}$ and the assumptions, we obtain

$$\begin{aligned}
&\left| \widehat{\mathrm{CCTE}}_{c,n}^{T_n}(a\mathbf{X}, bY) \; - \; \mathrm{CCTE}_c^{T_n}(a\mathbf{X}, bY) \right| \\
&\qquad = \frac{1}{b} \left| \mathbb{E}_n \left[Y | a\mathbf{X} \in aL_n(c)^{T_n} \right] - \mathbb{E}\left[Y | a\mathbf{X} \in aL(c)^{T_n} \right] \right|.
\end{aligned}$$

Hence the result. □

Acknowledgements The authors thank the two anonymous reviewers and an associated editor for their useful comments and suggestions. This work has been partially supported by the French research national agency (ANR) under the references ANR-08BLAN-0314-01 and ANR 2011 BS01 010 01 projet Calibration. The authors thank Yannick Baraud, Roland Diel, Christine Tuleau-Malo and Patricia Reynaud-Bouret for fruitful discussions.

References

Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. Math Finance 9(3):203–228, DOI 10.1111/1467-9965.00068

Baíllo A, Cuesta-Albertos J, Cuevas A (2001) Convergence rates in nonparametric estimation of level sets. Statistics and Probability Letters 53:27–35

Belzunce F, Castaño A, Olvera-Cervantes A, Suárez-Llorens A (2007) Quantile curves and dependence structure for bivariate distributions. Computational Statistics & Data Analysis 51(10):5112–5129, DOI 10.1016/j.csda.2006.08.017

Bentahar I (2006) Tail conditional expectation for vector-valued risks. Sfb 649 discussion papers, Humboldt University, Collaborative Research Center 649, URL http://econpapers.repec.org/RePEc:hum:wpaper:sfb649dp2006-029

Billingsley P (1995) Probability and measure, 3rd edn. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, a Wiley-Interscience Publication

Cai J, Li H (2005) Conditional tail expectations for multivariate phase-type distributions. Journal of Applied Probability 42(3):810–825

Cavalier L (1997) Nonparametric estimation of regression level sets. Statistics A Journal of Theoretical and Applied Statistics 29(2):131–160, DOI 10.1080/02331889708802579

Chebana F, Ouarda T (2011) Multivariate quantiles in hydrological frequency analysis. Environmetrics 22(1):63–78, DOI 10.1002/env.1027, URL http://dx.doi.org/10.1002/env.1027

Cousin A, Di Bernardino E (2013) On multivariate extensions of Value-at-Risk. Journal of Mulitivariate Analysis 119:32–46

Cuevas A, González-Manteiga W, Rodríguez-Casal A (2006) Plug-in estimation of general level sets. Australian & New Zealand Journal of Statistics 48(1):7–19, DOI 10.1111/j.1467-842X.2006.00421.x

Daouia A, Gardes L, Girard S, Lekina A (2010) Kernel estimators of extreme level curves. Test (20):311–333

Dedu S, Ciumara R (2010) Restricted optimal retention in stop-loss reinsurance under VaR and CTE risk measures. Proc Rom Acad Ser A Math Phys Tech Sci Inf Sci 11(3):213–217

Denuit M, Dhaene J, Goovaerts M, Kaas R (2005) Actuarial Theory for Dependent Risks. Wiley

Di Bernardino E, Laloë T, Maume-Deschamps V, 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

Draisma G, Drees H, Ferreira A, de Haan L (2004) Bivariate tail estimation: dependence in asymptotic independence. Bernoulli 10(2):251–280, DOI 10.3150/bj/1082380219, URL http://dx.doi.org/10.3150/bj/1082380219

Ekeland I, Galichon A, Henry M (2012) Comonotonic measures of multivariate risks. Tech. rep.

Embrechts P, Puccetti G (2006) Bounds for functions of multivariate risks. Journal of Multivariate Analysis 97(2):526–547, DOI 10.1016/j.jmva.2005.04.001, URL http://dx.doi.org/10.1016/j.jmva.2005.04.001

Fan J, Yao Q (2003) Nonlinear time series. Springer series in statistics, Springer, New York, NY

de Haan L, Huang X (1995) Large quantile estimation in a multivariate setting. J Multivariate Anal 53(2):247–263, DOI 10.1006/jmva.1995.1035

Hawkes J, Gouldby BP, Tawn JA, Owen MW (2002) The joint probability ofwaves andwater levels in coastal engineering design. Journal of Hydraulic Research 40(3):241–251

Imlahi L, Ezzerg M, Chakak A (1999) Estimación de la curva mediana de una cópula $C(x_1, \dots, x_n)$. Rev R Acad Cien Exact Fis Nat 93(2):241–250

Jouini E, Meddeb M, Touzi N (2004) Vector-valued coherent risk measures. Finance Stoch 8(4):531–552, DOI 10.1007/s00780-004-0127-6, URL http://dx.doi.org/10.1007/s00780-004-0127-6

Kohler M, Krzyzak A, Walk H (2009) Optimal global rates of convergence for nonparametric regression with unbounded data. J Statist Plann Inference 139(4):1286–1296, DOI 10.

1016/j.jspi.2008.07.012, URL http://dx.doi.org/10.1016/j.jspi.2008.07.012

Laloë T, Servien R (2013) Nonparametric estimation of regression level sets using kernel plug-in estimator. JKSS 42(3):301–311, DOI 10.1016/j.jkss.2012.10.001, URL http://dx.doi.org/10.1016/j.jkss.2012.10.001

Mason M, Polonik W (2009) Asymptotic normality of plug-in level set estimates. Annals of Applied Probability 19:1108–1142

Nappo G, Spizzichino F (2009) Kendall distributions and level sets in bivariate exchangeable survival models. Information Sciences 179:2878–2890, DOI 10.1016/j.ins.2009.02.007, URL http://portal.acm.org/citation.cfm?id=1558390.1558504

Nelsen RB (1999) An introduction to copulas, Lecture Notes in Statistics, vol 139. Springer-Verlag, New York

Rigollet P, Vert R (2009) Optimal rates for plug-in estimators of density level sets. Bernoulli Official Journal of the Bernoulli Society for Mathematical Statistics and Probability 15(4):1154–1178, DOI 10.3150/09-BEJ184, URL http://dx.doi.org/10.3150/09-BEJ184

Salvadori G, De Michele C, Durante F (2011) On the return period and design in a multivariate framework. Hydrology and Earth System Sciences 15(11):3293–3305, DOI 10.5194/hess-15-3293-2011, URL http://www.hydrol-earth-syst-sci.net/15/3293/2011/

Tau V, Dam GC (2011) Preliminary design study. Project group Flood Defence, Institute for Water and Environment Research pp 1–144

Tibiletti L (1993) On a new notion of multidimensional quantile. Metron International Journal of Statistics 51(3-4):77–83