ESAIM: PS 17 (2013) 236–256
DOI: 10.1051/ps/2011161

ESAIM: Probability and Statistics
www.esaim-ps.org

PLUG-IN ESTIMATION OF LEVEL SETS IN A NON-COMPACT SETTING WITH APPLICATIONS IN MULTIVARIATE RISK THEORY

Elena Di Bernardino[1], Thomas Laloë[2], Véronique Maume-Deschamps[1] and Clémentine Prieur[3]

Abstract. This paper deals with the problem of estimating the level sets $L(c) = \{F(x) \geq c\}$, with $c \in (0,1)$, of an unknown distribution function F on $\mathbb{R}^2_+$. A plug-in approach is followed. That is, given a consistent estimator F_n of F, we estimate $L(c)$ by $L_n(c) = \{F_n(x) \geq c\}$. In our setting, non-compactness property is *a priori* required for the level sets to estimate. We state consistency results with respect to the Hausdorff distance and the volume of the symmetric difference. Our results are motivated by applications in multivariate risk theory. In particular we propose a new bivariate version of the conditional tail expectation by conditioning the two-dimensional random vector to be in the level set $L(c)$. We also present simulated and real examples which illustrate our theoretical results.

Mathematics Subject Classification. 62G05, 62G20, 60E05, 91B30.

Received March 22, 2011. Revised July 13, 2011.

1. Introduction

The conditional tail expectation (CTE) is a commonly used tool in the univariate risk theory. Given a random variable X with continuous and strictly monotonic distribution function F_X, the CTE at level α is defined by

$$\mathrm{CTE}_\alpha(X) = \mathbb{E}[\, X \,|\, X \geq Q_X(\alpha)\,], \tag{1.1}$$

with $\alpha \in (0,1)$ and $Q_X(\alpha)$ the univariate quantile function of X at level α. In other words the $\mathrm{CTE}_\alpha(X)$ is defined by the expectation value of X conditionally to the fact that X belongs to a certain one-dimensional level set of F_X. For further details about this univariate risk measure we refer the reader *e.g.* to Denuit *et al.* [16].

In this paper we are interested to introduce a new bivariate generalization of CTE in (1.1) and to provide a consistent procedure to estimate it. Consequently we essentially need a generalization of the notion of quantile in dimension higher than one. Therefore we remark that the quantile function is a relevant tool in statistics and

Keywords and phrases. Level sets, distribution function, plug-in estimation, Hausdorff distance, conditional tail expectation.

[1] Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 50 Avenue Tony Garnier, 69366 Lyon, France. elena.di-bernardino@univ-lyon1.fr; veronique.maume@univ-lyon1.fr

[2] Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice Cedex 02, France. thomas.laloe@unice.fr

[3] Université Joseph Fourier, Tour IRMA, MOISE-LJK B.P. 53 38041 Grenoble, France. clementine.prieur@imag.fr

Article published by EDP Sciences

© EDP Sciences, SMAI 2013

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

probability, but its practical use is mostly confined to the univariate case where it is easily interpretable and calculable. The bivariate (or multivariate) characterization has received some attention however, it is much less common, due to a number of theoretical and practical reasons. In particular the main difficulty is that there does not exist a natural ordering in n-dimensions, for $n > 1$. For this reason, in the last decade several attempts have been made to provide a multidimensional generalization of the univariate quantile function. We refer to Serfling [31] for a complete review on the topic. For example Massé and Theodorescu [23] defined multivariate quantiles as halfplanes and Koltchinskii [21] provided a general treatment of multivariate quantiles as inversions of mappings.

As a possible generalization of the univariate quantile function, Tibiletti [32], Fernández–Ponce and Suárez–Lloréns [18] and Belzunce *et al.* [4] defined a multivariate quantile as a set of points which accumulate the same probability for a fixed orthant. They called it level curves or quantile curves. In this sense de Haan and Huang [14] model a risk-problem of flood in the bivariate setting using an estimator of level curves of bivariate distribution function. In this paper we will follow this approach in order to propose a new risk measure which generalizes the CTE in (1.1). To this end we need to estimate the level sets of a bivariate distribution function F. This leads us into the general field of level sets estimation.

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 recently. In particular the estimation of density level sets has been studied in Polonik [25], Tsybakov [33], Baíllo *et al.* [3], Baíllo [2], Cadre [7], Rigollet and Vert [27], Biau *et al.* [5]. The estimation of regression level sets in a compact setting has been analyzed in Cavalier [9], Laloë [22]. An alternative approach, based on the geometric properties of the compact support sets, has been presented by Hartigan [20], Cuevas and Fraiman [11], Cuevas and Rodríguez–Casal [12]. The problem of estimating general level sets under compactness assumptions has been discussed by Cuevas *et al.* [13]. The asymptotic behavior of minimum volume sets and of a generalized quantile process is analyzed by Polonik [26].

Considering the level sets of a bivariate distribution function, the commonly assumed property of compactness for these sets is no more reasonable. Then, differing from the literature cited above, we need to work in a non-compact setting and this requires special attention in the statement of our problem.

Our general approach will be the following: first, we provide a consistent estimator of the level set

$$L(c) = \{F(x) \geq c\}, \quad \text{for } c \in (0,1).$$

Therefore we consider a *plug-in* approach (*e.g.* see Baíllo *et al.* [3]; Rigollet and Vert [27]; Cuevas *et al.* [13]), that is, $L(c)$ is estimated by

$$L_n(c) = \{F_n(x) \geq c\}, \quad \text{for } c \in (0,1),$$

where F_n is a consistent estimator of F. 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.

In order to provide a consistency results we define two proximity criteria between sets. A standard choice is the volume of the symmetric difference. Another natural criterion is given by the Hausdorff distance that corresponds to an intuitive notion of "physical proximity" between sets. Our consistency properties are stated with respect to these two criteria under reasonable assumptions on F and F_n (Thms. 3.1–4.1). These results are based on a slight modification of Proposition 3.1 in the Ph.D. thesis of Rodríguez–Casal [28] (Prop. 2.1).

We then introduce our new bivariate version of the Conditional Tail Expectation by conditioning the two-dimensional random vector to be in the level set $L(c)$. This new risk measure is based on the bivariate Value-at-Risk proposed in Embrechts and Puccetti [17] and Nappo and Spizzichino [24] and it is different from the existing bivariate generalizations of the univariate conditional tail expectation (*e.g.* see Cai and Li [8]). Indeed we do not use any aggregate risk in the portfolio in order to analyze the bivariate risk's issue. Our bivariate version of the CTE deals with the simultaneous joint damages considering the dependence structure of data in a bivariate specific risk's area ($L(c)$). We propose an estimator for this new risk measure using plug-in approach for level sets and provide a consistency result (Thm. 5.3).

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

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 Sections 3 and 4. In Section 5 we introduce a new bivariate generalization of CTE in (1.1) and we provide a consistent estimator for this risk measure. Illustrations with simulated and real data are presented in Section 6. Section 7 summarizes and briefly mentions directions for future research. Finally, some auxiliary results and more technical proofs are postponed to Section 8.

2. Notation and preliminaries

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

Let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{R}_+^{2\,*} = \mathbb{R}_+^2 \setminus (0,0)$, $\mathcal{F}$ the set of continuous distribution functions $\mathbb{R}_+^2 \to [0,1]$ and $F \in \mathcal{F}$. Given an *i.i.d* sample $\{X_i\}_{i=1}^n$ in $\mathbb{R}_+^2$ with distribution function F, we denote by $F_n(\cdot) = F_n(X_1, X_2, \ldots, X_n, \cdot)$ an estimator of F based on the finite sample $(X_1, X_2, \ldots, X_n)$. We restrict ourselves to $\mathbb{R}_+^2$ for convenience but the following results are completely adaptable in $\mathbb{R}^2$. This choice is motivated essentially by our application in risk theory proposed in Section 5, where random variables will be losses then defined in a positive support.

Define, for $c \in (0,1)$, the *upper c-level set* of $F \in \mathcal{F}$ and its plug-in estimator

$$L(c) = \{x \in \mathbb{R}_+^2 : F(x) \geq c\}, \quad L_n(c) = \{x \in \mathbb{R}_+^2 : F_n(x) \geq c\},$$

and

$$\{F = c\} = \{x \in \mathbb{R}_+^2 : F(x) = c\}.$$

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

$$L(c)^T = \{x \in [0,T]^2 : F(x) \geq c\}, \;\; L_n(c)^T = \{x \in [0,T]^2 : F_n(x) \geq c\},$$

$$\{F = c\}^T = \{x \in [0,T]^2 : F(x) = c\}.$$

Furthermore, for any $A \subset \mathbb{R}_+^2$ we denote by ∂A its boundary.

Note that $\{F = c\}$ can be a portion of quadrant $\mathbb{R}_+^2$ instead of a set of Lebesgue measure null in $\mathbb{R}_+^2$ (that is the presence of plateau at level c). In the following we will introduce suitable conditions in order to avoid the presence of plateau in the graph of F.

In the metric space $(\mathbb{R}_+^2, d)$, where d stands for the Euclidean distance, we denote by $B(x,\rho)$ the closed ball centered on x and with positive radius ρ. Let $B(S,\rho) = \bigcup_{x \in S} B(x,\rho)$, with S a closed set of $\mathbb{R}_+^2$. For $r > 0$ and $\zeta > 0$, define

$$E = B(\{x \in \mathbb{R}_+^2 : \mid F(x) - c \mid \leq r\}, \zeta),$$

and, for a twice differentiable function F,

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

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 $\|(HF)_x\|$ its matrix norm induced by the Euclidean norm.

For sake of completeness, we recall that if A_1 and A_2 are compact sets in $(\mathbb{R}_+^2, d)$, the Hausdorff distance between A_1 and A_2 is defined by

$$d_H(A_1, A_2) = \inf\{\rho > 0 : A_1 \subset B(A_2, \rho), A_2 \subset B(A_1, \rho)\},$$

or equivalently by

$$d_H(A_1, A_2) = \max\left\{ \sup_{x \in A_1} d(x, A_2), \sup_{x \in A_2} d(x, A_1) \right\},$$

where $d(x, A_2) = \inf_{y \in A_2} \| x - y \|$.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

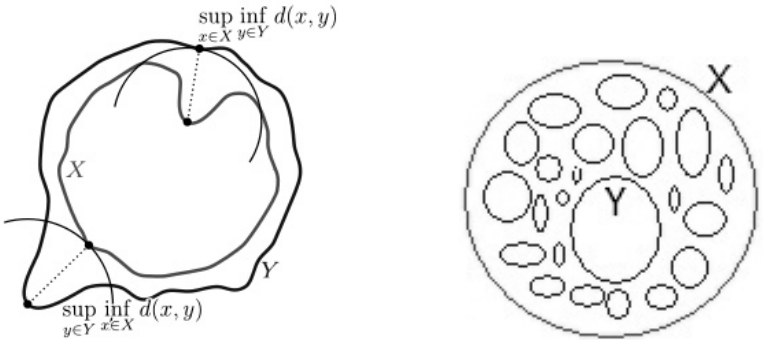


FIGURE 1. (left) Hausdorff distance between sets; (right) particular case where $d_H(X,Y)$ does not capture the shape properties of X and Y.

The above expression is well-defined even when A_1 and A_2 are just closed (not necessarily compact) sets but in this case the value $d_H(A_1, A_2)$ could be infinity. Then in our setting, in order to avoid these situations we need a truncated version of Assumption T in Cuevas *et al.* [13] or Tsybakov [33]. More precisely we introduce the following assumption.

H: There exist $\gamma > 0$ and $A > 0$ such that, if $|t - c| \leq \gamma$ then $\forall\ T > 0$ such that $\{F = c\}^T \neq \emptyset$ and $\{F = t\}^T \neq \emptyset$,

$$d_H(\{F = c\}^T, \{F = t\}^T) \leq A\,|t - c|\,.$$

Assumption **H** is satisfied under mild conditions. Proposition 2.1 below is a slight modification of Proposition 3.1 in the Ph.D. thesis of Rodríguez–Casal [28] in order to deal with non-compact sets.

Proposition 2.1. *Let $c \in (0,1)$. Let $F \in \mathcal{F}$ be twice differentiable on $\mathbb{R}^{2*}_+$. Assume there exist $r > 0$, $\zeta > 0$ such that $m^\nabla > 0$ and $M_H < \infty$. Then F satisfies Assumption* **H**, *with $A = \frac{2}{m^\nabla}$.*

The proof is postponed to Section 8.

Remark 2.2. Under assumptions of Proposition 2.1, F is continuous and $m^\nabla > 0$, there is no plateau in the graph of F for each level t such that $|t - c| \leq r$. Furthermore from Theorem 1 in Rossi [30] we know that each half-line in $\mathbb{R}^2_+$, parallel to one of the axis, meets $\{F = t\}$ in a connected set of points. As a consequence we obtain that $\{F = t\}$ is a curve in the quadrant $\mathbb{R}^2_+$. In particular, from $m^\nabla > 0$, for a fixed x we have to consider all corresponding values y in a specific interval (we refer the interested reader to Rem. 2.1 in Rossi [30] or Rem. 3.1 in Rossi [29]). In this case the plane curve $\{F = t\}$ is not the graph of a function but it has the following monotonic property. We consider (x, y), $(x', y') \in \{F = t\}$, if $x < x'$ then $y \geq y'$, if $y < y'$ then $x \geq x'$. In particular if we suppose that each component of $(\nabla F)_x$ is greater than zero in E then $\{F = t\}$ is a monotone decreasing curve in $\mathbb{R}^2_+$. Finally under assumptions of Proposition 2.1 we obtain $\partial L(c)^T = \{F = c\}^T = \{F = c\} \cap [0,T]^2$ (we refer for details to Thm. 3.2 in Rodríguez–Casal [28]).

3. Consistency in terms of the Hausdorff distance

In this section we study the consistency properties of $L_n(c)^T$ with respect to the Hausdorff distance between $\partial L_n(c)^T$ and $\partial L(c)^T$. The Hausdorff distance corresponds to an intuitive notion of "physical proximity" between sets (see Fig. 1 (left) for a graphical illustration). However the metric d_H is not always completely successful in capturing the shape properties: two sets can be very close in d_H and still have quite different shapes. For instance in Figure 1 (right) we have X, the outside ball, and Y the set of balls inside X. The distance $d_H(X,Y)$ is small even if X and Y are quite different in many important aspects.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

In order to avoid these situations, following Cuevas and Rodríguez–Casal [12] and Cuevas *et al.* [13], a way to reinforce the notion of visual proximity between two sets provided by d_H is to impose the proximity of the respective boundaries. We will follow this approach in Theorem 3.1 below.

From now on we note, for $n \in \mathbb{N}^*$,

$$\|F - F_n\|_\infty = \sup_{x \in \mathbb{R}^2_+} \mid F(x) - F_n(x) \mid,$$

and for $T > 0$

$$\|F - F_n\|_\infty^T = \sup_{x \in [0,T]^2} \mid F(x) - F_n(x) \mid .$$

The following result can be considered an adapted version of Theorem 2 in Cuevas *et al.* [13].

Theorem 3.1. *Let $c \in (0,1)$. Let $F \in \mathcal{F}$ be twice differentiable on $\mathbb{R}^{2*}_+$. Assume that there exist $r > 0$, $\zeta > 0$ such that $m^\triangledown > 0$ and $M_H < \infty$. Let $T_1 > 0$ such that for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Let $(T_n)_{n\in\mathbb{N}^*}$ be an increasing sequence of positive values. Assume that, for each n and for almost all samples of size n, F_n is a continuous function and that*

$$\|F - F_n\|_\infty \to 0, \quad a.s. \tag{3.1}$$

Then

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) = O(\|F - F_n\|_\infty), \qquad a.s.$$

The proof is postponed to Section 8. Under assumptions of Theorem 3.1, $d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n})$ converges to zero and the quality of our plug-in estimator is obviously related to the quality of the estimator F_n.

Remark 3.2. Theorem 3.1 provides an asymptotic result for a fixed level c. In particular following the proof of Theorem 3.1 (postponed to Sect. 8) we remark that, for n large enough,

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) \leq 6\, A\, \|F - F_n\|_\infty^{T_n}, \quad \text{a.s.},$$

where $A = \frac{2}{m^\triangledown}$. Note that in the case c is close to one the constant A could be large. In this case, we will need a large number of data to get a reasonable estimation.

Note that the empirical distribution estimator does not satisfy continuity assumption imposed in Theorem 3.1. However, in order to overcome this problem it can be considered a smooth version of this estimator (*e.g.* see Chaubey and Sen [10]), which satisfies convergence in (3.1). However we remark that the continuity assumption, imposed in Theorem 3.1, will not be necessary in the following results.

4. L_1 CONSISTENCY

The previous section was devoted to the consistency of L_n in terms of the Hausdorff distance. We consider now another consistency criterion: the consistency 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 $\mathbb{R}^+_2$ by

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

where λ stands for the Lebesgue measure on $\mathbb{R}^2$ 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

$$v_n \int_{[0,T_n]^2} \mid F - F_n \mid^p \ \lambda(\mathrm{d}x) \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

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

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

We now establish our consistency result in terms of the volume of the symmetric difference. We can interpret the following theorem as a generalization of Theorem 3 in Cuevas *et al.* [13] (Sect. 3) in the case of non-compact level sets. In addition, we provide also a convergence rate for the symmetric difference between $L(c)^{T_n}$ and $L_n(c)^{T_n}$.

Theorem 4.1. *Let $c \in (0,1)$. Let $F \in \mathcal{F}$ be twice differentiable on $\mathbb{R}_+^{2*}$. Assume that there exist $r > 0$, $\zeta > 0$ such that $m^{\nabla} > 0$ and $M_H < \infty$. Assume that for each n, with probability one, 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 and that for all $t : \mid t-c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Then it holds that*

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

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

The proof is postponed to Section 8. Theorem 4.1 provides a convergence rate, which is closely related to the choice of the sequence T_n. A sequence T_n whose divergence rate is large implies a convergence rate p_n quite slow. Note that, as in Theorem 3 in Cuevas *et al.* [13], Theorem 4.1 above does not require any continuity assumption on F_n.

Remark 4.2. Assumptions on gradient vector and Hessian matrix of F, in Theorems 3.1 and 4.1, are satisfied for a quite large class of classical bivariate distribution functions; for instance independent copula and exponential marginals, Farlie–Gumbel–Morgenstern (FGM), Clayton or Survival Clayton copulas and Burr marginals.

We now suppose that there exists a positive increasing sequences $(v_n)_{n\in\mathbb{N}^*}$ such that $v_n \, \|F - F_n\|_\infty \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Then

$$\forall \; p \geq 1, \quad w_n \int_{[0,T_n]^2} \mid F - F_n \mid^p \; \lambda(\mathrm{d}x) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \qquad \text{with} \qquad w_n = \frac{v_n^p}{T_n^2}.$$

In this case, from Theorem 4.1 we get

$$p_n = o\left(w_n^{\frac{1}{p+1}} / T_n^{\frac{p}{p+1}}\right) = o\left(v_n^{\frac{p}{p+1}} / T_n^{\frac{p+2}{p+1}}\right). \tag{4.1}$$

From these considerations we can introduce the following remark.

Remark 4.3. Let F_n the bivariate empirical distribution function. Then it holds that $v_n \, \|F - F_n\|_\infty \xrightarrow[n\to\infty]{\mathbb{P}} 0$, with $v_n = o(\sqrt{n})$. From Theorem 4.1 and using the convergence rate in (4.1) with $p = 2$, we obtain $p_n = o(n^{\frac{1}{3}} / T_n^{\frac{4}{3}})$. For instance, for $F(x,y) = (1 - \mathrm{e}^{-x})(1 - \mathrm{e}^{-2\,y})$ and using the bivariate empirical distribution function F_n we obtain $p_n = o(n^{\frac{1}{3}} / \ln(n)^{\frac{4}{3}})$, with $p = 2$ and $T_n = \ln(n)$.

5. Application in bivariate risk theory

From the usual definition in the univariate setting we know that the quantile function Q_X provides a point which accumulates a probability α to the left tail and $1-\alpha$ to the right tail. More precisely, given an univariate continuous and strictly monotonic loss distribution function F_X,

$$Q_X(\alpha) = F_X^{-1}(\alpha), \quad \forall\, \alpha \in (0,1). \tag{5.1}$$

The notion of univariate quantile function Q_X is used in risk theory to define an univariate measure of risk: the Value-at-Risk (VaR). This measure is defined as

$$\mathrm{VaR}_\alpha(X) = Q_X(\alpha), \quad \forall\, \alpha \in (0,1). \tag{5.2}$$

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

As presented in the introduction of this paper, following the general ideas of Embrechts and Puccetti [17] and Nappo and Spizzichino [24] an intuitive generalization of the VaR measure in the case of a bidimensional loss distribution function F is given by its α-quantile curves. More precisely:

Definition 5.1. For $\alpha \in (0,1)$ and $F \in \mathcal{F}$, the bidimensional Value-at-Risk at probability level α is the boundary of its α-level set, *i.e.* $\mathrm{VaR}_\alpha(F) = \partial L(\alpha)$.

As pointed out in Tibiletti [32], imposing $\alpha = \frac{1}{2}$ in Definition 5.1, we get a natural extension of the bidimensional median. For details about a parametric formulation of the quantile curve $\partial L(\alpha)$ see Belzunce *et al.* [4]. For details about its properties see for instance Fernández–Ponce and Suárez–Lloréns [18] (and references therein) and Nappo and Spizzichino [24].

Then, using a bivariate estimator F_n as in Sections 3 and 4, we can define our plug-in estimator of the bivariate Value-at-Risk by

$$\mathrm{VaR}_\alpha(F_n) = \partial L_n(\alpha), \quad \text{for } \alpha \in (0,1).$$

Moreover, under assumptions of Theorem 3.1, we obtain a consistency result, with respect to the Hausdorff distance, for the $\mathrm{VaR}_\alpha(F_n)$ on the quadrant $\mathbb{R}_+^2$ *i.e.*

$$d_H(\mathrm{VaR}_\alpha(F)^{T_n}, \mathrm{VaR}_\alpha(F_n)^{T_n}) = O(\|F - F_n\|_\infty), \quad \text{a.s.}$$

As in the univariate case, the two-dimensional VaR at a predetermined level α does not give any information about thickness of the upper tail of the distribution function. This is a considerable shortcoming of VaR measure because in practice we are not only concerned with frequency of the default but also with the severity of loss in case of default. In other words we are interested to analyze the behavior of X and Y not only on the boundary but also in the whole α-level set.

In dimension one, in order to overcome this problem, another risk measure has recently received growing attention in insurance and finance literature: conditional tail expectation (CTE). Following Artzner *et al.* [1] and Dedu and Ciumara [15], for a continuous loss distribution function F_X the CTE at level α is defined by

$$\mathrm{CTE}_\alpha(X) = \mathbb{E}[\, X \,|\, X \geq \mathrm{VaR}_\alpha(X)\,],$$

where $\mathrm{VaR}_\alpha(X)$ is the univariate VaR in (5.2). For a comprehensive treatment and for references to the extensive literature on $\mathrm{VaR}_\alpha(X)$ and $\mathrm{CTE}_\alpha(X)$ one may refer to Denuit *et al.* [16].

Several bivariate generalizations of the classical univariate CTE have been proposed; mainly using as conditioning events the total risk or some univariate aggregate extreme risk in the portfolio (for instance the sum, the max or the min between two risks):

$$\mathbb{E}[\,(X,Y) \,|\, X+Y > Q_{X+Y}(\alpha)\,], \quad \mathbb{E}[\,(X,Y) \,|\, \min\{X,Y\} > Q_{\min\{X,Y\}}(\alpha)] \quad \text{and}$$
$$\mathbb{E}[\,(X,Y) \,|\, \max\{X,Y\} > Q_{\max\{X,Y\}}(\alpha)\,].$$

The interested reader is referred to Cai and Li [8] for further details. Using the same approach of Definition 5.1, we are interested to preserve the complete information about dependence structure between X and Y and the bi-dimensional aspect of our problem. To this end, starting from these general considerations, we propose to study a new bivariate version of the conditional tail expectation (Def. 5.2 below).

Let us first introduce the following assumption:

A2: X and Y are positive random variables fulfilling $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$. The random vector (X,Y) has a λ-density function f such that $\int f^{1+r} \mathrm{d}\lambda < \infty$, for some $r > 0$.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

Definition 5.2. Consider a random vector (X,Y) satisfying Assumption **A2**, with associate distribution function $F \in \mathcal{F}$. For $\alpha \in (0,1)$, we define

(1) the bivariate α-conditional tail expectation

$$\mathrm{CTE}_\alpha(X,Y) = \mathbb{E}[(X,Y)|(X,Y) \in L(\alpha)] = \begin{pmatrix} \mathbb{E}[\,X \,|\, (X,Y) \in L(\alpha)\,] \\ \mathbb{E}[\,Y \,|\, (X,Y) \in L(\alpha)\,] \end{pmatrix}.$$

(2) the estimated bivariate α-conditional tail expectation

$$\widehat{\mathrm{CTE}}_\alpha(X,Y) = \begin{pmatrix} \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}} \\ \frac{\sum_{i=1}^n Y_i 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)\}}} \end{pmatrix}. \tag{5.3}$$

Note that this bivariate conditional tail expectation is a natural extension of the univariate one. Moreover, if X and Y are identically distributed with a symmetric copula then $\mathbb{E}[\,X \,|\, (X,Y) \in L(\alpha)\,] = \mathbb{E}[\,Y \,|\, (X,Y) \in L(\alpha)\,]$.

Contrarily to the generalizations of the univariate CTE above, our $\mathrm{CTE}_\alpha(X,Y)$ does not use an aggregate variable (sum, min, max ...) in order to analyze the bivariate risk's issue. By contrast, using a geometric approach, $\mathrm{CTE}_\alpha(X,Y)$ rather deals with the simultaneous joint damages considering the bivariate dependence structure of data in a specific risk's area ($L(\alpha)$).

Let $\alpha \in (0,1)$ and $F \in \mathcal{F}$. We introduce truncated versions of the theoretical and estimated CTE_α:

$$\mathrm{CTE}_\alpha^T(X,Y) = \mathbb{E}[(X,Y)|(X,Y) \in L(\alpha)^T],$$

$$\widehat{\mathrm{CTE}}_\alpha^T(X,Y) = \begin{pmatrix} \frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}} \\ \frac{\sum_{i=1}^n Y_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^T\}}} \end{pmatrix},$$

where $L(\alpha)^T$ and $L_n(\alpha)^T$ are the truncated versions of theoretical and estimated upper α-level set defined in Section 2.

Theorem 5.3. *Under Assumption* **A2**, *Assumptions of Theorem 4.1 and with the notation of Theorem 4.1, it holds that*

$$\beta_n \left| \mathrm{CTE}_\alpha^{T_n}(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y) \right| \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0, \tag{5.4}$$

where $\beta_n = \min\{\, p_n^{\frac{r}{2(1+r)}}, a_n \,\}$, *with* $r > 0$ *satisfying Assumption* **A2** *and* $a_n = o(\sqrt{n})$.

The convergence in (5.4) must be interpreted as a componentwise convergence. In the case of a bounded density function $f_{(X,Y)}$ we obtain $\beta_n = \min\{\sqrt{p_n}, a_n\}$.

Using Theorem 5.3 and Remark 4.3, we obtain the following result in the case of the bivariate empirical distribution function.

Corollary 5.4. *Let* F_n *the bivariate empirical distribution function. Under Assumptions Theorem 5.3 and with the notation of Theorem 5.3, it holds that*

$$\beta_n \left| \mathrm{CTE}_\alpha^{T_n}(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y) \right| \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0,$$

where $\beta_n = o\left(n^{\frac{r}{6(1+r)}} / T_n^{\frac{2r}{3(1+r)}} \right)$, *with* $r > 0$ *satisfying Assumption* **A2**.

In the case of a bounded density function $f_{(X,Y)}$, $\beta_n = o\left(n^{\frac{1}{6}} / T_n^{\frac{2}{3}}\right)$.

Remark 5.5. Starting from Corollary 5.4, it could be interesting to consider the convergence $\left| \mathrm{CTE}_\alpha(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y) \right|$. We remark that in this case the speed of convergence will also depend on the

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

convergence rate to zero of $\left|\ \mathrm{CTE}_\alpha(X,Y) - \mathrm{CTE}_\alpha^{T_n}(X,Y)\right|$, then, in particular of $\mathbb{P}[(X,Y) \in L(\alpha) \setminus L(\alpha)^{T_n}]$ for $n \to \infty$. More precisely $\left|\mathrm{CTE}_\alpha(X,Y) - \mathrm{CTE}_\alpha^{T_n}(X,Y)\right|$ decays to zero at least with a convergence rate $(\mathbb{P}[X \geq T_n \text{ or } Y \geq T_n])^{-1}$. We now provide an illustration (in a simple case) on how to choose T_n in order to consider the convergence $\left|\mathrm{CTE}_\alpha(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)\right|$. Let $F(x,y) = (1-\mathrm{e}^{-x})(1-\mathrm{e}^{-y})$, *i.e.* independent and exponentially distributed marginals with parameter 1, with bounded density function in $\mathbb{R}_+^2$. Let F_n the bivariate empirical distribution function. In this case all assumptions of Corollary 5.4 are satisfied. Then we obtain that $\left|\mathrm{CTE}_\alpha(X,Y) - \widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)\right|$ decays to zero at least with a convergence rate $\beta_n = o(n^{\frac{1}{6}}/\ln(n)^{\frac{2}{3}})$, with a choice of sequence $T_n = \ln(n)$. We remark that $(\mathbb{P}[X \geq T_n \text{ or } Y \geq T_n])^{-1}$ is increasing in T_n, whereas the speed of convergence is decreasing in T_n. This kind of compromise provides an illustration on how to choose T_n, apart from satisfying the assumptions of consistency results in Section 4 (see Sect. 6.2.1 for several numerical illustrations of this aspect).

6. Illustrations

6.1. Estimation of the level sets

In this section we confront our estimator of level sets of the distribution function with various simulated samples. The plug-in estimation of level sets is constructed using the empirical estimator F_n. We consider two distribution functions which satisfy assumptions of Theorem 4.1: independent copula with exponential marginals and Survival Clayton copula with Burr marginals.

Following Remark 4.3 we consider $T_n = \ln(n)$ and we obtain $p_n = o(n^{\frac{1}{3}}/\ln(n)^{\frac{4}{3}})$. We also take $T_n = n^{0.45}$. This second sequence provides an illustration for the case of T_n too large (see Tabs. 1b and 2b, 4b and 5b).

We take a random grid of 10 000 points in $[0,T_n]^2$ and a distribution with independent and exponentially distributed marginals with parameter 1 and 2 respectively. We provide a Monte Carlo approximation for $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$ (averaged on 100 iterations), for different values of α and n. The results are gathered in Table 1.

In order to illustrate the influence of the choice of T_n, in Table 2 we show $p_n\,\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$, with $p_n = o(n^{\frac{1}{3}}/\ln(n)^{\frac{4}{3}})$ in the case $T_n = \ln(n)$ (Tab. 2a) and $T_n = n^{0.45}$ (Tab. 2b). Results in Table 2a set out how $p_n = o(n^{\frac{1}{3}}/\ln(n)^{\frac{4}{3}})$ is at least the convergence rate of $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$ in this particular case. This illustrates not only the consistency property but also the speed of convergence of our plug-in estimator. As expected in Table 2b we see that taking T_n too large, we can not expect to get a good approximation of $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$.

In Tables 3a and 3b we calculate the theoretical $\lambda(L(\alpha)^{T_n})$, for $T_n = \ln(n)$ and $T_n = n^{0.45}$ respectively, and for different values of n and α. These quantities provide an illustration of the truncation of level sets $L(\alpha)$, with respect to T_n and α.

Table 4 provides a Monte Carlo approximation for $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$ (averaged on 100 iterations), in the case of a Survival Clayton copula with parameter 1 and Burr $(2,1)$ marginals. In Table 5 we show $p_n\,\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$, with $p_n = o(n^{\frac{1}{3}}/\ln(n)^{\frac{4}{3}})$ in the case $T_n = \ln(n)$ (Tab. 5a) and $T_n = n^{0.45}$ (Tab. 5b). We obtain results that are analogous to those Table 2.

As expected, the greater n is, the better the estimations are (see Tabs. 1a and 4b). Moreover we note that for big values of α we need more data to get a good estimation of the level sets. This may come from the fact that when α grows the gradient of the distribution function decreases to zero and the constant A grows significantly (see proof of Theorem 4.1 in Sect. 8).

Remark 6.1. As remarked in Section 4 the choice of the sequence T_n is essentially a compromise: if T_n is large we have a quite slow convergence rate p_n for $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$ and if T_n is quite small we obtain a faster convergence rate. However choosing T_n too small we substantially truncate our level sets. This aspect can be problematic in the case c is close to one. Finally the theoretical optimal criterion for the choice of sequence T_n in Theorem 4.1 is an open interesting problem that we leave for a future work.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

TABLE 1. Distribution with independent and exponentially distributed marginals with parameter 1 and 2 respectively. Approximated $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$.

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	0.099	0.089	0.078
0.24	0.226	0.176	0.075
0.38	0.248	0.183	0.143
0.52	0.324	0.223	0.217
0.66	0.429	0.259	0.232
0.80	0.613	0.371	0.332

(b) $T_n = \ln(n)$

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	0.331	0.326	0.223
0.24	0.519	0.391	0.249
0.38	0.591	0.469	0.396
0.52	1.057	0.906	0.881
0.66	1.222	0.989	0.904
0.80	1.541	1.367	1.334

(b) $T_n = n^{0.45}$

TABLE 2. Distribution with independent and exponentially distributed marginals with parameter 1 and 2 respectively. Approximated $p_n\, \lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$, with $p_n = o(n^{\frac{1}{3}}/\ln(n)^{\frac{4}{3}})$.

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	0.069	0.068	0.065
0.24	0.156	0.134	0.063
0.38	0.172	0.139	0.121
0.52	0.225	0.169	0.183
0.66	0.298	0.199	0.195
0.80	0.426	0.282	0.279

(a) $T_n = \ln(n)$

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	0.229	0.248	0.188
0.24	0.361	0.298	0.209
0.38	0.411	0.357	0.339
0.52	0.734	0.689	0.743
0.66	0.849	0.752	0.762
0.80	1.071	1.039	1.124

(b) $T_n = n^{0.45}$

TABLE 3. Theoretical $\lambda(L(\alpha)^{T_n})$, for different values of n and α. Distribution with independent and exponentially distributed marginals with parameter 1 and 2 respectively.

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	37.472	46.459	56.405
0.24	35.777	44.588	54.358
0.38	33.837	42.436	51.995
0.52	31.505	39.837	49.129
0.66	28.513	36.487	45.421
0.80	24.201	31.621	40.003

(a) $T_n = \ln(n)$

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	142.205	248.516	433.872
0.24	139.039	244.378	428.449
0.38	135.328	239.495	422.021
0.52	130.771	233.464	414.045
0.66	124.779	225.488	403.451
0.80	115.851	213.505	387.436

(b) $T_n = n^{0.45}$

TABLE 4. Distribution with Survival Clayton copula with parameter 1 and Burr$(2, 1)$ marginals. Approximated $\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$.

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	0.208	0.171	0.126
0.24	0.252	0.241	0.149
0.38	0.339	0.288	0.132
0.52	0.349	0.335	0.245
0.66	0.499	0.364	0.353
0.80	0.771	0.644	0.605

(b) $T_n = \ln(n)$

α	$n = 500$	$n = 1000$	$n = 2000$
0.10	0.697	0.633	0.536
0.24	0.893	0.872	0.809
0.38	0.971	0.911	0.879
0.52	1.001	0.982	1.229
0.66	1.569	1.522	1.413
0.80	2.377	2.269	2.175

(b) $T_n = n^{0.45}$

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

TABLE 5. Distribution with Survival Clayton copula with parameter 1 and Burr $(2,1)$ marginals. Approximated $p_n\,\lambda(L(\alpha)^{T_n} \triangle L_n(\alpha)^{T_n})$, with $p_n = o(n^{\frac{1}{3}}/\ln(n)^{\frac{4}{3}})$.

α	$n=500$	$n=1000$	$n=2000$
0.10	0.144	0.129	0.106
0.24	0.175	0.183	0.126
0.38	0.235	0.219	0.111
0.52	0.242	0.254	0.206
0.66	0.347	0.277	0.297
0.80	0.535	0.489	0.501

(b) $T_n = \ln(n)$

α	$n=500$	$n=1000$	$n=2000$
0.10	0.484	0.481	0.452
0.24	0.620	0.663	0.682
0.38	0.674	0.692	0.741
0.52	0.695	0.746	1.036
0.66	1.089	1.157	1.191
0.80	1.651	1.725	1.834

(b) $T_n = n^{0.45}$

6.2. Estimation of $\mathrm{CTE}_{\alpha}(X,Y)$

6.2.1. Simulated data

In order to evaluate the performance of our estimator we present here some simulated cases of estimation of $\mathrm{CTE}_\alpha(X,Y)$, for different values of level α. To compare the estimated results with the theoretical ones we consider cases for which we can calculate (with Maple) the explicit value of the theoretical $\mathrm{CTE}_\alpha(X,Y)$. However our estimator can be applied to much more general cases. By being able to get a good estimator for the level sets, we can expect a good estimator for $\mathrm{CTE}_\alpha(X,Y)$.

In the following we compare $\widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)$ with the theoretical $\mathrm{CTE}_\alpha(X,Y)$, for different distributions: Independent copula with exponentially distributed marginals; Clayton copula with parameter 1, with exponential and Burr $(4,1)$ univariate marginals. The plug-in estimation of level sets is constructed using the empirical estimator F_n of the bivariate distribution function, with $n=1000$.

In the following we denote $\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n}}(X,Y) = \big(\overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}}(X,Y), \overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,2}}(X,Y)\big)$ the mean (coordinate by coordinate) of $\widehat{\mathrm{CTE}}_\alpha^{T_n}(X,Y)$ on 100 simulations. We denote $\widehat{\sigma} = (\widehat{\sigma}_1, \widehat{\sigma}_2)$ the empirical standard deviation (coordinate by coordinate) with

$$\widehat{\sigma}_1 = \sqrt{\frac{1}{99}\sum_{j=1}^{100}\left(\widehat{\mathrm{CTE}}_\alpha^{T_n,1}(X,Y)_j - \overline{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}}(X,Y)\right)^2}$$

relatives to the first coordinate (resp. $\widehat{\sigma}_2$ relatives to the second one).

We denote $\mathrm{RMSE} = (\mathrm{RMSE}_1, \mathrm{RMSE}_2)$ the relative mean square error (coordinate by coordinate) with

$$\mathrm{RMSE}_1 = \sqrt{\frac{1}{100}\sum_{j=1}^{100}\left(\frac{\widehat{\mathrm{CTE}}_\alpha^{T_n,1}(X,Y)_j - \mathrm{CTE}_\alpha^{T_n,1}(X,Y)}{\mathrm{CTE}_\alpha^{T_n,1}(X,Y)}\right)^2}$$

relatives to the first coordinate of $\mathrm{CTE}_\alpha^{T_n}(X,Y)$ (resp. RMSE_2 relatives to the second one).

For the Independent copula with exponentially distributed marginals, from Corollary 5.4 and Remark 5.5, we take $T_n = \ln(n)$ (see Tabs. 6 and 8). This choice provides a convergence rate $\beta_n = o(n^{\frac{1}{6}}/\ln(n)^{\frac{2}{3}})$. Furthermore, as in Section 6.1, we consider the case with $T_n = n^{0.45}$. This second sequence provides an illustration for the case of T_n too large (see Tabs. 7 and 9). Following the same approach of Remark 5.5, in the case of Clayton copula with parameter 1, with exponential and Burr$(4,1)$ univariate marginals we have $\beta_n = o(n^{\frac{1}{6}}/(n^{0.2})^{\frac{2}{3}})$, with a sequence $T_n = n^{0.2}$ (see Tab. 10). Also in this case, we show the behavior of our estimator for $T_n = n^{0.45}$ (see Tab. 11).

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

TABLE 6. (X, Y) with independent and exponentially distributed components with parameter 2; $T_n = \ln(n)$.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\overline{\widehat{\mathrm{CTE}_\alpha}^{T_n}}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(0.627, 0.627)	(0.585, 0.637)	(0.022, 0.021)	(0.035, 0.039)
0.24	(0.761, 0.761)	(0.766, 0.759)	(0.031, 0.031)	(0.041, 0.046)
0.38	(0.896, 0.896)	(0.864, 0.917)	(0.036, 0.037)	(0.053, 0.049)
0.52	(1.051, 1.051)	(1.027, 1.071)	(0.051, 0.053)	(0.055, 0.052)
0.66	(1.246, 1.246)	(1.252, 1.238)	(0.075, 0.067)	(0.063, 0.065)
0.80	(1.531, 1.531)	(1.497, 1.549)	(0.123, 0.117)	(0.083, 0.077)

TABLE 7. (X, Y) with independent and exponentially distributed components with parameter 2; $T_n = n^{0.45}$.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\overline{\widehat{\mathrm{CTE}_\alpha}^{T_n}}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(0.627, 0.627)	(0.603, 0.656)	(0.031, 0.031)	(0.062, 0.068)
0.24	(0.761, 0.761)	(0.774, 0.731)	(0.061, 0.071)	(0.082, 0.130)
0.38	(0.896, 0.896)	(0.927, 0.871)	(0.072, 0.076)	(0.087, 0.119)
0.52	(1.051, 1.051)	(1.086, 1.128)	(0.094, 0.082)	(0.095, 0.107)
0.66	(1.246, 1.246)	(1.281, 1.322)	(0.127, 0.101)	(0.102, 0.101)
0.80	(1.531, 1.531)	(1.545, 1.611)	(0.157, 0.161)	(0.105, 0.117)

TABLE 8. (X, Y) with independent and exponentially distributed components with parameter 1 and 2 respectively; $T_n = \ln(n)$.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\overline{\widehat{\mathrm{CTE}_\alpha}^{T_n}}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(1.255, 0.627)	(1.222, 0.638)	(0.044, 0.022)	(0.043, 0.039)
0.24	(1.521, 0.761)	(1.488, 0.811)	(0.069, 0.023)	(0.051, 0.042)
0.38	(1.792, 0.896)	(1.797, 0.911)	(0.075, 0.038)	(0.044, 0.046)
0.52	(2.102, 1.051)	(2.082, 1.047)	(0.104, 0.052)	(0.052, 0.052)
0.66	(2.492, 1.246)	(2.461, 1.255)	(0.139, 0.071)	(0.057, 0.056)
0.80	(3.061, 1.531)	(3.011, 1.544)	(0.251, 0.125)	(0.084, 0.082)

TABLE 9. (X, Y) with independent and exponentially distributed components with parameter 1 and 2 respectively; $T_n = n^{0.45}$.

α	$\mathrm{CTE}_\alpha(X,Y)$	$\overline{\widehat{\mathrm{CTE}_\alpha}^{T_n}}(X,Y)$	$\widehat{\sigma}$	RMSE
0.10	(1.255, 0.627)	(1.253, 0.624)	(0.061, 0.043)	(0.051, 0.054)
0.24	(1.521, 0.761)	(1.514, 0.803)	(0.074, 0.049)	(0.048, 0.075)
0.38	(1.792, 0.896)	(1.793, 0.948)	(0.096, 0.055)	(0.053, 0.084)
0.52	(2.102, 1.051)	(2.087, 1.111)	(0.118, 0.076)	(0.056, 0.092)
0.66	(2.492, 1.246)	(2.497, 1.311)	(0.169, 0.108)	(0.068, 0.101)
0.80	(3.061, 1.531)	(3.056, 1.602)	(0.313, 0.153)	(0.102, 0.111)

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

TABLE 10. (X, Y) with Clayton copula with parameter 1, F_X exponential distribution with parameter 1, F_Y Burr$(4, 1)$ distribution; $T_n = n^{0.2}$.

α	$\text{CTE}_\alpha(X, Y)$	$\overline{\widehat{\text{CTE}_\alpha}^{T_n}}(X, Y)$	$\widehat{\sigma}$	RMSE
0.10	(1.188, 1.229)	(1.049, 1.192)	(0.032, 0.021)	(0.019, 0.033)
0.24	(1.448, 1.366)	(1.283, 1.379)	(0.053, 0.224)	(0.019, 0.063)
0.38	(1.727, 1.505)	(1.525, 1.471)	(0.046, 0.031)	(0.019, 0.031)
0.52	(2.049, 1.666)	(1.803, 1.625)	(0.058, 0.041)	(0.023, 0.034)
0.66	(2.454, 1.875)	(2.129, 1.823)	(0.071, 0.054)	(0.035, 0.039)
0.80	(3.039, 2.202)	(2.591, 2.144)	(0.111, 0.105)	(0.055, 0.054)

TABLE 11. (X, Y) with Clayton copula with parameter 1, F_X exponential distribution with parameter 1, F_Y Burr$(4, 1)$ distribution; $T_n = n^{0.45}$.

α	$\text{CTE}_\alpha(X, Y)$	$\overline{\widehat{\text{CTE}_\alpha}^{T_n}}(X, Y)$	$\widehat{\sigma}$	RMSE
0.10	(1.188, 1.229)	(1.189, 1.238)	(0.061, 0.035)	(0.039, 0.029)
0.24	(1.448, 1.366)	(1.462, 1.365)	(0.065, 0.037)	(0.046, 0.031)
0.38	(1.727, 1.505)	(1.751, 1.536)	(0.082, 0.046)	(0.049, 0.037)
0.52	(2.049, 1.666)	(2.063, 1.713)	(0.091, 0.061)	(0.051, 0.045)
0.66	(2.454, 1.875)	(2.457, 1.951)	(0.117, 0.104)	(0.057, 0.068)
0.80	(3.039, 2.202)	(3.037, 2.322)	(0.192, 0.165)	(0.063, 0.108)

TABLE 12. Evolution of $\widehat{\sigma}$ and RMSE in terms of sample size n for $\alpha = 0.9$; (X, Y) independent and exponentially distributed components with parameter 1 and 2 respectively. $T_n = \ln(n)$.

n	500	1000	1500	2000	2500	5000
$\widehat{\sigma}$	(0.614, 0.359)	(0.444, 0.308)	(0.431, 0.295)	(0.377, 0.168)	(0.241, 0.123)	(0.216, 0.121)
RMSE	(0.168, 0.189)	(0.123, 0.163)	(0.115, 0.161)	(0.099, 0.089)	(0.077, 0.079)	(0.063, 0.057)

TABLE 13. Evolution of $\widehat{\sigma}$ and RMSE in terms of sample size n for $\alpha = 0.9$; (X, Y) independent and exponentially distributed components with parameter 1 and 2 respectively. $T_n = n^{0.45}$.

n	500	1000	1500	2000	2500	5000
$\widehat{\sigma}$	(0.919, 0.419)	(0.568, 0.319)	(0.511, 0.294)	(0.382, 0.239)	(0.348, 0.223)	(0.307, 0.151)
RMSE	(0.242, 0.221)	(0.151, 0.172)	(0.133, 0.165)	(0.101, 0.144)	(0.093, 0.129)	(0.092, 0.108)

Following we show that for high levels (here $\alpha = 0.9$), one needs to use large samples to get reasonable estimates of CTE_α. We consider for the purpose (X, Y) independent and exponentially distributed with respective parameters 1 and 2. The theoretical value is $\text{CTE}_{0.9}(X, Y) = (3.78, 1.89)$. We take $T_n = \ln(n)$ (see Tab. 12). In this case we need between 2000 and 2500 data to get the same performances as for lower level (see Tab. 8 above). Furthermore we consider the case $T_n = n^{0.45}$ (see Tab. 13). In this case we need between 2500 and 5000 data to get the same performances as for lower level (see Tab. 9 above).

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

The bad properties of the estimate for sample sizes less than 2000 (in the case of $T_n = \ln(n)$) can be explained by the fact that for high levels, the constant A is large (see Proof of Theorem 4.1 in Sect. 8), but also by the fact that for large values of α the empirical estimate F_n of F may not be the most appropriate.

6.2.2. Real data

We consider here the estimation of CTE_α in a real case: **Loss-ALAE data** in the log scale (for details see Frees and Valdez [19]). Each claim consists of an indemnity payment (the loss, X) and an allocated loss adjustment expense (ALAE, Y). Examples of ALAE are the fees paid to outside attorneys, experts, and investigators used to defend claims.

The data size is $n = 1500$. Let $T_n = n^{0.4}$. Again our estimator is constructed using the empirical estimator F_n. Results about the estimation of $\mathrm{CTE}_\alpha^{T_n}$ are gathered in Table 14. In Figure 2 we represent data, estimated level sets and estimated bivariate conditional tail expectation for several values of α.

In this real setting the estimation of CTE_α can be used in order to quantify the mean of the Loss (resp. ALAE) in the log scale conditionally to the fact that Loss and ALAE data belong jointly to the specific risk's area $L(\alpha)$.

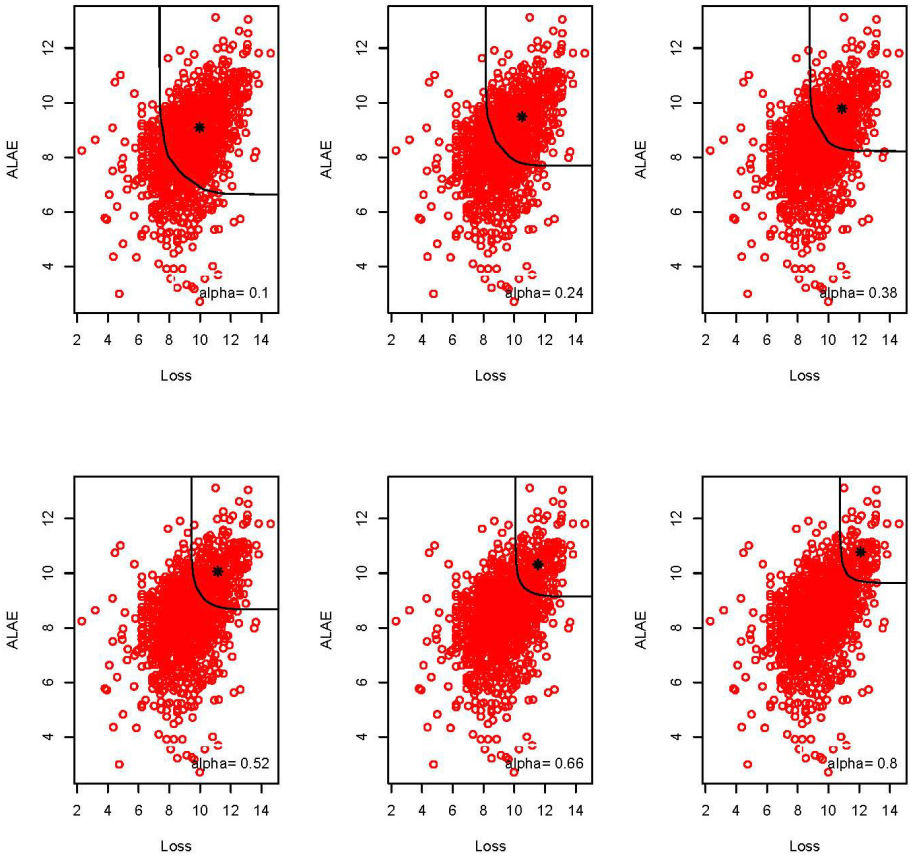


FIGURE 2. Loss-ALAE data in log scale: boundary of estimated level sets (line) and $\widehat{\mathrm{CTE}}_\alpha^{T_n}$ (star) for different values of α.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

TABLE 14. $\widehat{\text{CTE}}_{\alpha}^{T_n}$ for Loss-ALAE data in log scale, with different values of level α.

α	0.10	0.24	0.38
$\widehat{\text{CTE}}_{\alpha}^{T_n}$	(9.937, 9.252)	(10.361, 9.566)	(10.731, 9.728)
α	0.52	0.66	0.80
$\widehat{\text{CTE}}_{\alpha}^{T_n}$	(11.096, 10.011)	(11.518, 10.315)	(12.057, 10.758)

7. Conclusion

In this paper we have provided convergence results for the plug-in estimator of the level sets of an unknown distribution function on $\mathbb{R}_+^2$ in terms of Hausdorff distance and volume of the symmetric difference. In this setting we have proposed and estimated a new bivariate risk measure: $\text{CTE}_{\alpha}(X, Y)$. A future work comparing our bivariate Conditional Tail Expectation with existing bivariate generalizations of the CTE in terms of classical properties (monotonicity, translation invariance, homogeneity, ...), dependence structure, behavior with respect to different risk scenarios, is in preparation.

8. Proofs

Proof of Proposition 2.1. Take $T > 0$ such that for all $t : \mid t - c \mid \leq r,\ \{F = t\}^T \neq \emptyset$, (from assumptions of Proposition 2.1 we know that such an T there exists).

Let $x \in \{z \in [0,T]^2 : \mid F(z) - c \mid \leq r\}$. Define for $\overline{\lambda} \in \mathbb{R}$

$$y_{\overline{\lambda}} \equiv y_{\overline{\lambda},x} = x + \overline{\lambda}\frac{(\nabla F)_x}{\|(\nabla F)_x\|},$$

so $\|y_{\overline{\lambda}} - x\| = \mid \overline{\lambda} \mid$. From the differentiability properties of F and using Taylor's formula we have, for $\mid \overline{\lambda} \mid < \zeta$

$$F(y_{\overline{\lambda}}) = F(x) + (\nabla F)_x^T(y_{\overline{\lambda}} - x) + \frac{1}{2}(y_{\overline{\lambda}} - x)^T(HF)_{\overline{x}}(y_{\overline{\lambda}} - x),$$

with $\overline{x}$ a point on the segment between x and $y_{\overline{\lambda}}$. So

$$F(y_{\overline{\lambda}}) = F(x) + \overline{\lambda}\|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2\|(\nabla F)_x\|^2}(\nabla F)_x^T(HF)_{\overline{x}}(\nabla F)_x.$$

By the Cauchy–Schwarz inequality we obtain

$$F(y_{\overline{\lambda}}) \geq F(x) + \overline{\lambda}\|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2\|(\nabla F)_x\|}\|(HF)_{\overline{x}}(\nabla F)_x\|$$

and

$$F(y_{\overline{\lambda}}) \leq F(x) + \overline{\lambda}\|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2\|(\nabla F)_x\|}\|(HF)_{\overline{x}}(\nabla F)_x\|.$$

Since $\|(HF)_{\overline{x}}(\nabla F)_x\| \leq \|(HF)_{\overline{x}}\|\|(\nabla F)_x\|$, we have

$$F(x) + \overline{\lambda}\|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2}\|(HF)_{\overline{x}}\| \leq F(y_{\overline{\lambda}}) \leq F(x) + \overline{\lambda}\|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2}\|(HF)_{\overline{x}}\|.$$

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

Since $\overline{x} \in E$ and $M_H < \infty$ we obtain

$$F(x) + \overline{\lambda}\|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2} M_H \leq F(y_{\overline{\lambda}}) \leq F(x) + \overline{\lambda}\|(\nabla F)_x\| + \frac{\overline{\lambda}^2}{2} M_H. \tag{8.1}$$

For $0 < \overline{\lambda} < \zeta$, we have from the left side of (8.1)

$$F(y_{\overline{\lambda}}) \geq F(x) + \overline{\lambda}\|(\nabla F)_x\| - \frac{\overline{\lambda}^2}{2} M_H \geq F(x) + \overline{\lambda} m^{\nabla} - \frac{\overline{\lambda}^2}{2} M_H.$$

We assume now that $M_H > 0$ (the case $M_H = 0$ is trivial).

For $x \in \{z \in [0,T]^2 : \mid F(z) - c \mid \leq r\}$ and $0 < \overline{\lambda} < \zeta \wedge \frac{m^{\nabla}}{M_H}$ we have then

$$F(y_{\overline{\lambda}}) \geq F(x) + \frac{\overline{\lambda}}{2} m^{\nabla}, \tag{8.2}$$

Similarly using the right side of (8.1) we obtain for $0 < \overline{\lambda} < \zeta \wedge \frac{m^{\nabla}}{M_H}$ and $x \in \{z \in [0,T]^2 : \mid F(z) - c \mid \leq r\}$,

$$F(y_{-\overline{\lambda}}) \leq F(x) - \frac{\overline{\lambda}}{2} m^{\nabla}. \tag{8.3}$$

Define

$$\gamma = \left(\frac{m^{\nabla}}{4} \left(\zeta \wedge \frac{m^{\nabla}}{M_H} \right) \right) \wedge r > 0.$$

Suppose that $t = c + \varepsilon$, $0 < \varepsilon \leq \gamma$. Let $x \in [0,T]^2$ such that $F(x) = t = c + \varepsilon$ then $x \in \{z \in [0,T]^2 : \mid F(z) - c \mid \leq r\}$. Take now

$$0 < \overline{\lambda} = \frac{2\varepsilon}{m^{\nabla}} < \zeta \wedge \frac{m^{\nabla}}{M_H}.$$

We obtain from (8.3)

$$F(y_{-\overline{\lambda}}) \leq F(x) - \frac{\overline{\lambda}}{2} m^{\nabla} = c + \varepsilon - \varepsilon = c.$$

From the continuity property of F, we deduce that there exists y between x and $y_{-\overline{\lambda}}$ such that $F(y) = c$ and we have

$$\|x - y\| \leq \|x - y_{-\overline{\lambda}}\| = \mid \overline{\lambda} \mid = \frac{2\varepsilon}{m^{\nabla}} = \frac{2}{m^{\nabla}} \mid t - c \mid .$$

So we have proved that

$$\sup_{x \in \{F=t\}^T} d(x, \{F = c\}^T) \leq \frac{2}{m^{\nabla}} \mid t - c \mid .$$

Similarly, take $x \in [0,T]^2$ such that $F(x) = c$ and use (8.2) to obtain

$$\sup_{x \in \{F=c\}^T} d(x, \{F = t\}^T) \leq \frac{2}{m^{\nabla}} \mid t - c \mid .$$

The proof in case $t < c$ is completely analogous. So F satisfies Assumption **H** (see Sect. 2) with $A = \frac{2}{m^{\nabla}}$. □

Proof of Theorem 3.1. Under assumptions of Theorem 3.1, we can always take $T_1 > 0$ such that for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_1} \neq \emptyset$.

Then for each n, for all $t : \mid t - c \mid \leq r$, $\partial L(t)^{T_n}$ is a non-empty (and compact) set on $\mathbb{R}^2_+$.

In each $[0,T_n]^2$, from Proposition 2.1, Assumption **H** (Sect. 2) is satisfied with

$$\gamma = \left(\frac{m^{\nabla}}{4} \left(\zeta \wedge \frac{m^{\nabla}}{M_H} \right) \right) \wedge r > 0 \quad \text{and} \quad A = \frac{2}{m^{\nabla}}.$$

First we have to find a bound for $\sup_{x \in \partial L(c)^{T_n}} d(x, \partial L_n(c)^{T_n})$.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

Take $x \in \partial L(c)^{T_n}$ and define $\varepsilon_n = 2\,\|F - F_n\|_\infty^{T_n}$. Using $\|F - F_n\|_\infty \to 0$, a.s., for $n \to \infty$, then $\varepsilon_n \to 0$, a.s., for $n \to \infty$. So with probability one there exists n_0 such that $\forall\, n \geq n_0$, $\varepsilon_n \leq \gamma$.

Since for all $t : |t - c| \leq r$ $\partial L(t)^{T_n} \neq \emptyset$, from Assumption **H**, there exist $u_n \equiv u_x^{\varepsilon_n}$ and $l_n \equiv l_x^{\varepsilon_n}$ in $[0, T_n]^2$ such that

$$F(u_n) = c + \varepsilon_n; \quad d(x, u_n) \leq A\,\varepsilon_n,$$
$$F(l_n) = c - \varepsilon_n; \quad d(x, l_n) \leq A\,\varepsilon_n.$$

Suppose now $\|F - F_n\|_\infty^{T_n} > 0$ (the other case is a trivial one). In this case

$$F_n(u_n) = c + \varepsilon_n + F_n(u_n) - F(u_n) \geq c + \varepsilon_n - \|F - F_n\|_\infty^{T_n} = c + 2\|F - F_n\|_\infty^{T_n} - \|F - F_n\|_\infty^{T_n} > c,$$

and in a similar way we can prove that $F_n(l_n) < c$.

As $F_n(l_n) < c$ and $F_n(u_n) > c$, with u_n and l_n in $[0, T_n]^2$, there exists $z_n \in \partial L_n(c)^{T_n} \cap B(u_n, d(u_n, l_n))$ with

$$d(z_n, x) \leq d(z_n, u_n) + d(u_n, x) \leq d(u_n, l_n) + d(u_n, x) \leq d(u_n, x) + d(x, l_n) + d(u_n, x) \leq 3\,A\,\varepsilon_n = 6\,A\|F - F_n\|_\infty^{T_n}.$$

Hence, for $n \geq n_0$

$$\sup_{x \in \partial L(c)^{T_n}} d(x, \partial L_n(c)^{T_n}) \leq\ 6\,A\,\|F - F_n\|_\infty^{T_n}.$$

Let us now bound $\sup_{x \in \partial L_n(c)^{T_n}} d(x, \partial L(c)^{T_n})$.

Take $x \in \partial L_n(c)^{T_n}$. From the a.s. continuity of F_n we obtain $F_n(x) = c$, a.s., so

$$|\,F(x) - c\,| \leq |\,F(x) - F_n(x)\,| \leq\ \|F - F_n\|_\infty^{T_n}\ \leq\ \varepsilon_n, \quad \text{a.s.}$$

Remember that $\forall\ n \geq n_0$, $\varepsilon_n \leq \gamma$, a.s. Then from Assumption **H** $d(x, \partial L(c)^{T_n}) \leq\ A\ |\,F(x) - c\,| \leq\ A\,\|F - F_n\|_\infty^{T_n}$. We can conclude that with probability one, for $n \geq n_0$

$$\sup_{x \in \partial L_n(c)^{T_n}} d(x, \partial L(c)^{T_n}) \leq\ A\,\|F - F_n\|_\infty^{T_n}.$$

We obtain for $n \geq n_0$, $d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) \leq\ 6\,A\,\|F - F_n\|_\infty^{T_n}$, then

$$d_H(\partial L(c)^{T_n}, \partial L_n(c)^{T_n}) = O(\|F - F_n\|_\infty^{T_n}), \quad \text{a.s.}$$

Hence the result. □

Proof of Theorem 4.1. Under assumptions of Theorem 4.1, we can always take $T_1 > 0$ such that for all $t : |t - c| \leq r$, $\partial L(t)^{T_1} \neq \emptyset$. Then for each n, for all $t : |t - c| \leq r$, $\partial L(t)^{T_n}$ is a non-empty (and compact) set on $\mathbb{R}_+^2$.

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 $L(c)^{T_n} \triangle L_n(c)^{T_n}$, $Q_{\varepsilon_n} = \{x \in [0, T_n]^2 : |\,F - F_n\,| \leq \varepsilon_n\}$ and $\widetilde{Q}_{\varepsilon_n} = \{x \in [0, T_n]^2 : |\,F - F_n\,| > \varepsilon_n\}$ are measurable and

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) = \lambda(L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n}) + \lambda(L(c)^{T_n} \triangle L_n(c)^{T_n} \cap \widetilde{Q}_{\varepsilon_n}). \tag{8.4}$$

Since $L(c)^{T_n} \triangle L_n(c)^{T_n} \cap Q_{\varepsilon_n} \subset \{x \in [0, T_n]^2 : c - \varepsilon_n \leq F < c + \varepsilon_n\}$ we obtain

$$\lambda(L(c)^{T_n} \triangle L_n(c)^{T_n}) \leq \lambda(\{x \in [0, T_n]^2 : c - \varepsilon_n \leq F < c + \varepsilon_n\}) + \lambda(\widetilde{Q}_{\varepsilon_n}).$$

From Assumption **H** (Sect. 2) and Proposition 2.1, if $2\,\varepsilon_n \leq \gamma$ then

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

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

So we can write

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

If we now choose

$$\varepsilon_n = o\left(\frac{1}{p_n\, T_n}\right) \tag{8.5}$$

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

$$p_n\, \lambda(\{x \in [0,T_n]^2 : c-\varepsilon_n \le F < c+\varepsilon_n\}) \underset{n\to\infty}{\to} 0.$$

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

$$p_n\, \lambda(\widetilde{Q}_{\varepsilon_n}) = p_n \int 1_{\{x\in[0,T_n]^2:\, |F-F_n|>\varepsilon_n\}}\, \lambda(\mathrm{d}x) \le \frac{p_n}{\varepsilon_n^p} \int_{[0,T_n]^2} |\, F-F_n\,|^p\ \lambda(\mathrm{d}x). \tag{8.6}$$

Take ε_n such that

$$\varepsilon_n = \left(\frac{p_n}{v_n}\right)^{\frac{1}{p}}. \tag{8.7}$$

Then, from Assumption **A1**, in Section 4, we obtain $p_n\, \lambda(\widetilde{Q}_{\varepsilon_n}) \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0$. As $p_n = o(v_n^{1/p+1}/T_n^{p/p+1})$ we can choose ε_n that satisfies (8.5) and (8.7). Hence the result. □

Proof of Theorem 5.3. We only prove the result for the first coordinate of $\mathrm{CTE}_\alpha(X,Y)$ (the proof is similar for the second one). First we remark that, in the following, probabilities involving events which depend on $L_n(\alpha)^{T_n}$, are conditional probabilities to the sample (X_i,Y_i), for $i=1,\dots n$. We will denote $\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]$ purely for reasons of notational convenience.

We introduce these two preliminary results (Lems. 8.1 and 8.2):

Lemma 8.1. *Under Assumption* **A2**, *Assumptions of Theorem 4.1 and with the notation of Theorem 4.1, it holds that*

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

with $r>0$ *satisfying Assumption* **A2**.

Proof of Lemma 8.1. From Assumption **A2** and Theorem 4.1 we obtain

$$p_n^{\frac{r}{2(1+r)}} \left|\mathbb{P}[(X,Y)\in L(\alpha)^{T_n} \,\triangle\, L_n(\alpha)^{T_n}]\right| \le p_n^{\frac{r}{2(1+r)}}\, d_\lambda\big(L(\alpha)^{T_n}, L_n(\alpha)^{T_n}\big)^{\frac{r}{1+r}}\, \|\, f\,\|_{1+r} \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0. \tag{8.8}$$

As a straightforward consequence we find

$$p_n^{\frac{r}{2(1+r)}} \left|\mathbb{P}[(X,Y)\in L(\alpha)^{T_n}] - \mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]\right| \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0.$$

Using Assumption **A2** we also obtain

$$p_n^{\frac{r}{2(1+r)}} \left|\int_{L(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y) - \int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)\right|$$
$$\le p_n^{\frac{r}{2(1+r)}}\ \mathbb{E}[X^2]^{\frac{1}{2}}\ d_\lambda\big(L(\alpha)^{T_n}, L_n(\alpha)^{T_n}\big)^{\frac{r}{2(1+r)}}\ \|\, f\,\|_{1+r}^{\frac{1}{2}} \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0. \tag{8.9}$$

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

Then

$$
\begin{aligned}
p_n^{\frac{r}{2(1+r)}} & \left|\mathbb{E}[X|(X,Y)\in L(\alpha)^{T_n}]-\mathbb{E}[X|(X,Y)\in L_n(\alpha)^{T_n}]\right| \\
&= p_n^{\frac{r}{2(1+r)}}\left|\int_{L(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)\ \mathbb{P}[(X,Y)\in L(\alpha)^{T_n}]^{-1}\right. \\
&\quad \left.-\int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)\ \mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]^{-1}\right| \\
&\le \frac{p_n^{\frac{r}{2(1+r)}}}{\mathbb{P}[(X,Y)\in L(\alpha)^{T_n}]\,\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]}\left(\mathbb{P}[(X,Y)\in L(\alpha)^{T_n}]\right. \\
&\quad \times\left|\int_{L(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)-\int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)\right| \\
&\quad \left.+\int_{L(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)\left|\mathbb{P}[(X,Y)\in L(\alpha)^{T_n}]-\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]\right|\right). \qquad (8.10)
\end{aligned}
$$

Using (8.8)–(8.9) we obtain that (8.10) converges to zero in probability, for $n\to\infty$. Hence the result. □

Lemma 8.2. *Under Assumption* **A2**, *Assumptions of Theorem 4.1 and with the notation of Theorem 4.1, it holds that*

$$
a_n\left|\mathbb{E}[\,X\,|\,(X,Y)\in L_n(\alpha)^{T_n}\,]-\frac{\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right|\underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0,
$$

for $a_n = o(\sqrt{n})$.

Proof of Lemma 8.2. We can write

$$
\begin{aligned}
a_n & \left|\mathbb{E}[X|(X,Y)\in L_n(\alpha)^{T_n}]-\frac{\displaystyle\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\displaystyle\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right| \\
&= a_n\left|\frac{\displaystyle\int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)}{\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]}-\frac{\displaystyle\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\displaystyle\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}\right|.
\end{aligned}
$$

Under assumptions of Lemma 8.2, from the central limit theorem for triangular arrays (*e.g.* Thm. 27.2 in Billingsley [6]) we obtain

$$
\begin{aligned}
& a_n\left|\mathbb{P}[(X,Y)\in L_n(\alpha)^{T_n}]-\frac{1}{n}\sum_{i=1}^n 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}\right|\underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0, \\
& a_n\left|\int_{L_n(\alpha)^{T_n}} x\, f_{X,Y}(x,y)\lambda(\mathrm{d}x\,\mathrm{d}y)-\frac{1}{n}\sum_{i=1}^n X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}\right|\underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0.
\end{aligned}
$$

Hence the result. □

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

Then to prove Theorem 5.3 we can write (5.4) as

$$\begin{aligned}
&\beta_n \left| \mathbb{E}[X|(X,Y) \in L(\alpha)^{T_n}] - \frac{\sum_{i=1}^{n} X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^{n} 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}} \right| \\
&\quad \leq \beta_n \left| \mathbb{E}[X|(X,Y) \in L(\alpha)^{T_n}] - \mathbb{E}[X|(X,Y) \in L_n(\alpha)^{T_n}] \right| \\
&\quad + \beta_n \left| \mathbb{E}[X|(X,Y) \in L_n(\alpha)^{T_n}] - \frac{\sum_{i=1}^{n} X_i 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}}{\sum_{i=1}^{n} 1_{\{(X_i,Y_i)\in L_n(\alpha)^{T_n}\}}} \right|.
\end{aligned}$$

The result is a straightforward application of Lemmas 8.1 and 8.2. □

Acknowledgements. The authors thank two anonymous referees for relevant remarks and constructive comments on a previous version of the paper. The authors thank Alberto Rodríguez–Casal for his fruitful comments. They thank also Stéphane Loisel and Ragnar Norberg for helpful discussions. This work has been partially supported by the French research national agency (ANR) under the reference ANR-08BLAN-0314-01. The fourth author thanks the BNP Paribas Cardif Chair "Management de la modélisation" for financial support. The views expressed in this document are the author's own and do not necessarily reflect those endorsed by BNP Paribas Cardif.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, Coherent measures of risk. *Math. Finance* **9** (1999) 203–228.
[2] A. Baíllo, Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.* **65** (2003) 411–417.
[3] A. Baíllo, J.A. Cuesta-Albertos and A. Cuevas, Convergence rates in nonparametric estimation of level sets. *Statist. Probab. Lett.* **53** (2001) 27–35.
[4] F. Belzunce, A. Castaño, A. Olvera-Cervantes and A. Suárez-Llorens, Quantile curves and dependence structure for bivariate distributions. *Comput. Stat. Data Anal.* **51** (2007) 5112–5129.
[5] G. Biau, B. Cadre and B. Pelletier, A graph-based estimator of the number of clusters. *ESAIM: PS* **11** (2007) 272–280.
[6] P. Billingsley, *Probability and measure*. Wiley Series in Probability and Mathematical Statistics, 3th edition, John Wiley & Sons Inc., A Wiley-Interscience Publication, New York (1995).
[7] B. Cadre, Kernel estimation of density level sets. *J. Multivar. Anal.* **97** (2006) 999–1023.
[8] J. Cai and H. Li, Conditional tail expectations for multivariate phase-type distributions. *J. Appl. Probab.* **42** (2005) 810–825.
[9] L. Cavalier, Nonparametric estimation of regression level sets. *Statistics (Berl. DDR)* **29** (1997) 131–160.
[10] Y.P. Chaubey and P.K. Sen, Smooth estimation of multivariate survival and density functions. *J. Statist. Plann. Inference* **103** (2002) 361–376; C. R. Rao 80th birthday felicitation volume, Part I.
[11] A. Cuevas and R. Fraiman, A plug-in approach to support estimation. *Ann. Stat.* **25** (1997) 2300–2312.
[12] A. Cuevas and A. Rodríguez–Casal, On boundary estimation. *Adv. Appl. Probab.* **36** (2004) 340–354.
[13] A. Cuevas, W. González-Manteiga and A. Rodríguez–Casal, Plug-in estimation of general level sets. *Australian & New Zealand J. Statist.* **48** (2006) 7–19.
[14] L. de Haan and X. Huang, Large quantile estimation in a multivariate setting. *J. Multivar. Anal.* **53** (1995) 247–263.
[15] S. Dedu and R. Ciumara, Restricted optimal retention in stop-loss reinsurance under VaR and CTE risk measures. *Proc. of Rom. Acad. Ser. A* **11** (2010) 213–217.
[16] M. Denuit, J. Dhaene, M. Goovaerts and R. Kaas, *Actuarial Theory for Dependent Risks*. Wiley, (2005).
[17] P. Embrechts and G. Puccetti, Bounds for functions of multivariate risks. *J. Multivar. Anal.* **97** (2006) 526–547.
[18] J.M. Fernández-Ponce and A. Suárez-Lloréns, Central regions for bivariate distributions. *Austrian J. Stat.* **31** (2002) 141–156.
[19] E.W. Frees and E.A. Valdez, Understanding relationships using copulas. *North Amer. Actuar. J.* **2** (1998) 1–25.
[20] J.A. Hartigan, Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** (1987) 267–270.
[21] V.I. Koltchinskii, M-estimation, convexity and quantiles. *Ann. Statist.* **25** (1997) 435–477.
[22] T. Laloë, *Sur Quelques Problèmes d'Apprentissage Supervisé et Non Supervisé*. Ph.D. thesis, University Montpellier II (2009).
[23] J.-C. Massé and R. Theodorescu, Halfplane trimming for bivariate distributions. *J. Multivar. Anal.* **48** (1994) 188–202.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press

[24] G. Nappo and F. Spizzichino, Kendall distributions and level sets in bivariate exchangeable survival models. *Inform. Sci.* **179** (2009) 2878–2890.
[25] W. Polonik, Measuring mass concentrations and estimating density contour clusters – an excess mass approach. *Ann. Stat.* **23** (1995) 855–881.
[26] W. Polonik, Minimum volume sets and generalized quantile processes. *Stoch. Proc. Appl.* **69** (1997) 1–24.
[27] P. Rigollet and R. Vert, Optimal rates for plug-in estimators of density level sets. *Bernoulli.* **15** (2009) 1154–1178.
[28] A. Rodríguez–Casal. *Estimacíon de conjuntos y sus fronteras. Un enfoque geometrico*. Ph.D. thesis, University of Santiago de Compostela (2003).
[29] C. Rossi, Sulle curve di livello di una superficie di ripartizione in due variabili; on level curves of two dimensional distribution function. *Giornale dell'Istituto Italiano degli Attuari* **36** (1973) 87–108.
[30] C. Rossi, Proprietà geometriche delle superficie di ripartizione. *Rend. Mat. (6)* **9** (1976) 725–736 (1977).
[31] R. Serfling, Quantile functions for multivariate analysis: approaches and applications. *Stat. Neerlandica* **56** (2002) 214–232 Special issue: Frontier research in theoretical statistics (2000) (Eindhoven).
[32] L. Tibiletti, On a new notion of multidimensional quantile. *Metron* **51** (1993) 77–83.
[33] A.B. Tsybakov, On nonparametric estimation of density level sets. *Ann. Stat.* **25** (1997) 948–969.

https://doi.org/10.1051/ps/2011161 Published online by Cambridge University Press