

Estimation of extreme multivariate expectiles with functional covariates

Elena Di Bernardino ^{*,1}, Thomas Laloë¹, Cambyse Pakzad¹

Université Côte d'Azur, Laboratoire J.A. Dieudonné, UMR CNRS 7351, Nice, France

ARTICLE INFO

AMS 2020 subject classifications:

60G70

62H12

90C53

91G70

Keywords:

Dependence

Extreme value theory

Multivariate Expectiles

Multivariate Regular Variation

Multivariate risk measures

Optimization

ABSTRACT

The present article is devoted to the semi-parametric estimation of multivariate expectiles for extreme levels. The considered multivariate risk measures also include the possible conditioning with respect to a functional covariate, belonging to an infinite-dimensional space. By using the first order optimality condition, we interpret these expectiles as solutions of a multidimensional nonlinear optimum problem. Then the inference is based on a minimization algorithm of gradient descent type, coupled with consistent kernel estimations of our key statistical quantities such as conditional quantiles, conditional tail index and conditional tail dependence functions. The method is valid for equivalently heavy-tailed marginals and under a multivariate regular variation condition on the underlying unknown random vector with arbitrary dependence structure. Our main result establishes the consistency in probability of the optimum approximated solution vectors with a speed rate. This allows us to estimate the global computational cost of the whole procedure according to the data sample size. The finite-sample performance of our methodology is provided via a numerical illustration of simulated datasets.

1. Introduction

Risk measurement theory is an active branch of research with numerous applications in the fields of finance, insurance, economics and for the environment (hydrology, geology, ...). To study the extreme risk of a random phenomena, e.g., of a big loss on a financial position, the most popular way is to estimate quantiles, also known as Value-at-Risk (VaR), at high level. It has been argued that is an incoherent risk measure in view of Artzner et al. [2] (see also Acerbi [1]). Another drawback is that it relies on the frequency of tail events and not on their real magnitudes which is precisely what one would like to know. On the other hand, the second most famous risk measure, the expected shortfall, is not elicitable in the sense of Gneiting [25], meaning that it is not defined as the minimization of the expectation of some score function. The univariate expectiles are then introduced by Newey and Powell [35] and turn out to be the only law invariant risk measures which are elicitable by construction and coherent for a threshold level range, see Bellini et al. [5], Ziegel [43]. Concerning the univariate expectile, we also refer to Bellini and Bernardino [4], Daouia et al. [9,10]. In economic terms, they may be interpreted as ratios of expected gain/loss which found to be recognized in portfolio management, see Bellini and Bernardino [4], and represent the quantity of money to inject in a position to reach a prescribed ratio gain/loss.

A fundamental question for potential practice is the statistical estimation in the extreme regime of the risk measures at our disposal. In this context, the extreme regime is modeled by a risk level tending towards zero or one, and an assumption of heavy-tails

* Corresponding author.

E-mail address: Elena.Di.bernardino@unice.fr (E. Di Bernardino).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.jmva.2023.105292>

Received 14 April 2023; Received in revised form 18 December 2023; Accepted 20 December 2023

Available online 23 December 2023

0047-259X/© 2023 Elsevier Inc. All rights reserved.

on the underlying distributions, typically of Pareto-type, which best captures rare phenomena. The proper mathematical framework is the regular variation notion. Besides, it often happens in practical applications that the observations are recorded along with auxiliary information represented by a random covariate. Then, one would preferably take advantage of the extra information by focusing on the conditional extremes. This line of research was carried by Girard et al. [22,24] when the covariate is a random vector. A valuable improvement is Girard et al. [23] for univariate extreme expectiles when the covariate belongs to an infinite-dimensional or functional space. Furthermore, in several situations, one needs to simultaneously manage risks over different positions, requiring a multivariate version of risk measures which would take the underlying dependence structure into account. Recently, several possible multivariate extensions of expectiles emerged in the literature such as geometric expectiles in Herrmann et al. [27] and L^p -expectiles in Maume-Deschamps et al. [33]. In the present work, we focus on the specific case of the multivariate L^1 -expectiles of Maume-Deschamps et al. [33] in the extreme regime, which we abbreviate with the notation MEEs for Multivariate Extreme Expectiles. Herein, in Maume-Deschamps et al. [33] the authors construct Σ -expectiles as another extension possibility based on correlation matrices Σ that reduces to the L^1 -version if $\Sigma \equiv \mathbf{1}$. To estimate multivariate expectiles, they exploit in Maume-Deschamps et al. [33] the elicibility by focusing on the first order optimality condition, namely MEEs are points in the d -dimensional Euclidean space for which the score gradient vanishes. The latter only involving tail expectations through positive and negative parts, the authors achieve the estimation by means of Robbins–Monro’s stochastic approximation method for moderate levels of risk. Conversely, in the extreme regime, the same authors use in Maume-Deschamps et al. [34] classic regular variation tools to express MEEs ratios as solutions of a system of coupled nonlinear equations. In addition, an estimation procedure is given for L^1 -MEEs with equivalent regularly varying marginal tails, provided the tail dependence is either comonotonic or asymptotically independent. Then, the approach in Maume-Deschamps et al. [34] works well for specific dependence structures.

In this paper, we address the estimation of the L^1 -MEEs when a conditional covariate lying in a possibly infinite-dimensional space is available. We assume the equivalent regularly varying marginal tails hypothesis and that the underlying dependence structure and the marginal distributions are unknown. Again, the first order optimality condition yields a system of equations for which functional MEEs ratios are solutions. Equivalently, functional MEEs may be seen as roots of a certain loss function which can be turned into an optimization problem. Crucial quantities are involved regarding the tail behavior such as the conditional tail index and the conditional tail dependence function. Inspired by Beck et al. [3], plugging their empirical counterparts in the optimization problem results in the approximated loss function. Finally, we propose to apply a BFGS-gradient method. In our main result (see Theorem 1), we prove the consistency in probability with rate of the approximated loss function and of the associated optimum solution in this conditional functional setting. Contrarily to Beck et al. [3], in the present work, a special attention is devoted to the rates of convergence in the approximation of the underlying optimization problem. It allows to link the loss function approximation quality with the steps of the used gradient descent algorithm. As a result, we explicitly provide the speed rate at which the approximated optimum converges to its theoretical value, which essentially quantifies the estimation quality according to the sample size.

The paper is organized as follows. In Section 2, we present some necessary notation and the setting of our model. We introduce in Section 3 the formal definition of functional multivariate L^1 -expectiles and how they can be theoretically related to a non-linear optimum problem. We subsequently develop in Section 4 the statistical tools of the present paper. After introducing the different estimators, we construct the associated approximated optimum problem. In Section 5 we present the required hypotheses and we state our main result (see Theorem 1) about the convergence with rate of the approximated optimum problem to the theoretical one. We outline the sketch of the proof of our main result and we give a crucial intermediate convergence result on the loss function and its gradient is also given (see Proposition 1). We devote the last part of Section 5 to a discussion about the required hypotheses. In Section 6, we study the performance of our estimation procedure via numerical simulations. Section 7 is devoted to the proof of the main result. Auxiliary proofs and supplementary lemmas are postponed to Section 8. Further material about the second order regular variation condition is provided in Appendix.

2. Notation and preliminaries

In this work, we consider a Polish space $(E, \|\cdot\|_E)$ endowed with its Borel σ -algebra and a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $2 \leq d < +\infty$ and a random pair $(X, Y) \in \mathbb{R}^d \times E$, with $X = (X_j)_{1 \leq j \leq d}$, defined on Ω such that $X \in (L^1(\Omega))^d$. On account of [29, Theorem 3.2], such topological features on E ensures the existence of regular conditional probabilities $\mathbb{P}_y(\cdot) := \mathbb{P}(\cdot|Y = y)$ on \mathcal{A} for \mathbb{P}_Y -almost all $y \in E$, where $\mathbb{P}_Y = \mathbb{P} \circ Y^{-1}$ is the pushforward measure. As a result, we may denote the conditional cumulative distribution function (cdf) of X given $Y = y$ by $F_{X,y}(\mathbf{x}) := \mathbb{P}_y(\{X \leq \mathbf{x}\}) = \mathbb{P}(X \leq \mathbf{x}|Y = y)$, for $\mathbf{x} \in \mathbb{R}^d$ and $y \in E$, where the inequality is to be understood component-wise on \mathbb{R}^d .

We denote the conditional marginal distributions, which we suppose to be continuous, by

$$x_j \mapsto F_{j,y}(x_j) := F_{X,y}(+\infty, \dots, x_j, +\infty, \dots), \quad 1 \leq j \leq d,$$

and for $1 \leq j < k \leq d$, the conditional bivariate marginal distributions by

$$(x_j, x_k) \mapsto F_{j,k,y}(x_j, x_k) := F_{X,y}(+\infty, \dots, x_j, +\infty, \dots, x_k, +\infty, \dots). \tag{1}$$

We also introduce the conditional marginal quantile function of X_j given $Y = y$ and its conditional marginal quantile tail function by, respectively,

$$q_{j,y}(\alpha) := \inf \{x > 0, F_{j,y}(x) \geq \alpha\}, \quad \alpha \in (0, 1), \tag{2}$$

$$U_{j,y}(x) := q_{j,y}(1 - x^{-1}) = \inf \{t \in \mathbb{R}, F_{j,y}(t) \geq 1 - x^{-1}\}, \quad x > 0.$$

Under the hypothesis of continuity of conditional marginal distributions and by Sklar’s Theorem (1959) (see Sklar [40]), there exists an unique copula C_y such that, for $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$,

$$C_y(\mathbf{u}) = F_{X,y}\left(F_{1,y}^{-1}(u_1), \dots, F_{d,y}^{-1}(u_d)\right) = \mathbb{P}\left(\bigcap_{j=1}^d \{F_{j,y}(X_j) \leq u_j\} \mid Y = y\right),$$

for which we assume the regularity condition from Gijbels et al. [21], Segers [39]; namely, for each $j \in \{1, \dots, d\}$,

$$\mathbf{u} \mapsto \frac{\partial C_y}{\partial u_j}(\mathbf{u}) \text{ exists and is continuous on the set } \{\mathbf{u} \in [0, 1]^d : 0 < u_j < 1\}. \tag{3}$$

For $y \in E$ and $\mathbf{x} \in \mathbb{R}_+^d$, we define the conditional stable tail dependence function

$$L_y(\mathbf{x}) := \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(\bigcup_{j=1}^d \{1 - F_{j,y}(X_j) \leq t x_j\} \mid Y = y\right) = \lim_{t \downarrow 0} t^{-1} (1 - C_y(\mathbf{1} - t \mathbf{x})). \tag{4}$$

For any $y \in E$ and $\mathbf{x} \in [0, +\infty]^d \setminus \{(+\infty, \dots, +\infty)\}$, the associated conditional upper tail dependence function is given by

$$\lambda_y(\mathbf{x}) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(\bigcap_{j=1}^d \{1 - F_{j,y}(X_j) \leq t x_j\} \mid Y = y\right) = \|\mathbf{x}\|_1 - L_y(\mathbf{x}), \tag{5}$$

where $\|\cdot\|_1$ denotes the L^1 norm. Let $1 \leq j < k \leq d$, $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ and $\mathbf{u}^{j,k} = (\mathbf{1}_{\ell \notin \{j,k\}} + u_\ell \mathbf{1}_{\ell \in \{j,k\}})_{1 \leq \ell \leq d}$. By using (5) and for $\mathbf{x} \in \mathbb{R}_+^2$, we introduce the bivariate restrictions of the conditional copula and tail dependence functions as following

$$C_{j,k,y}(u_j, u_k) := C_y(\mathbf{u}^{j,k}) = \mathbb{P}\left(F_{j,y}(X_j) \leq u_j, F_{k,y}(X_k) \leq u_k \mid Y = y\right),$$

$$L_{j,k,y}(\mathbf{x}) := \lim_{t \downarrow 0} t^{-1} (1 - C_{j,k,y}(\mathbf{1} - t \mathbf{x})), \quad \lambda_{j,k,y}(\mathbf{x}) = \|\mathbf{x}\|_1 - L_{j,k,y}(\mathbf{x}) = \lim_{t \downarrow 0} t^{-1} \bar{C}_{j,k,y}(t \mathbf{x}),$$

where $\bar{C}_{j,k,y}$ is the survival copula defined by $\bar{C}_{j,k,y}(u, v) = u + v - 1 + C_{j,k,y}(1 - u, 1 - v)$ on $[0, 1]^2$. In the subsequent paper, we suppose that each bivariate tail dependence function is continuous. Furthermore, we quantify the convergence rate in (4), akin to Einmahl et al. [14], Schmidt and Stadtmüller [38], by assuming that for any $1 \leq j \neq k \leq d$, there exists $\mu_{j,k,y} > 0$ such that, as $t \downarrow 0$, and for any $T > 0$,

$$\sup_{\mathbf{x} \in [0, T]^d} \left| t^{-1} (1 - C_{j,k,y}(\mathbf{1} - t \mathbf{x})) - L_{j,k,y}(\mathbf{x}) \right| = O(t^{\mu_{j,k,y}}).$$

In particular, defining $\mu_y := \max_{1 \leq j \neq k \leq d} \{\mu_{j,k,y}\}$, we may write

$$\max_{1 \leq j \neq k \leq d} \sup_{\mathbf{x} \in [0, T]^d} \left| t^{-1} (1 - C_{j,k,y}(\mathbf{1} - t \mathbf{x})) - L_{j,k,y}(\mathbf{x}) \right| = O(t^{\mu_y}). \tag{6}$$

We now state hypothesis on the considered random vector X given $Y = y$: it has equivalent regularly varying marginal tails. More precisely, as in Beck et al. [3], Maume-Deschamps et al. [34], we assume that each conditional marginal function behaves the same way in the extreme regime.

Assumption 1 (Considered Conditional Marginal Tails Model). For any $1 \leq j \leq d$ and $y \in E$, there exists $\gamma_y > 0$, $\rho_{j,y} \leq 0$, $c_{j,y} > 0$, such that

$$(H1) \quad \bar{F}_{j,y}(\cdot) = \mathbb{P}(X_j > \cdot \mid Y = y) \in 2RV_{-\frac{1}{\gamma_y}, \frac{\rho_{j,y}}{\gamma_y}}(+\infty),$$

$$(H2) \quad \lim_{x \rightarrow \infty} \frac{\bar{F}_{j,y}(x)}{\bar{F}_{1,y}(x)} = c_{j,y} < +\infty.$$

Assumptions (H1) and (H2) are classical in the extreme value literature. For a discussion about Assumptions (H1) and (H2) the reader is referred to Appendix. This equivalent tails model is often used to model claim amounts in insurance and study dependent extreme events, and is also used in ruin theory models. The equivalent tails model includes in particular portfolios with identically distributed risks, and the case where there is a scale difference in distributions. This setting is introduced to study the asymptotic behavior of the considered multivariate expectiles in the multivariate regular variation framework (see in particular Maume-Deschamps et al. [34] and (9) below).

While Assumption 1 is stated conditionally to y , it is worthwhile to mention that it does not hold unconditionally unless all quantities depending on y such as $\bar{F}_{j,y}$, γ_y and $\rho_{j,y}$ are constant with respect to y . In this specific case, Assumption 1 reduces to equivalent tails assumption in Maume-Deschamps et al. [34].

3. Functional MEEs and optimum system

The multivariate risk measure studied in the present paper is introduced in the following definition.

Definition 1 (Multivariate L^1 -Expectile with Functional Covariate). Let $y \in E$. We define the multivariate L^1 -expectile with functional covariate Y at risk level $\alpha \in (0, 1)$ as the random vector in \mathbb{R}^d , such that

$$e_\alpha(X, y) := \arg \min_{x \in \mathbb{R}^d} \mathbb{E} \left[\alpha \|(X - x)_+\|_1^2 + (1 - \alpha) \|(X - x)_-\|_1^2 \mid Y = y \right] \tag{7}$$

$$= \arg \min_{(x_1, \dots, x_d) \in \mathbb{R}^d} \mathbb{E} \left[\alpha \left(\sum_{j=1}^d (X_j - x_j)_+ \right)^2 + (1 - \alpha) \left(\sum_{j=1}^d (X_j - x_j)_- \right)^2 \mid Y = y \right].$$

Definition 1 can be seen as a functional conditional extension of the multivariate L^1 -expectiles recently introduced in Beck et al. [3] (for further details the interested reader is referred to Eq. (2) and Conclusion section in Beck et al. [3]). Observe that the equality in (7) is assured by the strict convexity of $s_\alpha(t, x) = \alpha \|(t - x)_+\|_1^2 + (1 - \alpha) \|(t - x)_-\|_1^2$ in the variable x which transfers to $\mathbb{E}(s_\alpha(X, x))$ as well by Jensen inequality. Obviously, the conditional L^1 -expectile in (7) is a random vector in \mathbb{R}^d and we write $e_\alpha(X, y) = (e_\alpha^j(X, y))_{1 \leq j \leq d}$. Furthermore, concerning the endpoint, one has $x_F = \sup\{x, F(x) < 1\} = +\infty$, so that $e_\alpha^j(X, y) \rightarrow \infty$, for $\alpha \rightarrow 1$ and $1 \leq j \leq d$.

Interestingly, an equivalent definition of the functional conditional L^1 -expectile in (7) can be given using an first order condition of optimality. Indeed the L^1 -expectile is the unique solution in \mathbb{R}^d of the following system $\alpha \mathbb{E} \left[\|(X - x)_+\|_1 \mathbf{1}_{\{X_k > x_k\}} \mid Y = y \right] = (1 - \alpha) \mathbb{E} \left[\|(X - x)_-\|_1 \mathbf{1}_{\{X_k < x_k\}} \mid Y = y \right]$, $k \in \{1, \dots, d\}$, or equivalently

$$\frac{\mathbb{E} \left[\|(X - x)_+\|_1 \mathbf{1}_{\{X_k > x_k\}} \mid Y = y \right]}{\mathbb{E} \left[\|(X - x)_-\|_1 \mathbf{1}_{\{X_k < x_k\}} \mid Y = y \right]} = \frac{1 - \alpha}{\alpha}, \quad k \in \{1, \dots, d\}, \tag{8}$$

which can be interpreted as a ratio of an activity participation in the conditional positive scenarios and its participation in the negative ones. System in (8) clarifies the relevance of L^1 -expectile in the multivariate risk analysis. The functional expectile $e_\alpha(X, y)$ represents the vector of marginal levels that make the ratio in (8) constant for all the marginal assets, given a covariate realization. It balances the risk level between the portfolio's components and may be seen as a capital allocation tool. The interested reader is referred to [33, Section 2.1] for more details.

From (8), and using marginal tails behavior and the tail dependence functions, one can get an appropriate expression of the optimum system (8) in the extreme regime (see Maume-Deschamps et al. [34] in the unconditional setting). As suggested by Beck et al. [3], we may convert without effort the work of Maume-Deschamps et al. [34], especially their Propositions 3.1, 3.3 and 5.1, to our functional setup. Therefore, suppose that the random vector X given $Y = y$ is regularly varying (MRV) (see Appendix) of index $1/\gamma_y$ with conditional marginal tails satisfying Conditions (H1) and (H2). Then, as $\alpha \rightarrow 1$, any limiting vector $(\eta, \beta_2, \dots, \beta_d)$ of

$\left(\frac{1 - \alpha}{F_{1,y}(e_\alpha^1(X, y))}, \frac{e_\alpha^2(X, y)}{e_\alpha^1(X, y)}, \dots, \frac{e_\alpha^d(X, y)}{e_\alpha^1(X, y)} \right)$, for $y \in E$, satisfies the system with $k \in \{1, \dots, d\}$,

$$\frac{1}{1/\gamma_y - 1} - \eta \frac{\beta_k^{1/\gamma_y}}{c_{k,y}} = - \sum_{j \neq k} \left(\int_{\frac{\beta_j}{\beta_k}}^{+\infty} \lambda_{j,k,y} \left(\frac{c_{j,y}}{c_{k,y}} t^{-1/\gamma_y}, 1 \right) dt - \eta \frac{\beta_k^{1/\gamma_y - 1}}{c_{k,y}} \beta_j \right), \tag{9}$$

(see (3.5) in Maume-Deschamps et al. [34] for a comparison). In the case where $\Theta := (\eta, \beta_2, \dots, \beta_d)$ is unique, by using the quantile in (2), as $\alpha \rightarrow 1$, we have

$$e_\alpha(X, y) \sim q_{1,y}(\alpha) \eta^{\gamma_y} (1, \beta_2, \dots, \beta_d)^\top.$$

For a comprehensive review of the different definitions available for the MRV property, we refer to Resnick et al. [37]. This assumption on $X|Y$ mainly ensures the existence of the conditional bivariate upper tail dependence functions which we suppose continuous, see Section 2.3 in Maume-Deschamps et al. [34]. Notice the system (9) only displays pairwise interactions through the conditional bivariate tail dependence function. In view of (9), we introduce the associated optimum problem.

Definition 2 (Loss Function for the Optimum Problem of $e_\alpha(X, y)$). Let $y \in E$. We consider $\Theta := (\eta, \beta_2, \dots, \beta_d) > 0$ as (9). Consider also

$$\xi_y = (\gamma_y, c_{2,y}, \dots, c_{d,y}, \lambda_y) \tag{10}$$

with λ_y as in (5), γ_y as in Condition (H1) and $c_{j,y}$ as in Condition (H2). Define the vector $\varphi_y(\Theta, \xi_y) = (\varphi_{k,y}(\Theta, \xi_y))_{1 \leq k \leq d}$ where

$$\varphi_{k,y}(\Theta, \xi_y) := \frac{\gamma_y}{1 - \gamma_y} - \eta \frac{\beta_k^{1/\gamma_y}}{c_{k,y}} \left(1 + \sum_{j \neq k} \frac{\beta_j}{\beta_k} \right) + \sum_{j \neq k} \int_{\frac{\beta_j}{\beta_k}}^{+\infty} \lambda_{j,k,y} \left(\frac{c_{j,y}}{c_{k,y}} t^{-1/\gamma_y}, 1 \right) dt. \tag{11}$$

With these notations, one can formulate the system in (9) as $\varphi_y(\Theta, \xi_y) = \mathbf{0}$ (the equality being understood in \mathbb{R}^d). We introduce the loss function via the squared Euclidean norm

$$\mathcal{L}_{\xi_y}(\Theta) := \frac{1}{2} \|\varphi_y(\Theta, \xi_y)\|_2^2 = \frac{1}{2} \sum_{k=1}^d \varphi_{k,y}(\Theta, \xi_y)^2, \tag{12}$$

and an optimal vector Θ_y^* obtained by minimizing the loss function \mathcal{L}_{ξ_y} in (12),

$$\Theta_y^* \in \arg \min_{\Theta} \mathcal{L}_{\xi_y}(\Theta). \tag{13}$$

To minimize the optimum problem (13), we turn to the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm which belongs to the family of quasi-Newton optimization methods. This choice is motivated by practicality while, from a rigorous point of view, its global convergence is denied due to the loss function lacking of convexity property. A particular interest comes from the fact that no second derivatives computations are needed. In fact, we will rather consider the L-BFGS-B method, a refinement of the classic BFGS that incorporates bound constraints.

Since the underlying distribution is not known in practice, direct application of any minimization algorithm on the optimization problem (12)–(13) is not feasible. The next section is devoted to overcome this drawback.

4. Approximated optimum problem

Let (X_i, Y_i) for $1 \leq i \leq n$, with $X_i = (X_{ij})_{1 \leq j \leq d}$, be n independent copies of $(X, Y) \in \mathbb{R}^d \times E$. In the following the convergence in probability and in distribution, as n goes to infinity, are respectively denoted by $\xrightarrow{\mathbb{P}}$ and \xrightarrow{d} . Relying on this data sample, we consider an approximated version of the loss function for the optimum problem of $e_\alpha(X, y)$ previously introduced in Definition 2.

Definition 3 (Approximated Loss Function of $e_\alpha(X, y)$). By using (12) the approximated loss function can be written as $\mathcal{L}_{\hat{\xi}_{n,y}}(\Theta) = \frac{1}{2} \|\varphi_y(\Theta, \hat{\xi}_{n,y})\|_2^2$, where $\hat{\xi}_{n,y}$ is an estimator vector of ξ_y in (10). Furthermore, we define the associated optimum problem as

$$\hat{\Theta}_y^* \in \arg \min_{\Theta} \mathcal{L}_{\hat{\xi}_{n,y}}(\Theta). \tag{14}$$

We now aim to consistently estimate the loss function $\mathcal{L}_{\xi_y}(\Theta)$. To this purpose, we firstly build a consistent estimator vector $\hat{\xi}_{n,y} = (\hat{\gamma}_{n,y}, \hat{c}_{2,n,y}, \dots, \hat{c}_{d,n,y}, \hat{\lambda}_y)$ of ξ_y . According to (9), we need to estimate $\lambda_{j,k,y}$ (see Section 4.1 below) and $\gamma_y, c_y = (c_{2,y}, \dots, c_{d,y})$ (see Section 4.2). The interested reader is referred to the framework in Gijbels et al. [21], Girard et al. [23] (see also Gardes and Girard [19]).

4.1. Empirical counterpart for $\lambda_{j,k,y}$

Let $1 \leq j < k \leq d$. We consider the empirical counterpart of the bivariate conditional cdf $F_{j,k,y}$ in (1) (see, e.g., Daouia et al. [7], Ferraty et al. [16], Girard et al. [23])

$$\hat{F}_{j,k,n,y}(x_j, x_k) := \sum_{i=1}^n w_{i,n} \mathbf{1}_{\{X_{ij} \leq x_j, X_{ik} \leq x_k\}}, \quad x_j, x_k \in \mathbb{R}, \tag{15}$$

where the Nadaraya–Watson weights are given by

$$w_{i,n} := w_{i,n}(Y, y, h_n) = \frac{K\left(\frac{\|Y_i - y\|_E}{h_n}\right)}{\sum_{s=1}^n K\left(\frac{\|Y_s - y\|_E}{h_n}\right)}, \quad \sum_{i=1}^n w_{i,n} = 1, \tag{16}$$

with $K : \mathbb{R} \rightarrow \mathbb{R}^+$ a kernel function, i.e., positive and measurable, having support in $[0, 1]$.

We define the empirical estimator of the conditional copula and its bivariate restriction

$$\hat{C}_{n,y}(\mathbf{u}) := \sum_{i=1}^n w_{i,n} \mathbf{1}_{\{\bigcap_{j=1}^d \{\hat{F}_{j,n,y}(X_{ij}) \leq u_j\}\}}, \quad \hat{C}_{j,k,n,y}(u_j, u_k) := \hat{C}_{n,y}(\mathbf{u}^{j,k}) = \sum_{i=1}^n w_{i,n} \mathbf{1}_{\{\hat{F}_{j,n,y}(X_{ij}) \leq u_j, \hat{F}_{k,n,y}(X_{ik}) \leq u_k\}}, \tag{17}$$

where $\hat{F}_{j,n,y}, \hat{F}_{k,n,y}$ are the corresponding marginal distribution functions of $\hat{F}_{j,k,n,y}$ in (15). Furthermore, for $\mathbf{x} \in \mathbb{R}_+^2$ and k_n a deterministic intermediate sequence, i.e., $1 \ll k_n \ll n$, we consider the empirical estimators of the bivariate conditional stable tail dependence function and of the associated upper tail dependence function:

$$\hat{L}_{j,k,n,y}(\mathbf{x}) := \frac{n}{k_n} \left(1 - \hat{C}_{j,k,n,y}\left(\mathbf{1} - \frac{k_n}{n} \mathbf{x}\right)\right), \quad \hat{\lambda}_{j,k,n,y}(\mathbf{x}) = \|\mathbf{x}\|_1 - \hat{L}_{j,k,n,y}(\mathbf{x}).$$

4.2. Empirical counterparts for γ_y and c_y

Let $y \in E$. Assume $h_n \ll 1$ be a non-random positive sequence such that the small ball probability function is positive, i.e.,

$$\psi_y(h_n) := \mathbb{P}(\|Y - y\|_E \leq h_n) > 0. \tag{18}$$

The latter quantity has a major impact in the functional estimation framework as one can see for instance in our hypotheses in Section 5.1 below. Note that it heavily relies on the intrinsic metric on E .

Our plug-in estimation relies on functional estimators of the marginal quantile and tail index previously introduced by Gardes and Girard [19], Girard et al. [23]. Let us then define $\hat{q}_{j,y,n}(\alpha) := \inf \{x > 0, \hat{F}_{j,y,n}(x) \geq \alpha\}$, $\alpha \in [0, 1]$, the natural empirical counterpart of the conditional quantile function in (2), where $\hat{F}_{j,y,n}$ is the marginal distribution function associated to (15).

We pick an integer $1 \leq J < +\infty$ and a subdivision $0 < \tau_J < \tau_{J-1} < \dots < \tau_1 < \dots < \tau_1 \leq 1$, e.g. $\tau_i = 1/i$ and $J = 9$ are motivated by the discussion after Corollary 2 in Daouia et al. [8]. The tail index being the same over all margins, we restrict our attention to $j = 1$ and consider the so-called functional Hill estimator, first introduced by Gardes and Girard [19],

$$\hat{\gamma}_{n,y} := \frac{\sum_{i=1}^J \ln \hat{q}_{1,n,y}(1 - \tau_i \bar{\alpha}_n) - \ln \hat{q}_{1,n,y}(\alpha_n)}{-\sum_{i=1}^J \ln(\tau_i)}, \tag{19}$$

where $\bar{\alpha}_n = 1 - \alpha_n$. Concerning the tail ratio $c_{j,y}$ in Condition (H2), we simply extend in a functional framework the estimator proposed in Beck et al. [3], Maume-Deschamps et al. [34]

$$\hat{c}_{j,n,y} := \left(\frac{\hat{q}_{j,n,y}(\alpha_n)}{\hat{q}_{1,n,y}(\alpha_n)} \right)^{1/\hat{\gamma}_{n,y}}, \quad j \in \{2, \dots, d\}. \tag{20}$$

By using the inference procedure presented in Sections 4.1 and 4.2, we now focus on the convergence of the approximated optimum problem in Eq. (14).

5. Convergence of the approximated optimum problem for functional MEEs

In Section 5.1 below, we introduce conditions required in our main convergence result, i.e., Proposition 1. These conditions fall into porous categories. Because of the bias, the scaling sequence α_n and the smoothing parameter h_n are linked through different regimes involving in particular the small ball probability $\psi_y(h_n)$.

5.1. Required assumptions

We organize the presentation of our assumptions by separating them in three classes (I, II and III). The first class below is related to the bandwidth h_n for the small ball probability in (18), the risk level α_n and the marginal distribution function of X_1 .

(I.1) The bandwidth $h := h_n \ll 1$ satisfies the intermediate regime $n \bar{\alpha}_n \psi_y(h_n) \gg 1$.

(I.2) For any $1 \leq j \leq d$, there exists $\delta_j > 0$ such that for any $y \in E$,

$$\sqrt{n \bar{\alpha}_n \psi_y(h_n) \ln(\bar{\alpha}_n)} \sup_{\substack{z \geq (1-\delta_j) \eta_{j,y}(\alpha_n) \\ y' \in B(y, h_n)}} \frac{1}{\ln z} \left| \ln \frac{\bar{F}_{1,y'}(z)}{\bar{F}_{1,y}(z)} \right| \xrightarrow{n \rightarrow +\infty} 0.$$

(I.3) For any $y \in E$, it holds that $\sqrt{n \bar{\alpha}_n \psi_y(h_n) A_{1,y}(\bar{\alpha}_n^{-1})} \xrightarrow{n \rightarrow +\infty} p_{1,y} \in \mathbb{R}$, where $A_{1,y}$ is the auxiliary function of X_1 from Condition (H1) (see also Eq. (A.1)).

(I.4) For any $s \in [0, 1]$ and $0 < h_n \ll 1$ deterministic, the following pointwise convergence holds

$$\tau_{y,h_n}(s) := \frac{\psi_y(h_n s)}{\psi_y(h_n)} = \mathbb{P} \left(\|Y - y\|_E < s h_n \mid \|Y - y\|_E < h_n \right) \xrightarrow{h_n \rightarrow 0_+} \tau_{y,0}(s).$$

A second class of hypothesis involves the conditional bivariate distribution functions and the conditional bivariate upper tail dependence functions.

(II.1) For any $1 \leq j < k \leq d$ and $y \in E$, there exists $\eta_{j,k,y} > 0$ such that, for n large enough, as $t \rightarrow 0$,

$$\lambda_{j,k,y}(t, 1) = O(t^{\eta_{j,k,y}}), \quad \hat{\lambda}_{j,k,n,y}(t, 1) = O_{\mathbb{P}}(t^{\eta_{j,k,y}}).$$

Furthermore, it holds that $\gamma_y < \eta_y := \min_{j \neq k} \eta_{j,k,y}$.

(II.2) For any $1 \leq j < k \leq d$ and $y \in E$, it holds that $\sup_{y' \in E, \|y'\| \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^2} |F_{j,k,y+\delta y'}(\mathbf{x}) - F_{j,k,y}(\mathbf{x})| \xrightarrow{\delta \rightarrow 0_+} 0$. This means that the functional $F_{j,k,y}$ is continuous at the point y uniformly in \mathbf{x} . In particular, conditional marginal distributions are supposed to be continuous in \mathbf{x} and y .

(II.3) For any $1 \leq j < k \leq d$, analogously as in Ferraty et al. [17] we consider

$$\phi_{j,k,x,y}^F(s) := \mathbb{E}_Y \left[F_{j,k,Y}(\mathbf{x}) - F_{j,k,y}(\mathbf{x}) \mid \|Y - y\|_E = s \right], \quad \mathbf{x} = (x_j, x_k),$$

quantifying the expected difference $F_{j,k,Y}(\mathbf{x}) - F_{j,k,y}(\mathbf{x})$ when Y is forced to be at a distance s from the point $y \in E$. Assume that there exists $\beta > 0$ such that the function $s \mapsto \phi_{j,k,x,y}^F(s)$ with $\mathbf{x} = (x_j, x_k)$ satisfies, uniformly in \mathbf{x} ,

$$\phi_{j,k,x,y}^F(s) = s^\beta v_{j,k}(\mathbf{x}) + o(s^\beta), \quad s \rightarrow 0_+,$$

for some continuous function $\mathbf{u} \mapsto v_{j,k}(F_{j,y}^{-1}(u_j), F_{k,y}^{-1}(u_k))$ on $[0, 1]^2$. Furthermore, the bandwidth $h := h_n \ll 1$ satisfies $n \psi_y(h_n) h_n^{2\beta} = O(1)$.

A third class hypothesis focuses on the Kernel K in (16).

(III.1) K has support $[0, 1]$, is bounded in $(0, 1)$, has a continuous derivative on $[0, 1)$ with $K'(s) \leq 0$ and

$$M_1 := K(1) - \int_0^1 K'(s)\tau_{y,0}(s)ds > 0,$$

where $\tau_{y,0}$ is the pointwise limit as $h \rightarrow 0_+$ of $\tau_{y,h}$ in (I.4).

A discussion on the previous required assumptions is postponed to Section 5.3. We are now ready to state and prove our main result.

5.2. Main result

We begin by enunciating the multiple hypotheses needed for our main result. It is then followed by the different speed rates associated to the estimation part of the optimum problem and to the actual minimization algorithm (BFGS family). These scaling sequences are expressed in terms of copula estimation, small ball probability and regular variation.

Let a covariate value $y \in E$. Regarding the inference, we require that:

- the random vector $\mathbf{X} = (X_1, \dots, X_d)^\top$ given $Y = y$ is multivariate regularly varying with index $1/\gamma_y$ (see Appendix) and the conditional marginal tails satisfy Conditions (H1) and (H2).
- The condition in (3) and the first order copula condition in (6).
- Assumptions of classes I, II and III in Section 5.1.

Next, denote by $\hat{\Theta}_y^{(m)}$ the m th iteration step of the minimization algorithm applied to the approximated optimum problem (14). We make the following assumptions on the optimization procedure:

- for any $y \in E$, there exists a hyperrectangle $\mathcal{K} := \prod_{r=1}^d [\varepsilon_r, M_r] \subset (0, +\infty)^d$ such that for any $m \geq 1$ and $n \geq 1$ large enough, $\hat{\Theta}_y^{(m)} \in \mathcal{K}$.
- The minimization algorithm solves for the global minimum of optimum problem (13) and has complexity $O(\delta_{(m)})$, i.e., the computational cost of m steps of the algorithm is proportional to $\delta_{(m)}$.

Concerning the speed rates, for any $1 \leq j \leq d$, any deterministic $\alpha_n = 1 + o(1)$, $1 \ll k_n \ll n$ and $m_n \gg 1$, we define the following deterministic sequences, with $\mu_y > 0$ as in (6),

$$\bar{\delta}_{-1,n,y} := \left(\frac{n}{k_n}\right)^{\mu_y}, \quad \bar{\delta}_{0,n,y} := \sqrt{n\bar{\alpha}_n\psi_y(h)}, \quad \bar{\delta}_{j,n,y} := A_{j,y}^{-1}(\bar{\alpha}_n^{-1}). \tag{21}$$

Next, we introduce the speed rate $\delta_{n,y}$ as a scaling sequence such that

$$1 \ll \delta_{n,y} \ll \min_{-1 \leq j \leq d} \bar{\delta}_{j,n,y}. \tag{22}$$

We may now state our main result under the previous assumptions.

Theorem 1 (Approximated optimum convergence with rate). *Let $y \in E$. Let Θ_y^* the optimum of (13) and $\hat{\Theta}_y^{(m_n)}$ the m_n^{th} iteration step in the minimization algorithm applied to (14) where $m_n \gg 1$ is deterministic. Let also two scaling sequences $\delta_{n,y}$ and $\delta_{(m_n)}$. Then, under the previous hypotheses,*

$$\min(\delta_{n,y}, \delta_{(m_n)}) \left\| \hat{\Theta}_y^{(m_n)} - \Theta_y^* \right\|_1 \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

The proof of Theorem 1 is postponed below.

In order to estimate functional MEEs, we draw our attention on (9). A direct consequence is that, if unique, the limit $\Theta = (\eta, \beta_2, \dots, \beta_d)$ of $(\frac{1-\alpha}{F_{1,y}(e_\alpha^1(X,y))}, \frac{e_\alpha^2(X,y)}{e_\alpha^1(X,y)}, \dots, \frac{e_\alpha^d(X,y)}{e_\alpha^1(X,y)})$, as $\alpha \rightarrow 1$, satisfies the optimum problem (13) and

$$e_\alpha(X, y) \sim F_{1,y}^{-1}(\alpha)\eta^{\gamma_y} (1, \beta_2, \dots, \beta_d)^\top, \quad \alpha \rightarrow 1. \tag{23}$$

Notice that the optimum convergence in Beck et al. [3] is given via a non-interchangeable two limits result. The first concerns the convergence of the statistical approximation of the loss function, the second the convergence of the descent gradient algorithm. The interested reader is referred to Algorithm 2 and Corollary 3 in Beck et al. [3].

Conversely, our result unifies both limits $n, m \rightarrow +\infty$ into a single limit, i.e.,

$$\hat{\Theta}_y^{(m_n)} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \Theta_y^*. \tag{24}$$

The speed rate $\delta_{n,y}$ in Theorem 1, is associated to the statistical approximation of the loss function. The second rate $\delta_{(m)}$ is inherited of the used minimization algorithm. Considering the double limit $m := m_n \gg 1$, the global speed rate of (24) naturally involves a compromise between the two scaling sequences. For sake of clarity, in Remark 1 below, we present some more practical expressions on the speed rate $\delta_{n,y}$.

Remark 1 (About the Rate $\delta_{n,y}$). Under List (I.3), the condition on the scaling sequence $\delta_{n,y}$ in (22) becomes

$$1 \ll \delta_{n,y} \ll \min_{\substack{-1 \leq j \leq d \\ j \neq -1}} \tilde{\delta}_{j,n,y}. \tag{25}$$

Notice that explicit expressions for the auxiliary functions $A_{k,y}$ are available (e.g., Mao and Hu [30]) for particular distributions. For instance, in Section 6, we consider the Lomax (or Pareto II) conditional marginal distributions, which we generically denote by $X_j|Y = y \sim \text{Lomax}(1/\gamma_y, s_j)$ where $\gamma_y, s_j > 0$ for $y \in E$ and $1 \leq j \leq d$. For this model, the survival conditional distribution function is $\bar{F}_{j,y}(x) = (1 + x/s_j)^{-1/\gamma_y} \in 2\text{RV}_{-1/\gamma_y, -1}$, or equivalently, the conditional tail quantile function is $U_{j,y}(t) = s_j(t^{\gamma_y} - 1) \in 2\text{RV}_{\gamma_y, -\gamma_y}$ with auxiliary function $A_y(t) = \gamma_y t^{-\gamma_y}$. In this case, the speed rate in (25) has a simpler expression

$$1 \ll \delta_{n,y} \ll \min \{ \delta_{-1,n,y}, \delta_{0,n,y}, \gamma_y(1 - \alpha_n)^{-\gamma_y} \}.$$

Further, if one considers $k_n = n\psi_y(h_n)$ satisfying $1 \ll k_n \ll n$, one may write $\min(\delta_{0,n,y}, \delta_{-1,n,y}) = (n\psi_y(h_n))^{\min(\frac{1}{2}, \mu_y)}$ and the relation

$$1 \ll \delta_{n,y} \ll \min \left(\gamma_y(1 - \alpha_n)^{-\gamma_y}, (n\psi_y(h_n))^{\min(\frac{1}{2}, \mu_y)} \right).$$

Another possibility is to suppress the auxiliary functions dependence by considering $\delta_{0,n,y} = \delta_{-1,n,y}(1 + o(1))$ as $n \rightarrow +\infty$, or equivalently $k_n \sim n(n\psi_y(h_n))^{-1/(2\mu_y)}$. This particular choice of k_n gives under the condition $1 \ll k_n \ll n$ the new regime $\psi_y(h_n)^{1/(2\mu_y)} \ll n^{1-1/(2\mu_y)}$.

Remark 2 (About the Minimization Algorithm and Rate $\delta_{(m)}$). At the moment, we are essentially facing two possibilities for quasi-Newton methods (which have less complexity than the classic Newton ones). The first one is the classic BFGS method as described by Algorithm 6.1 in Nocedal and Wright [36] or Algorithm 1 in Beck et al. [3]. For such algorithm, each iteration costs $O(d^2)$ arithmetic operations. Hence, the whole computational cost is $O(d^2m)$ where m is the number of iterations, and we set $\delta_{(m)} = d^2m$. On the other hand, one may preferably use the L-BFGS-B version, which is designed for bound constrained optimization with limited memory storage. The motivations for this particular algorithm are threefold. First, it seems reasonable to consider MEEs only in a certain range. Second, no information about the Hessian matrix or about the structure of the loss function are required. Besides, when $m := m_n \gg 1$ and n is the number of observations, the upcoming proof of Theorem 1 requires uniform convergence results. The computational cost of one iteration in the L-BFGS-B algorithm is linear in d , namely $O(p^2d)$ with $3 \leq p \leq 20$ controlling the amount of storage (see Zhu et al. [42]). Thus, we associate the speed rate $\delta_{(m)} = p^2dm$ to this case.

To prove Proposition 1, we proceed in a two-step demonstration.

- (i) Firstly, we build the approximated minimization problem in (14) by considering an estimator vector $\hat{\xi}_{n,y} = (\hat{\gamma}_{n,y}, \hat{c}_{2,n,y}, \dots, \hat{c}_{d,n,y}, \hat{\lambda}_y)$. Once the consistency of $\hat{\xi}_{n,y}$ is assured, we prove uniform convergence with associated speed rate of the approximated loss function in Definition 3 and its gradient function to their theoretical counterparts.
- (ii) Then we link the minimization algorithm with the estimation procedure by inducing a dependency of the iteration steps on the data sample size n . Namely, we allow m to depend on n so that $m = m_n \gg 1$. This translates the fact that we are simultaneously consistently approximating the optimum problem (13) and solving the resulting approximated optimum problem (14) using the BFGS minimization algorithm.

Proof of Theorem 1. The proof of Theorem 1 heavily relies on the convergence of the loss function and its gradient. Moreover, the validity of step ii) can be proved if the gradient convergence is uniform. These results are gathered in the following intermediate proposition. The proof of Proposition 1 is postponed to Section 7.

Proposition 1 (Approximated loss function convergence with rate). Assume that assumptions of Theorem 1 hold true. Then,

$$\delta_{n,y} \left| \mathcal{L}_{\hat{\xi}_y}(\boldsymbol{\theta}) - \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \tag{26}$$

$$\delta_{n,y} \left\| \nabla \mathcal{L}_{\hat{\xi}_y}(\boldsymbol{\theta}) - \nabla \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta}) \right\|_1 \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \tag{27}$$

Furthermore if we restrict $\boldsymbol{\theta} \in \prod_{j=1}^d [\varepsilon_j, M_j]$ with $0 < \varepsilon_j < M_j < +\infty$ for any $1 \leq j \leq d$, we get

$$\delta_{n,y} \sup_{\boldsymbol{\theta} \in \prod_{j=1}^d [\varepsilon_j, M_j]} \left| \mathcal{L}_{\hat{\xi}_y}(\boldsymbol{\theta}) - \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \tag{28}$$

$$\delta_{n,y} \sup_{\boldsymbol{\theta} \in \prod_{j=1}^d [\varepsilon_j, M_j]} \left\| \nabla \mathcal{L}_{\hat{\xi}_y}(\boldsymbol{\theta}) - \nabla \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta}) \right\|_1 \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \tag{29}$$

Moreover, in Theorem 1, we assume that the minimization algorithm solves for the global minimum of (13). Once (28) and (29) are established (see also Eq. (37)), one may decompose with the triangular inequality

$$\left\| \hat{\boldsymbol{\theta}}_y^{(m_n)} - \boldsymbol{\theta}_y^* \right\|_1 \leq \left\| \hat{\boldsymbol{\theta}}_y^{(m_n)} - \boldsymbol{\theta}_y^{(m_n)} \right\|_1 + \left\| \boldsymbol{\theta}_y^{(m_n)} - \boldsymbol{\theta}_y^* \right\|_1.$$

The first quantity translates the cost of estimating the m_n th step of the minimization algorithm, while the second quantity is purely inherent to the nature of the used minimization algorithm. Concerning the minimization algorithm, we may consider any suitable gradient descent algorithm with complexity $\delta_{(m)}$, so the whole cost is

$$\left\| \boldsymbol{\theta}_y^{(m_n)} - \boldsymbol{\theta}_y^* \right\|_1 = O(\delta_{(m_n)}^{-1}).$$

By using the classical steps in the BFGS quasi-Newton optimization algorithm (see, e.g., Algorithm 1 in Beck et al. [3]), once the uniform convergence (29) holds, it immediately follows that, independently of m_n ,

$$\left\| \hat{\boldsymbol{\theta}}_y^{(m_n)} - \boldsymbol{\theta}_y^{(m_n)} \right\|_1 = O_{\mathbb{P}}\left(\delta_{n,y}^{-1}\right).$$

Hence the desired result. \square

5.3. Discussion of Hypotheses

Before diving into the details of hypotheses classes I,II and III, we begin this section with the basic requirements for our model. As noted by Hua and Joe [28], the second order condition in (H1) excludes the Pareto type I with cdf $F(x) = 1 - (k/x)^\alpha$, $x \geq k$, as possible marginal. Nonetheless, the Pareto type II, or Lomax, is admissible (for other examples, the reader is referred to Remark 5).

On the other hand, the regularity condition (3) on the dependence structure is mild and classic. The convergence rate requirement (6) is also frequent in the literature and, as noted in Di Bernardino et al. [12], holds for any $\mu_y \geq 1$ when C_y is twice continuously differentiable over $[0, 1]^2$. Besides and more specifically, Einmahl et al. [13] shows that (6) holds for any $\mu_y \in (0, \frac{d}{\beta})$ in the so-called parallel meta-elliptical model Fang et al. [15].

Conditions are also required on the small ball probabilities of $Y \in E$, which are in fact related to its concentration properties and the (semi)-metric Ferraty and Vieu [18]. Overall, this setting admits standard examples in the literature, typically inspired from wavelength modeling or finance, e.g., trigonometric random polynomials and Brownian motion-based processes.

Quantiles are the building blocks of the estimation procedure as they play a crucial role in the estimation of the tail index and ratios. If Conditions (H1), (III.1), (I.1), (I.2) are satisfied, one can prove that the functional estimator $\hat{q}_{k,n,y}(\alpha_n)$ of the extreme quantile $q_{k,y}(\alpha_n)$ is asymptotically Gaussian for $\alpha_n \rightarrow 1$ as $n \rightarrow \infty$ (see for instance Proposition 1 in Girard et al. [23] and Theorem 2 in Gardes and Girard [19]), i.e.,

$$\tilde{\delta}_{0,n,y} \left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{q_{k,y}(\alpha_n)} - 1 \right) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, 1), \quad 1 \leq k \leq d, \tag{30}$$

where $\tilde{\delta}_{0,n,y}$ is as in (21), provided that $n\bar{\alpha}_n\psi_y(h_n) \gg 1$ (see Condition (I.1)). In this particular situation, $q_{k,y}(\alpha_n)$ is referred to as an intermediate conditional quantile. This intermediate regime prevents the estimation at arbitrarily large levels and ensures that there will be a growing number of observations larger than the conditional quantile to be estimated (see also discussion below (32)). One can overcome this issue using extreme value extrapolation tools for intermediate functional conditional quantiles. The interested reader is referred for instance to Section 4 in Girard et al. [23].

If in addition (I.3) is fulfilled, [23, Theorem 4] extends the Gaussian fluctuations for the functional tail index (see also Theorem 4 in Gardes and Girard [19]),

$$\tilde{\delta}_{0,n,y} (\hat{\gamma}_{n,y} - \gamma_y) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}\left(\frac{p_{1,y}}{\ln(J!)} \sum_{j=2}^J \frac{j^{\rho_{1,y}} - 1}{\rho_{1,y}}, \frac{J(J-1)(2J-1)}{6 \log(J!)} \gamma^2(y) \right), \tag{31}$$

with $\tilde{\delta}_{0,n,y}$ as in (21), γ_y as in Condition (H1) with associated estimator $\hat{\gamma}_{n,y}$ as in (19), $\rho_{1,y}$ the second order tail index from Condition (H1), $p_{1,y} \in \mathbb{R}$ as in Condition (I.3) and J a positive integer (e.g., $J = 9$).

Concerning the fluctuations of the empirical copula in this functional setting, one may invoke [21, Theorem 3]. When each marginal $F_{j,y}$ is continuous and Assumptions (3), (I.1), (I.4), (II.2), (II.3), (III.1) are satisfied, it yields

$$\sqrt{n\psi_y(h_n)} \sup_{\mathbf{u} \in [0,1]^2} \left| \hat{C}_{j,k,n,y}(\mathbf{u}) - C_{j,k,y}(\mathbf{u}) \right| = O_{\mathbb{P}}(1), \tag{32}$$

where $\psi_y(h_n)$ is the small probability in (18), $C_{j,k,y}$ the conditional bivariate copula (of the margins j, k) with associated estimator $\hat{C}_{j,k,n,y}$ defined in (17). Concerning the kernel, condition (III.1) gathers requirements for our asymptotic tools (30)–(31)–(32) to hold. It may be decomposed in two parts. The first one may be qualified of type I according to [18, Definition 4.1] and concerns the univariate objects. While the second is about the copula and may first be found in Ferraty et al. [17]. A classic example of admissible kernel is $K(u) = \frac{1}{2}(1 - u^2)\mathbf{1}_{[0,1]}(u)$.

We devote the rest of this section to illustrate the role of the required hypothesis in our Theorem 1, in particular in order to guarantee the crucial convergences in (30)–(32).

- In view of [19, Lemma 4], the intermediate regime in (I.1) is a necessary and sufficient condition for the almost sure presence of at least one sample point in the region $B(y, h_n) \times (q_{j,y}(\alpha_n), +\infty)$ for any $1 \leq j \leq d$, that is, $q_{j,y}(\alpha_n)$ is within the sample. Therefore, it makes sure that, as the number of data sample points increases, the number of observations larger than the conditional quantile increases.

- The hypothesis (III.1) gathers conditions on the kernel for the simultaneous application of convergences in (30)–(31)–(32). It is hence a combination of Type I kernel, according to [18, Definition 4.1], required for (30)–(31), and of continuous kernel versions.
- Concerning univariate quantiles and tail index, we require Conditions (I.2) and (I.3) for applying (30)–(31) and proving Proposition 2 and Lemma 2, 4, 5. The first Condition (I.2) allows to control the oscillations of the survival distribution function when the covariate is varying. For a deeper discussion, we refer to the end of Section 2 in Girard et al. [23]. On the other hand, as well-known in the extreme literature, the Hill type estimators require a second order regular variation condition (see also Condition (H2)) in order to determine their rate convergence and fluctuations. The estimator in (19) of the tail index is built upon the first marginal empirical quantile without any second order related object involved. This lacking in the tail approximation generates a bias that we should assume via (I.3) to be negligible in the scale $\tilde{\delta}_{n,y,0}$. Inherited from Theorem 7 in Girard et al. [23], Condition (I.3) links the asymptotic of the first marginal auxiliary function $A_{1,y}$ with the small ball probability regime.
- Condition (II.1) is an integrability requirement, used in Proposition 2, for manipulating conditional stable tail dependence functions and their estimators as they appear through integrals in the loss function (12).
- Assumptions (I.4), (II.2), (II.3) are technical requirements for (32) related to the functional aspect and appearing in Proposition 2 and Lemma 5. For instance, the authors of Gijbels et al. [21] used Condition (II.2) for tightness and equicontinuity in the considered empirical processes. In particular, the object $\tau_{y,h}$ in List (I.4) was first introduced in Ferraty et al. [17] with the same condition on the limiting function $\tau_{y,0}$. The connection with assumption on M_1 in Condition (III.1) comes down from the fact that kernel expectations are key objects (see Gardes and Girard [19], Girard et al. [23]) in functional kernel methods and may be written as $\mathbb{E}_Y[K(h_n^{-1}\|Y - y\|_E)] = K(1)\psi_y(h_n) - \int_0^1 K'(s)\psi_y(sh_n)ds$ (see [17, Lemma 2]). The identification of $\tau_{y,0}$ is driven by the asymptotic behavior of $s \mapsto \psi_y(s) := \mathbb{P}(\|Y - y\|_E \leq s)$. As stated in Ferraty and Vieu [18], the class of fractal (infinite-dimensional) processes is the one for which the small ball probabilities behave as $\psi_y(s) \sim C_y s^a$, for $s \rightarrow 0$ and $C_y, a > 0$. For such processes, we have $\tau_{y,0}(s) = s^a$. This includes the multivariate Euclidean case $E = \mathbb{R}^p$, with $a = p$. On the other hand, for the class of processes such that $\psi_y(s) \sim C_y s^a \exp(-C/s^b)$, for $s \rightarrow 0$ and some $C_y, C, a, b > 0$, one has $\tau_{y,0}(s) = \delta_1(s)$ with δ_1 being the Dirac delta function at point 1 (see Ferraty et al. [17]). As noted in Gijbels et al. [21], for those processes, $\int_0^1 K'(s)\tau_{y,0}(s) ds$ is zero so that one usually assumes $K(1) > 0$ in this case. In the general case, the condition $K'(\cdot) \leq 0$ is asked. For more details about $\tau_{y,h}$ and β in Condition (II.3), the interested reader is referred to Section 2.3, especially Proposition 1 and Section 4, in Ferraty et al. [17], and Section 3.2 in Gijbels et al. [21].
- Note that the required assumptions of the present work are not formulated directly on the tail index γ_y . Instead of its continuity on E as in de Haan and Zhou [26], or even Lipschitz property as in Gardes and Girard [19] in connection with the Karamata representation, our set of hypotheses requires several regularity conditions on the conditional cumulative distribution function of X given $Y = y$ which is supposed to be continuous in x and y (see, for instance, Condition (II.2)).

6. Finite-sample results

We propose here to illustrate the finite-sample performances of our estimation method through a parametric model.

6.1. Considered model and parameters selection

The considered covariate model: We mimic Gardes and Girard [19], Girard et al. [23] by considering Z uniformly distributed on $[1/4, 1]$ and the following random process as covariate lying in $E = L^2([0, 1])$:

$$Y_Z(t) := \cos(2\pi Zt), \quad t \in [0, 1]. \tag{33}$$

Endowing the space E with its usual metric, we have $\|Y_Z\|_{L^2}^2 = \int_0^1 Y_Z^2(t)dt = \frac{1}{2} \left(1 + \frac{\sin(4\pi Z)}{4\pi Z}\right)$. In the following, we suppose that we observe the functional covariate $t \mapsto Y_Z(t)$ on a equispaced grid $0 = t_0 < t_1 < \dots < t_m = 1$ with $m = 101$. So that, the L^2 -norm will in fact be computed through a discretized version. In particular, for any $y \in E$ of the form $y_z(t) = \cos(2\pi zt)$ with $z \in [1/4, 1]$, one has $\|Y_{Z_i} - y_z\|_{L^2,d}^2 = \frac{1}{m} \sum_{k=1}^m (Y_{Z_i}(t_k) - y_z(t_k))^2$ with $Y_{Z_i}(t_k) = \cos(2\pi Z_i t_k)$ and $(Z_i)_{1 \leq i \leq n}$ being n independent observations of Z . We may consider the conditional tail index,

$$\gamma_Z := (8 \|Y_Z\|_{L^2}^2 - 3)/2.5.$$

Note that the realizations γ_z take values in the interval $[0.06, 0.61]$. Subsequently, we occasionally identify the functional covariate Y with its latent associated Z and so on for their realizations $y \in E$ and $z \in [1/4, 1]$. In the rest of this section, as a slight abuse of the notation, we denote the conditional tail index γ_y instead of γ_z , for the sake of readability with respect to the initial notation of the present work.

The considered conditional distributional model: As discussed in Section 5.3, we should slightly modify Section 4.2 of Beck et al. [3] by opting for conditional Lomax/centered Pareto type II marginal distributions linked through a survival Clayton dependence structure. Let us consider, for $y \in E$ and $s = (s_j)$, with $s_j = 1.25(1 + j)$, $j \geq 0$ a scale sequence similar to Beck et al. [3], a random vector $X = (X_1, \dots, X_d)^\top$ such that, conditionally to $Y = y$, the marginals X_j , $1 \leq j \leq d$, are Lomax($1/\gamma_y, s_j$)-distributed, i.e.,

$$\bar{F}_{j,y}(x) = \left(1 + \frac{x}{s_j}\right)^{-1/\gamma_y}, \quad x \geq 0.$$

We consider the survival Clayton copula with parameter $\gamma_y > 0$,

$$C_y^{Cl,S}(\mathbf{u}) = \max \left(\left(\sum_{j=1}^d u_j^{-\gamma_y} - d + 1 \right)^{-1/\gamma_y}, 0 \right), \quad \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d.$$

We denote this model as $X|Y = y \sim \text{Lomax-Clayton}(1/\gamma_y, s)$. Interesting, one may explicit several quantities related to the tail behavior and the dependence structure. Indeed, a quick computation yields the tail dependence function, for $1 \leq j \neq k \leq d$ and $(x_1, x_2) \in \mathbb{R}^+ \times \mathbb{R}^+$,

$$\lambda_{j,k,y}^{Cl} = \lim_{t \downarrow 0} \frac{C_{j,k,y}^{Cl,S}(tx_1, tx_2)}{t} = \lim_{t \downarrow 0} (x^{-\gamma_y} + y^{-\gamma_y} - t^{\gamma_y})^{-1/\gamma_y} = (x_1^{-\gamma_y} + x_2^{-\gamma_y})^{-1/\gamma_y},$$

and also the tail equivalence constants, $c_{j,y} = \left(\frac{s_j}{s_1}\right)^{1/\gamma_y}$ for $2 \leq j \leq d$.

Concerning the regular variation properties, it is given in [30, Example 5.1] that $\bar{F}_{j,y} \in 2\text{RV}_{-1/\gamma_y, -1(+\infty)}$ or equivalently $U_{j,y} = s_j(t^{\gamma_y} - 1) \in 2\text{RV}_{-\gamma_y, -\gamma_y}$ with auxiliary function $A_{j,y} : t \mapsto \gamma_y t^{-\gamma_y}$. Using Taylor expansion, one can derive the rate convergence μ_y in (6), i.e., $\mu_y = 1$.

Sampling procedure and input parameters: The sampling method at hand is by mixture Maume-Deschamps et al. [32]. One first generates G a Gamma distribution with shape parameter $1/\gamma_y$ and scale parameter 1. Then, one draws d independent exponential distributions with rate parameter G/s_j which in turn yields the desired model. Indeed, the model $X|Y = y$ may be described in terms of a compound process Maume-Deschamps et al. [31],

$$\mathbb{P} \left(\bigcap_{j=1}^d \{X_j|Y = y > x_j\} | G = g \right) = \prod_{j=1}^d e^{-g x_j}, \quad \mathbf{x} = (x_1, \dots, x_d) \geq 0.$$

Unless specified otherwise, we study the performance of our expectiles estimator over 500 independent replications of the data sample with size $n \in \{500, 2000\}$. The risk level is set to $\alpha_n = 0.968$ for $n = 500$, and $\alpha_n = 0.99$ for $n = 2000$, the intermediate sequence related to the estimation of λ_y is chosen to be $k_n = n^{0.7}$ (the interested reader is referred to Fig. 3 in Beck et al. [3]) and the subdivision length $J = 9$ motivated by Daouia et al. [7]. Let us consider the functional covariate value y associated to the particular choice of latent covariate value $z = \frac{1}{2}$. In this case, it yields $y(t) = \cos(\pi t) \in L^2([0, 1])$ which in turn leads to $\gamma_y = \frac{2}{5}$.

Kernels and bandwidths selection: Throughout the study, we consider the Nadaraya-Watson weights (16) with triangular kernel $K(t) = (1.9 - 1.8t)\mathbf{1}_{[0,1]}(t)$ akin to Girard et al. [23]. Unlike for the tail index and tail ratio which are solely related to the marginal tail behavior, we consider for the dependence part parabolic kernel $K(t) = (1 - t^2)\mathbf{1}_{[0,1]}(t)$ as in Gijbels et al. [21], especially in regard of List (III.1). As in Girard et al. [23], we select the bandwidth depending on $y \in E$ in order to take into account the heterogeneity of y on its support. To this aim, we completely mime the cross-validation procedure already developed in Section 6.1 in Girard et al. [23] for the same marginal distributions setting. This leads us to propose a choice of h_n related to y to guarantee that the 280 nearest neighbors for $n = 2000$, and the 100 nearest neighbors for $n = 500$, are included in the L^2 -ball centred in y and with radius h_n . Notice that this nearest neighbors choice of h_n forces the empirical small ball probability to be equal to $280/2000 = 0.14$ and $100/500 = 0.2$, respectively.

6.2. Equivalent marginal tails in $d = 2$ case

Explicit optimum solutions and associated expectiles: The Lomax quantile is $F_{j,y}^{-1}(x) = s_j((1-x)^{-\gamma_y} - 1)$, for $j \in \{1, 2\}$. In particular, the explicit tail ratios are $c_{1,y} = 1$ and $c_{2,y} = (3/2)^{5/2} = 2.7557$. Hence, an immediate application of Theorem 1 for our example case gives $e_\alpha(\mathbf{X}, y) \sim s_j((1-\alpha)^{-\gamma_y} - 1)\eta^{\gamma_y} (1, \beta_2)^\top$, as $\alpha \rightarrow 1$.

In the particular choice of latent covariate value $z = \frac{1}{2}$ with equivalent marginal tails and $\beta_1 = 1$, the loss function in (12) becomes

$$\mathcal{L}_{\xi_y}(\boldsymbol{\theta}) := \frac{1}{2} \sum_{k=1}^2 \varphi_{k,y}^2(\boldsymbol{\theta}, \xi_y), \quad \text{where} \quad \begin{cases} \varphi_{1,y}(\boldsymbol{\theta}, \xi_y) & := \frac{\gamma_y}{1-\gamma_y} - \eta \left(1 + \beta_2\right) + \int_{\beta_2}^{+\infty} \lambda_{1,2,y} \left(c_{2,y} t^{-1/\gamma_y}, 1\right) dt, \\ \varphi_{2,y}(\boldsymbol{\theta}, \xi_y) & := \frac{\gamma_y}{1-\gamma_y} - \eta \frac{\beta_2^{1/\gamma_y}}{c_{2,y}} \left(1 + \frac{1}{\beta_2}\right) + \int_{\frac{1}{c_{2,y}}}^{+\infty} \lambda_{1,2,y} \left(\frac{1}{c_{2,y}} t^{-1/\gamma_y}, 1\right) dt. \end{cases} \quad (34)$$

This problem admits $\boldsymbol{\theta}_y^* = (\eta^*, \beta_2^*) = (0.371, 1.675)$ as unique real solution. The solver is from the package Sympy in Python. Fig. 1 displays theoretical loss $\mathcal{L}_{\xi_y}(\boldsymbol{\theta})$ in (34) and associated estimated ones $\mathcal{L}_{\xi_{n,y}}(\eta, \beta_2)$ for one random data simulation with two different sample sizes (the red dot corresponds to the theoretical minimum solution point $\boldsymbol{\theta}_y^*$). In Table 1 we display the median of the maximum deviations between the true surface and the estimated one on 500 independent replications, for $n = 500$ and $n = 2000$. We see, in accordance with Proposition 1, the estimated surface getting closer to the true one when n increases.

In this numerical study for the Lomax-Clayton model, we are considering $k_n \approx n\psi_y(h_n)$ and $\mu_y \equiv 1$. Then, from Remark 1, one can explicitly characterize the convergence rate in Eq. (22) as follows $\delta_{n,y} \ll \min(\gamma_y(1 - \alpha_n)^{-\gamma_y}, \sqrt{n\psi_y(h_n)})$. Furthermore, since the adaptive choice of h_n and the associated value of the empirical small ball probability on the simulated samples, in the last column of Table 1 we can consider the rate $\delta_{n,y} = \gamma_y(1 - \alpha_n)^{-\gamma_y}$. The obtained numerical results of this study case seem to confirm the rate obtained in Proposition 1, since the maximum error multiplied by the rate decreases when n increases.

In Table 2 we illustrate the estimation of the targeted expectiles $e_\alpha(\mathbf{X}, y) = e_\alpha^1(\mathbf{X}, y), e_\alpha^2(\mathbf{X}, y)$ in (23). In order to illustrate the convergence to the theoretical expectile values for large n , we consider the particular case where α is fixed and not depending on n

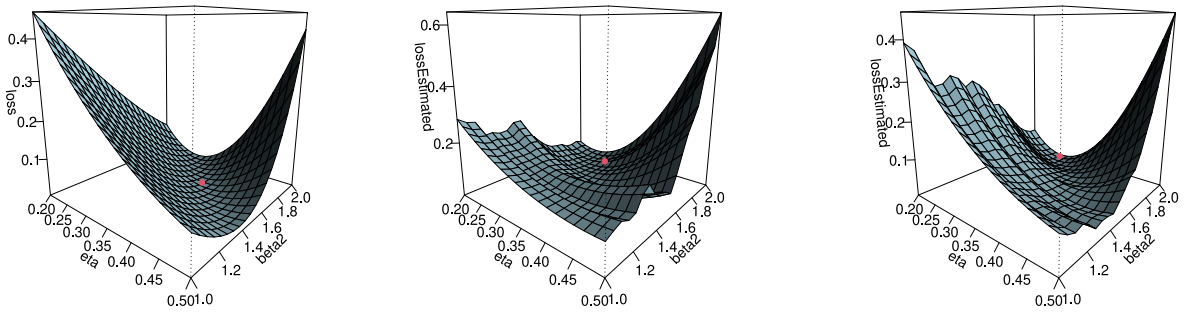


Fig. 1. Left panel: Theoretical loss surface $\mathcal{L}_{\xi_y}(\boldsymbol{\theta})$ in (34) as function of (η, β_2) . Center panel: one estimated loss surface for $n = 500$. Right panel : one estimated loss surface for $n = 2000$. The red dot corresponds to the theoretical minimum solution point $\boldsymbol{\theta}_y^* = (0.371, 1.675)$.

Table 1

Behavior of the supremum of $|\mathcal{L}_{\xi_y}(\boldsymbol{\theta}) - \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta})|$ over $N = 500$ independent replications of the data sample with respect of growing values of n and α_n . The considered rate is $\delta_{n,y} = \gamma_y(1 - \alpha_n)^{-\gamma_y}$.

n	α_n	$\sup_{\boldsymbol{\theta} \in \prod_{j=1}^d [e_j, M_j]} \mathcal{L}_{\xi_y}(\boldsymbol{\theta}) - \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta}) $	$\delta_{n,y} \sup_{\boldsymbol{\theta} \in \prod_{j=1}^d [e_j, M_j]} \mathcal{L}_{\xi_y}(\boldsymbol{\theta}) - \mathcal{L}_{\hat{\xi}_{n,y}}(\boldsymbol{\theta}) $
500	0.968	median = 0.2164 (sd = 0.1419)	median = 0.3439
2000	0.990	median = 0.1254 (sd = 0.0321)	median = 0.3166

Table 2

Behavior of the estimated $\hat{e}_\alpha^1(X, y)$ and $\hat{e}_\alpha^2(X, y)$ expectiles with respect to growing values of n and α_n .

n	α_n	$e_\alpha^1(X, y)$	$\hat{e}_\alpha^1(X, y)$	$e_\alpha^2(X, y)$	$\hat{e}_\alpha^2(X, y)$
500	0.968	4.9782	median = 4.5448 (sd = 0.3659)	8.3414	median = 8.3721 (sd= 0.7323)
2000	0.990	8.9231	median = 8.78191 (sd= 0.2929)	14.9513	median = 15.0724 (sd = 0.6785)

Table 3

Behavior of the estimated expectiles $\hat{e}_\alpha^1(X, y)$ and $\hat{e}_\alpha^2(X, y)$ with respect to growing values of n and fixed $\alpha = 0.990$. For the bigger value of n we considered the same h_n than in the case $n = 2000$.

n	α	$e_\alpha^1(X, y)$	$\hat{e}_\alpha^1(X, y)$	$e_\alpha^2(X, y)$	$\hat{e}_\alpha^2(X, y)$
500	0.990	8.9231	median = 8.3343 (sd = 0.6868)	14.9513	median = 15.5463 (sd= 1.3977)
2000	0.990	8.9231	median = 8.7819 (sd= 0.2929)	14.9513	median = 15.0724 (sd = 0.6785)
5000	0.990	8.9231	median = 8.8737 (sd= 0.2574)	14.9513	median = 15.0183 (sd = 0.5783)

(see Table 3). Notice that, we not only get closer to the true expectile values, but also that the variance of our estimator decrease with growing values of n .

On the estimation of ξ_y : Finally in Figs. 2 and 3 we illustrate the performance of the proposed consistent estimator vector $\hat{\xi}_{n,y} = (\hat{\gamma}_{n,y}, \hat{c}_{2,n,y}, \hat{\lambda}_y)$ of ξ_y (see Section 4). In particular in Fig. 2 we focus on the estimation of the conditional tail index $\hat{\gamma}_{n,y}$ and of the conditional tail ratio $\hat{c}_{2,n,y}$ for $n = 500$, $\alpha_n = 0.968$ (left panel) and $n = 2000$, $\alpha_n = 0.99$ (right panel). Fig. 3 is devoted to the comparison in finite-size samples of the tail dependence function estimation $\hat{\lambda}_y$ and the associated integral as presented in Section 4.1. The presented results empirically confirm the theoretical statements of our main result, i.e., the consistency of the tail index given in (31), the tail ratios in Lemma 2 and the integral of $\hat{\lambda}_y$ in Proposition 2.

6.3. Unit tail ratio in $d = 5$ case

We now assume that $c_{j,y} \equiv 1$ for any $1 \leq j \leq 5$ and $y \in E$ (e.g., $s_j = 5/2$, for all $1 \leq j \leq 5$). Again, Theorem 1 yields $e_\alpha(X, y) \sim s_1 \left((1 - \alpha)^{-\gamma_y} - 1 \right) \eta^{\gamma_y} (1, \beta_2, \beta_3, \beta_4, \beta_5)^\top$, as $\alpha \rightarrow 1$. The loss function may be written as

$$\mathcal{L}_{\xi_y}(\boldsymbol{\theta}) := \frac{1}{2} \sum_{k=1}^5 \varphi_{k,y}^2(\boldsymbol{\theta}, \xi_y), \tag{35}$$

where, with $\beta_1 = 1$ and $\varphi_{k,y}(\boldsymbol{\theta}, \xi_y) = \frac{\gamma_y}{1-\gamma_y} - \eta \beta_k^{1/\gamma_y} \left(1 + \sum_{j \neq k} \frac{\beta_j}{\beta_k} \right) + \sum_{j \neq k} \int_{\frac{\beta_j}{\beta_k}}^{+\infty} \lambda_{j,k,y} \left(t^{-1/\gamma_y}, 1 \right) dt = 0$ for $1 \leq k \leq 5$.

The optimum problem associated to the loss function in (35) admits a real solution $\boldsymbol{\theta}_y^* = \left(\frac{4}{5} \frac{2^{1-1/\gamma_y + \gamma_y}}{1-\gamma_y}, 1, 1, 1, 1 \right) \in \mathbb{R}^5$. Obtained performances of estimations of ξ_y and associated loss function are completely analogous to the previous study and then omitted here. They are essentially based on the same marginals for γ_y, c_j and on the bivariate vectors for the pairwise upper tail dependence functions λ_{jk} . This analogous behavior is completely obvious given the similar shape of the loss functions in (34) and (35).

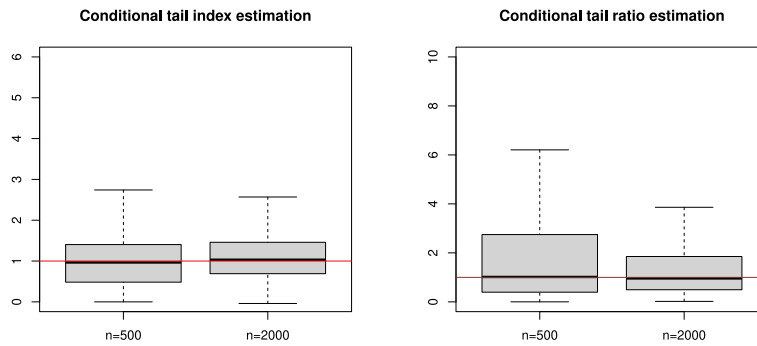


Fig. 2. Boxplots of the estimation of $\hat{\gamma}_{n,y}/\gamma_y$ with associate covariate $z = 0.5$ and of the tail ratio $\hat{c}_{2,n,y}/c_2$, with $\hat{c}_{2,n,y}$ in (20) $c_2 = 2.7557$. We have $\alpha_n = 0.968$ when $n = 500$ (left boxplot), and $\alpha_n = 0.99$ when $n = 2000$ (right boxplot). All boxplots are made from 500 Monte Carlo simulations.

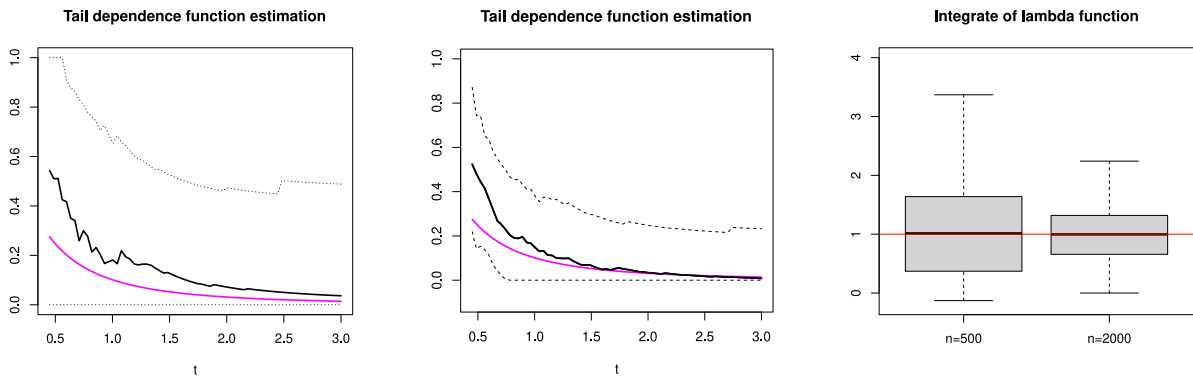


Fig. 3. Two first panels: estimation of $\lambda_{1,2,y}((c_{1,y}/c_{2,y})t^{-1/r}, 1)$ for respectively $n = 500$ and $n = 2000$. We display the theoretical function (magenta full line), with associated covariate $z = 0.5$, $c_{1,y} = 1$, $c_{2,y} = 2.7557$. Black full line corresponds to the averaged median behavior on 500 Monte Carlo simulations for the considered estimator $\hat{\lambda}_{1,2,n,y}$, for $t \in [\beta_1/\beta_2, 3]$, with $\beta_1 = 1$ and $\beta_2 = 1.675$. Dashed black lines to the 95% and 5% quantiles. Third panel: Boxplots on 500 Monte Carlo simulations of the ratio between empirical estimator and theoretical $\int_{\beta_1/\beta_2}^{+\infty} \lambda_{1,2,y}((c_{1,y}/c_{2,y})t^{-1/r}, 1) dt$ $n = 500$ and $n = 2000$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Proof of Proposition 1

Soliciting Δ -method (see, e.g., Theorem 3.8 in Vaart [41]), the proof of (26) and (27) is reduced to the establishment of the building block

$$\delta_{n,y} \max_{1 \leq k \leq d} \left| \varphi_{k,y}(\boldsymbol{\theta}, \boldsymbol{\xi}_y) - \varphi_{k,n,y}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}_{n,y}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \tag{36}$$

The proof of (28) and (29) relies upon, for $0 < \varepsilon_r < M_r < +\infty$, $1 \leq r \leq d$,

$$\delta_{n,y} \max_{1 \leq k \leq d} \sup_{\boldsymbol{\theta} \in \prod_{r=1}^d [\varepsilon_r, M_r]} \left| \varphi_{k,y}(\boldsymbol{\theta}, \boldsymbol{\xi}_y) - \varphi_{k,n,y}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}_{n,y}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \tag{37}$$

In view of (11), the proof of (36) and (37) comes down by proving the following convergence

$$\delta_{n,y} \max \left(\left| \Delta_{n,y}^{(1)} \right|, \left| \Delta_{k,n,y}^{(2)}(\boldsymbol{\theta}) \right|, \left| \Delta_{j,k,n,y}^{(3)}(\boldsymbol{\theta}) \right| \right) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad 1 \leq j \neq k \leq d, \quad y \in E, \tag{38}$$

and its uniform counterpart

$$\delta_{n,y} \sup_{\boldsymbol{\theta} \in \prod_{r=1}^d [\varepsilon_r, M_r]} \max \left(\left| \Delta_{n,y}^{(1)} \right|, \left| \Delta_{k,n,y}^{(2)}(\boldsymbol{\theta}) \right|, \left| \Delta_{j,k,n,y}^{(3)}(\boldsymbol{\theta}) \right| \right) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \tag{39}$$

where $\boldsymbol{\theta} = (\eta, \beta_2, \dots, \beta_d) \in (0, +\infty)^d$, $\beta_1 = 1$ and

$$\Delta_{n,y}^{(1)} := \frac{\gamma_y}{1 - \gamma_y} - \frac{\hat{\gamma}_{n,y}}{1 - \hat{\gamma}_{n,y}}, \quad \Delta_{k,n,y}^{(2)}(\boldsymbol{\theta}) = \eta \left(1 + \sum_{j \neq k} \frac{\beta_j}{\beta_k} \right) \left(\frac{\beta_k^{1/\gamma_y}}{c_{k,y}} - \frac{\beta_k^{1/\hat{\gamma}_{n,y}}}{\hat{c}_{k,n,y}} \right),$$

$$\Delta_{j,k,n,y}^{(3)}(\Theta) = \sum_{j \neq k} \left(\int_{\frac{\beta_j}{\beta_k}}^{+\infty} \hat{\lambda}_{j,k,n,y} \left(\frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} t^{-1/\hat{\gamma}_{n,y}}, 1 \right) - \lambda_{j,k,y} \left(\frac{c_{j,y}}{c_{k,y}} t^{-1/\gamma_y}, 1 \right) dt \right). \tag{40}$$

A straightforward application of the Δ -method and (31) implies that for any deterministic $1 \ll \delta_{0,n,y} \ll \tilde{\delta}_{0,n,y}$ and independently on $\Theta \in (0, +\infty)^d$,

$$\delta_{0,n,y} \left| \Delta_{n,y}^{(1)} \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

This readily provides (38) and (39) for $\Delta_{n,y}^{(1)}$. The quantities $\Delta_{k,n,y}^{(2)}$ and $\Delta_{j,k,n,y}^{(3)}$ are respectively tackled in the upcoming Lemma 2 and Proposition 2. One can obtain the identities for the gradient loss function by adapting the setting in Appendix A.3. in Beck et al. [3] to our functional covariate case. By using these gradient expressions we can show the gradient convergence in (28)–(29). Indeed, the random quantities $\hat{\gamma}_y$ and $\hat{c}_{k,y}$ appear through algebraic expressions in the gradient loss function in the same form as in (11).

The Karamata’s results, Theorem B.1.4 - B.1.5, Equation (B.1.14), Remark B.1.7 in De Haan and Ferreira [11], see also Theorem 1.3.1 in Bingham et al. [6], allow us to write

$$\ell_{U_{j,y}}(x) := \ell_{j,y}(x) = \kappa_j(y) b_{j,y}(x) \exp \left(\int_{x_0}^x \frac{b_{j,y}(t) - 1}{t} dt \right), \tag{41}$$

for some arbitrary $x_0 > 0$, some constant $\kappa_j(y) \in (0, +\infty)$ and

$$b_{j,y}(x) = \frac{x \ell_{j,y}(x)}{\int_{x_0}^x \ell_{j,y}(t) dt}. \tag{42}$$

Studying the Karamata representation (41) of the slowly varying part $\ell_{j,y}$ through the point of view of Π -Class (see Definition 4), we may show in the following Lemma 1 an asymptotic expansion on $\ell_{j,y}$.

Notice that Lemma 1 in the case $\rho_{j,y} < 0$ is a straightforward generalization to the conditional setting of Lemma 1 in Hua and Joe [28]. The proof in the specific case where $\rho_{j,y} = 0$ is postponed to Section 8.1. For examples of distributions satisfying the second order regular variation condition with $\rho_{j,y} = 0$ we interested reader is referred to Remark 5.

Lemma 1. Assume that for any $y \in E$, for any $1 \leq j \leq d$, $U_{j,y} \in 2RV_{-\gamma_y, \rho_{j,y}}(+\infty)$, $\gamma_y > 0$, $\rho_{j,y} \leq 0$, with auxiliary function $A_{j,y} \in RV_{\rho_{j,y}}(+\infty)$ or equivalently Condition (H1). Then, for some constant $\kappa_j(y) \in (0, +\infty)$ independent of x ,

$$\ell_{j,y}(x) = \kappa_j(y) (1 + O(A_{j,y}(x))), \quad x \rightarrow +\infty.$$

Once the slowly varying part of the tail quantile function is determined, we may derive, under the second order regular variation property (see List (H2)), the speed rate in the consistency of the tail ratios.

Lemma 2. Let $\Theta = (\eta, \beta_2, \dots, \beta_d) > 0$, $y \in E$, $\alpha_n = 1 + o(1)$, $h_n \ll 1$ deterministic such that the intermediate regime $n \bar{\alpha}_n \psi_y(h_n) \gg 1$ holds true. For each marginal $2 \leq k \leq d$, let $\delta_{k,n,y}$ be a scaling sequence such that

$$1 \ll \delta_{k,n,y} \ll \min_{s \in \{0,1,k\}} \tilde{\delta}_{s,n,y}, \quad \text{where } \tilde{\delta}_{s,n,y} \text{ is defined in (21)}. \tag{43}$$

Then, as $n \rightarrow +\infty$, under Assumptions (H1), (H2), (III.1), (I.1), (I.2), (I.3),

$$\delta_{k,n,y} \left| c_{k,y} - \hat{c}_{k,y} \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \tag{44}$$

Furthermore, it holds that, for $1 \leq k \leq d$,

$$\delta_{k,n,y} \left| \Delta_{k,n,y}^{(2)}(\Theta) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad \delta_{k,n,y} \sup_{\Theta \in \prod_{r=1}^d [\epsilon_r, M_r]} \left| \Delta_{k,n,y}^{(2)}(\Theta) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0,$$

for $0 < \epsilon_r < M_r < +\infty$, $1 \leq r \leq d$.

As defined in (43), the auxiliary marginal scaling sequences $\delta_{k,n,y}$ may be combined to recover the global rate $\delta_{n,y}$ defined in (22) by considering the minimum over $\tilde{\delta}_{-1,n,y}$ and over every $\delta_{k,n,y}$ for $k \geq 2$. The proof of Lemma 2 is postponed to Section 8.1. Next, we consider the integral of stable tail dependence functions and prove the following convergences.

Proposition 2. Let $\delta_{n,y}$ as in (22). Let $y \in E$, $1 \leq j \neq k \leq d$, $\Theta = (\eta, \beta_2, \dots, \beta_d) \in (0, +\infty)^d$ and $\beta_1 = 1$. If Assumptions (H1), (H2), (3), (6) and Assumptions of classes I, II, III in Section 5.1 are satisfied, then

$$\delta_{n,y} \left| \Delta_{j,k,n,y}^{(3)}(\Theta) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0,$$

and for any $0 < \epsilon_r < M_r < +\infty$, $1 \leq r \leq d$,

$$\delta_{n,y} \sup_{\Theta \in \prod_{r=1}^d [\epsilon_r, M_r]} \left| \Delta_{j,k,n,y}^{(3)}(\Theta) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

The proof of Proposition 2 can be found in Section 8.1.

8. Auxiliary proofs and results

We gather in this section the proofs of the results needed for Proposition 1 as well as further auxiliary results required for the intermediate steps of the proof.

8.1. Proofs of lemmas of Section 7

We introduce a particular proxy version to the second order regular variation property to be connected to slowly varying functions when the second order tail index is zero. In this context, we constructively specify the auxiliary function and shall make use of an uniform Drees-type inequality. This is of importance in order to address in depth the Karamata representation applied to the slowly varying part of the conditional marginal quantile tail function.

Definition 4 (Class Π).[Definition B.2.4 in De Haan and Ferreira [11]] A measurable function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ belongs to the class $\Pi(a)$ if there exists an auxiliary function $a : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that for $x > 0$,

$$\lim_{t \rightarrow +\infty} \frac{f(tx) - f(t)}{a(t)} = \ln(x). \tag{45}$$

Remark 3. The auxiliary function is slowly varying, namely $a \in \text{RV}_0(+\infty)$, by Theorem B.2.7 in De Haan and Ferreira [11]. By Theorem B.2.12 in De Haan and Ferreira [11], $f \in \Pi(a)$ is equivalent to $f \in \Pi(\tilde{a})$ with $\tilde{a}(t) = f(t) - \frac{1}{t} \int_0^t f(s)ds$. Hence, if $f \in \Pi(a)$ for some auxiliary function a , we can always choose instead the alternative auxiliary function \tilde{a} . More can be said, if in addition $x^\gamma f(x) \in 2\text{RV}_{\gamma,0}$, $\gamma > 0$. In this case, the auxiliary function A in Definition A.2 can be chosen as $A(t) = \tilde{A}(t) := \frac{\tilde{a}(t)}{f(t)}$. According to Proposition B.2.17 in De Haan and Ferreira [11], if $f \in \Pi(a)$, then for any $\epsilon, \delta > 0$, there exists $x_0 = x_0(\epsilon, \delta) \geq 0$, such that for any t and x with $tx \geq x_0$, the following uniform bound holds

$$\left| \frac{\ell(tx) - \ell(x)}{\tilde{a}(x)} - \ln(t) \right| \leq \epsilon (\max(t^\delta, t^{-\delta})), \tag{46}$$

with the relation $\tilde{A}\ell = \tilde{a}$.

Proof of Lemma 1. Let $1 \leq j \leq d$ and $y \in E$. The case $\rho_{j,y} < 0$ is a straightforward generalization to the conditional setting of Lemma 1 in Hua and Joe [28]. We may now assume the second order tail index of $U_{j,y}$ to be zero, i.e., $\rho_{j,y} = 0$.

According to the definition of the second order regular variation, there exists auxiliary functions $A_{j,y} := AU_{j,y} \in \text{RV}_0(+\infty)$ for $1 \leq j \leq d$ with ultimately constant sign and $\lim_{t \rightarrow +\infty} A_{j,y}(t) = 0$ such that

$$\lim_{t \rightarrow +\infty} \frac{1}{A_{j,y}(t)\ell_{j,y}(t)} (\ell_{j,y}(tx) - \ell_{j,y}(t)) = \ln(x), \quad x > 0.$$

Since $\ell_{j,y}$ is non-negative, we immediately deduce that

$$\ell_{j,y} \in \Pi(a_{j,y}), \quad a_{j,y}(t) := \text{sign} \{A_{j,y}(t)\} A_{j,y}(t)\ell_{j,y}(t) \in \text{RV}_0(+\infty), \quad a_{j,y} \geq 0.$$

By Remark 3, the definition of the second order regular variation and the relation $U_{j,y}(x) = x^{\gamma_y} \ell_{j,y}(x)$, we can modify the auxiliary functions so that

$$\ell_{j,y}(\cdot) \in \Pi(\tilde{a}_{j,y}), \quad \tilde{a}_{j,y}(t) := \ell_{j,y}(t) - \frac{1}{t} \int_0^t \ell_{j,y}(s) ds,$$

and $U_{j,y}(\cdot) \in 2\text{RV}_{-1/\gamma_y,0}(+\infty)$, with auxiliary function

$$A_{j,y}(t) = \frac{\tilde{a}_{j,y}(t)}{\ell_{j,y}(t)} = 1 - \int_0^1 \frac{\ell_{j,y}(tu)}{\ell_{j,y}(t)} du \in \text{RV}_0(+\infty).$$

Fix $\epsilon > 0$ and $\delta \in (0, 1)$. Let $x_0(\epsilon, \delta) > 0$ such that (46) holds. From Potter's Theorem B.1.9 in De Haan and Ferreira [11] with $\delta_1 = 1$ and $\delta_2 = 1/2$, let $x_0(\delta_1, \delta_2) > 0$ such that Potter's bounds hold as well. We finally set $x_0 := \max(x_0(\epsilon, \delta), x_0(\delta_1, \delta_2))$. In the rest of the proof, we consider this particular choice of x_0 in (41) and investigate (42).

With the substitution $u = \frac{t}{x}$ in (42), one can express $b_{j,y}^{-1}(x) = \int_{\frac{x_0}{x}}^1 \frac{\ell_{j,y}(ux)}{\ell_{j,y}(x)} du$. Then, with $A_{j,y} \in \text{RV}_0(+\infty)$,

$$\frac{1}{A_{j,y}(x)} \left(b_{j,y}^{-1}(x) - 1 + \frac{x_0}{x} \right) = \int_0^1 \frac{1}{A_{j,y}(x)} \left[\frac{\ell_{j,y}(ux)}{\ell_{j,y}(x)} - 1 \right] \mathbf{1}_{[\frac{x_0}{x}, 1]}(u) du.$$

For $u \in (0, 1)$ and x such that $ux \geq x_0$, the bound (46) yields

$$\left| \frac{1}{A_{j,y}(x)} \left[\frac{\ell_{j,y}(ux)}{\ell_{j,y}(x)} - 1 \right] \mathbf{1}_{[\frac{x_0}{x}, 1]}(u) \right| \leq -\ln(u) + \epsilon u^{-\delta} := f_{\epsilon, \delta}(u).$$

The function $u \mapsto f_{\varepsilon,\delta}(u)$ is integrable on $(0, 1)$ as $-1 < -\delta < 0$. Altogether with the pointwise convergence (45), the dominated convergence theorem gives

$$\int_0^1 \frac{1}{A_{j,y}(x)} \left[\frac{\ell_{j,y}(ux)}{\ell_{j,y}(x)} - 1 \right] \mathbf{1}_{\left[\frac{x_0}{x}, 1\right]}(u) du \xrightarrow{x \rightarrow +\infty} \kappa_0 := \int_0^1 \ln(u) du = -1 < +\infty.$$

The previous convergence holds with limiting constant $-\kappa_0$ instead of κ_0 by accordingly adjusting the sign of the auxiliary function $A_{j,y}$. Opting for this convention, we subsequently set $\kappa_0 = 1$. One may then write

$$b_{j,y}^{-1}(x) = 1 + A_{j,y}(x) - \frac{x_0}{x} + o(A_{j,y}(x)).$$

Since $x \mapsto \frac{1}{x} \in RV_{-1}(+\infty)$ and $A_{j,y} \in RV_0(+\infty)$, it follows that $b_{j,y}(x) = 1 + A_{j,y}(x) + o(A_{j,y}(x))$, $x \rightarrow +\infty$.
Going back to (41),

$$\ell_{j,y}(x) = \kappa_j(y) \left(1 + A_{j,y}(x) + o(A_{j,y}(x)) \right) \exp \left(\int_{x_0}^x \frac{A_{j,y}(t) + o(A_{j,y}(t))}{t} dt \right).$$

Repeating the same argument of dominated convergence but in the first order regular variation context for the auxiliary function $A_{j,y} \in RV_0(+\infty)$ with Potter’s Theorem B.1.9 in De Haan and Ferreira [11],

$$\int_{x_0}^x \frac{A_{j,y}(t)}{t} dt = A_{j,y}(x) + o(A_{j,y}(x)), \quad x \rightarrow +\infty.$$

Therefore, a Taylor expansion as $x \rightarrow +\infty$ yields the asymptotic representation

$$\ell_{j,y}(x) = \kappa_j(y) \left(1 + O(A_{j,y}(x)) \right), \quad A_{j,y} \in RV_0(+\infty), \quad \kappa_j(y) \in (0, +\infty). \quad \square$$

Proof of Lemma 2. Let $y \in E$ and $2 \leq k \leq d$. First, observe that straight consequences of (30) as well as Δ -method applied to (31) are

$$\frac{\hat{q}_{k,n,y}(\alpha_n)}{q_{k,y}(\alpha_n)} = 1 + O_{\mathbb{P}} \left(\tilde{\delta}_{0,n,y}^{-1} \right), \quad \hat{\gamma}_{n,y}^{-1} - \gamma_y^{-1} = O_{\mathbb{P}} \left(\tilde{\delta}_{0,n,y}^{-1} \right). \tag{47}$$

One may write

$$\ln \left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{\hat{q}_{1,n,y}(\alpha_n)} \right) = \ln \left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{q_{k,y}(\alpha_n)} \right) + \ln \left(\frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)} \right) + \ln \left(\frac{q_{1,y}(\alpha_n)}{\hat{q}_{1,n,y}(\alpha_n)} \right).$$

In view of (47) and after Taylor expanding, this becomes

$$\ln \left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{\hat{q}_{1,n,y}(\alpha_n)} \right) = \ln \left(\frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)} \right) + O_{\mathbb{P}} \left(\tilde{\delta}_{0,n,y}^{-1} \right). \tag{48}$$

We investigate the logarithm term in the RHS of (48) and precise its asymptotic behavior. Regarding the quantiles, it is clear that, for $\alpha_n \ll 1$ and each $1 \leq j \leq d$, we have $q_{j,y}(\alpha_n) \gg 1$ and $q_{j,y} \in RV_{\gamma_y}(1)$. As a functional version of Lemma 3.2 in Maume-Deschamps et al. [34] which may be proved along the same steps, we have that, if $t = O(s)$, for any $1 \leq i, j \leq d$,

$$\frac{\bar{F}_{i,y}(s)}{\bar{F}_{j,y}(t)} \sim \frac{c_{i,y}}{c_{j,y}} \left(\frac{s}{t} \right)^{-1/\gamma_y}, \quad t \rightarrow +\infty.$$

It follows that $1 = \frac{\bar{F}_{k,y}(q_{k,y}(\alpha_n))}{\bar{F}_{1,y}(q_{1,y}(\alpha_n))} = c_{k,y} \left(\frac{q_{1,y}(\alpha_n)}{q_{k,y}(\alpha_n)} \right)^{1/\gamma_y} + o(1)$, $n \rightarrow +\infty$. Hence,

$$c_{k,y} = \left(\frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)} \right)^{1/\gamma_y} + o(1), \quad \text{i.e.,} \quad \frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)} = c_{k,y}^{\gamma_y} + o(1).$$

On the other hand, one may write the ratio in terms of slowly varying functions

$$\frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)} = \frac{U_{k,y}(\bar{\alpha}_n^{-1})}{U_{1,y}(\bar{\alpha}_n^{-1})} = \frac{\ell_{k,y}(\bar{\alpha}_n^{-1})}{\ell_{1,y}(\bar{\alpha}_n^{-1})}.$$

By Lemma 1, for $1 \leq j \neq k \leq d$ and as $x \rightarrow +\infty$, the following asymptotic holds after a Taylor expansion

$$\frac{\ell_{j,y}(x)}{\ell_{k,y}(x)} = \frac{\kappa_j(y)}{\kappa_k(y)} \left[1 + O \left(\max(A_{j,y}(x), A_{k,y}(x)) \right) \right].$$

We deduce the tail ratio representation $c_{k,y} = (\kappa_k(y)/\kappa_1(y))^{1/\gamma_y}$ and that for $1 \leq k \leq d$,

$$0 \leq \limsup_{n \rightarrow +\infty} \min(\tilde{\delta}_{1,n,y}, \tilde{\delta}_{k,n,y}) \left(\frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)} - c_{k,y}^{\gamma_y} \right) < +\infty. \tag{49}$$

We can state (49) in an alternative way $c_{k,y}^{\gamma_y} \frac{q_{1,y}(\alpha_n)}{q_{k,y}(\alpha_n)} = 1 + O\left(\max\left(\tilde{\delta}_{1,n,y}^{-1}, \tilde{\delta}_{k,n,y}^{-1}\right)\right)$, $n \rightarrow +\infty$. Combining (31), (47), (48) and (49) with the definition of $\hat{c}_{k,y}$ in (20), it yields

$$\ln(\hat{c}_{j,n,y}) = \frac{1}{\hat{\gamma}_{n,y}} \ln\left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{\hat{q}_{1,n,y}(\alpha_n)}\right) = \frac{1}{\hat{\gamma}_{n,y}} \ln\left(\frac{q_{k,y}(\alpha_n)}{q_{1,y}(\alpha_n)}\right) + O_{\mathbb{P}}\left(\tilde{\delta}_{0,n,y}^{-1}\right) = \frac{\gamma_y}{\hat{\gamma}_{n,y}} \ln c_{k,y} + o_{\mathbb{P}}\left(\max\left(\tilde{\delta}_{1,n,y}^{-1}, \tilde{\delta}_{k,n,y}^{-1}\right)\right) + O_{\mathbb{P}}\left(\tilde{\delta}_{0,n,y}^{-1}\right).$$

Limit in (44) is proved after an application of the Δ -method. We now consider the quantity

$$\frac{\beta_k^{1/\hat{\gamma}_{n,y}}}{\hat{c}_{k,n,y}} \left(\frac{\beta_k^{1/\gamma_y}}{c_{k,y}}\right)^{-1} = \exp\left[-\hat{\gamma}_{n,y}^{-1} \ln\left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{\hat{q}_{1,n,y}(\alpha_n)}\right) + \left(\hat{\gamma}_{n,y}^{-1} - \gamma_y^{-1}\right) \ln(\beta_k) + \ln(c_{k,y})\right]. \tag{50}$$

Combining again (31), (47), (48) and (49), and Taylor expanding in (50), it yields

$$\frac{\beta_k^{1/\hat{\gamma}_{n,y}}}{\hat{c}_{k,n,y}} - \frac{\beta_k^{1/\gamma_y}}{c_{k,y}} = \frac{c_{k,y}^{\gamma_y}}{\gamma_y} O_{\mathbb{P}}\left(\max\left(\tilde{\delta}_{1,n,y}^{-1}, \tilde{\delta}_{k,n,y}^{-1}\right)\right) + O_{\mathbb{P}}\left(\tilde{\delta}_{0,n,y}^{-1}\right).$$

The previous asymptotic equality is valid for any $1 \leq k \leq d$. Moreover, the asymptotic holds uniformly over $\beta_k \in [\varepsilon_k, M_k]$ with $0 < \varepsilon_k < M_k < +\infty$. Indeed, in view of the term $\ln(\beta_k)$ in (50), we discriminate two cases: when $0 < \beta_k \neq 1$ and when $\beta_k = 1$.

In the former case, it is clear that $|\ln \beta_k|$ is positive and bounded by a constant depending on ε_k, M_k .

In the second case, the speed rate $\tilde{\delta}_{0,n,y}$ induced by the difference $\hat{\gamma}_{n,y}^{-1} - \gamma_y^{-1}$ apparently vanishes but is reintroduced by the term $\ln\left(\frac{\hat{q}_{k,n,y}(\alpha_n)}{\hat{q}_{k,n,y}(\alpha_n)}\right)$. Now, the term $\eta(1 + \sum_{j \neq k} \beta_j / \beta_k)$ is positive and bounded when $\Theta = (\eta, \beta_2, \dots, \beta_d) \in \prod_{k=1}^d [\varepsilon_k, M_k]$ and $\beta_1 = 1$. Hence, the second part of Lemma 2 is proved. \square

Proof of Proposition 2. In the following, for the sake of readability, we drop the finite sum in (40). We start by decomposing $\Delta_{j,k,n,y}^{(3)}$ defined in (40) as $\Delta_{j,k,n,y}^{(3)} = \Delta_{j,k,n,y}^{(3,1)} + \Delta_{j,k,n,y}^{(3,2)}$ with

$$\begin{aligned} \Delta_{j,k,n,y}^{(3,1)} &= \int_{\frac{\beta_j}{\beta_k}}^{+\infty} \lambda_{j,k,y} \left(\frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} t^{-1/\hat{\gamma}_{n,y}}, 1\right) dt - \int_{\frac{\beta_j}{\beta_k}}^{+\infty} \lambda_{j,k,y} \left(\frac{c_{j,y}}{c_{k,y}} t^{-1/\gamma_y}, 1\right) dt, \\ \Delta_{j,k,n,y}^{(3,2)} &= \int_{\frac{\beta_j}{\beta_k}}^{+\infty} \hat{\lambda}_{j,k,n,y} \left(\frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} t^{-1/\hat{\gamma}_{n,y}}, 1\right) dt - \int_{\frac{\beta_j}{\beta_k}}^{+\infty} \lambda_{j,k,y} \left(\frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} t^{-1/\hat{\gamma}_{n,y}}, 1\right) dt, \end{aligned}$$

where $\hat{\gamma}_{n,y}$ and $\hat{c}_{j,n,y}$, $1 \leq j \leq d$, are defined in (19) and (20). Note that $\lambda_{j,k,y}(0, 1) = 1 - L_{j,k,y}(0, 1) = 0$.

In view of List (II.1), the mapping $t \mapsto t^{-(\gamma_y+1)} \lambda_{j,k,y}(t, 1)$ is integrable in the proximity of 0 so we may consider the change of variable

$$\frac{c_{j,y}}{c_{k,y}} t^{-1/\gamma_y} = u \iff t = \left(\frac{c_{k,y}}{c_{j,y}} u\right)^{-\gamma_y}, \quad dt = -\gamma_y \left(\frac{c_{k,y}}{c_{j,y}}\right)^{-\gamma_y} u^{-(\gamma_y+1)} du.$$

It yields the alternative form

$$\Delta_{j,k,n,y}^{(3,1)} = -v_y \int_0^{M_y(\Theta)} \{\lambda_{j,k,y}(\hat{s}_{j,k,n,y}(t), 1) - \lambda_{j,k,y}(t, 1)\} dt, \quad \Delta_{j,k,n,y}^{(3,2)} = -\hat{v}_{n,y} \int_0^{\hat{M}_{n,y}(\Theta)} \{\hat{\lambda}_{j,k,n,y}(t, 1) - \lambda_{j,k,y}(t, 1)\} dt,$$

with

$$\hat{s}_{j,k,n,y}(t) := \frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} \left(\frac{c_{k,y}}{c_{j,y}}\right)^{\gamma_y/\hat{\gamma}_{n,y}} \cdot t^{\gamma_y/\hat{\gamma}_{n,y}} := \hat{\sigma}_{n,y} \cdot t^{\gamma_y/\hat{\gamma}_{n,y}}, \tag{51}$$

$$M_y(\Theta) := \frac{c_{j,y}}{c_{k,y}} \left(\frac{\beta_j}{\beta_k}\right)^{-1/\gamma_y}, \quad v_y := \gamma_y \left(\frac{c_{k,y}}{c_{j,y}}\right)^{-\gamma_y}, \quad \hat{M}_{n,y}(\Theta) := \frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} \left(\frac{\beta_j}{\beta_k}\right)^{-1/\hat{\gamma}_{n,y}}, \quad \hat{v}_{n,y} := \hat{\gamma}_{n,y} \left(\frac{\hat{c}_{k,n,y}}{\hat{c}_{j,n,y}}\right)^{-\hat{\gamma}_{n,y}}.$$

In order to prove Proposition 2, it is enough to show that, for $s \in \{1, 2\}$,

$$\delta_{n,y} \left| \Delta_{j,k,n,y}^{(3,s)}(\Theta) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \tag{52}$$

and the uniform counterparts, for $0 < \varepsilon_r < M_r < +\infty$, $1 \leq r \leq d$, $\delta_{n,y} \sup_{\Theta \in \prod_{r=1}^d [\varepsilon_r, M_r]} \left| \Delta_{j,k,n,y}^{(3,s)}(\Theta) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$.

From (31) and (44), for any $y \in E$, $1 \leq j \leq d$ and $t \geq 0$, we have $\hat{\gamma}_{n,y} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \gamma_y$, $\hat{c}_{j,n,y} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} c_{j,y}$. In particular, the continuous mapping theorem readily implies that $\hat{s}_{j,k,n,y}(t) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} t$ and also

$$\hat{M}_{n,y}(\Theta) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} M_y(\Theta), \quad \hat{v}_{n,y}(t) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} v_y. \tag{53}$$

It is classic knowledge that the stable tail dependence function $\lambda_{j,k,y}$ is 1-Lipschitz, hence, the case $\Delta_{n,y}^{(1)}$ in (52) follows from Lemma 4. On the other hand, the case $\Delta_{k,n,y}^{(2)}$ in (52) is a mere consequence of the uniform boundedness over $\beta := (1, \beta_2, \dots, \beta_d) \in \prod_{j=1}^d [\varepsilon_j, M_j]$ with $0 < \varepsilon_j < M_j < +\infty$, for any j , of the quantities in (53) and the uniform convergence in probability of Lemma 5. \square

8.2. Auxiliary lemmas

In this section we introduce and prove three auxiliary results. Recall the scaling sequence $\delta_{k,n,y}$ for each margins $1 \leq k \leq d$ defined in (43). Firstly we may have a closer look to the previous convergence of $\hat{s}_{j,k,n,y}$ in (51). It follows from (31) and (44) that $\delta_{k,n,y} (\hat{\sigma}_{n,y} - 1) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. An application of the multivariate Δ -method yields the following pointwise convergence in probability $\delta_{k,n,y} (\hat{s}_{j,k,n,y}(t) - t) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. Leaning on the monotonicity of the mapping $t \mapsto \hat{s}_{j,k,n,y}(t)$, we improve the pointwise convergence in probability $\hat{s}_{j,k,n,y}(t) - t = o_{\mathbb{P}}(1)$ to an uniform version by a probabilistic Dini-type argument.

Lemma 3. For any $0 < T < +\infty$ and $1 \leq j \neq k \leq d$, it holds that $\sup_{t \in [0, T]} \left| \hat{s}_{j,k,n,y}(t) - t \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$.

Proof of Lemma 3. Let $T > 0$ and $\varepsilon > 0$. Let also an integer $N := N(\varepsilon) \geq 1$ such that $T < \varepsilon N$. We consider a subdivision with step less than ε of the compact set $[0, T] = \cup_{\ell=0}^{m-1} [\frac{\ell}{N}T, \frac{\ell+1}{N}T]$. In particular, for any $t \in [0, T]$, there exists $0 \leq \ell_1 < \ell_2 \leq N$ such that $\frac{\ell_1}{N}T \leq t \leq \frac{\ell_2}{N}T$. Fix $\omega \in \Omega$ for the moment and assume that for any $0 \leq \ell \leq N$, $\left| \hat{s}_{j,k,n,y} \left(\frac{\ell}{N}T \right) (\omega) - \frac{\ell}{N}T \right| < \varepsilon$. Since the mapping $x \mapsto \hat{s}_{j,k,n,y}(x)(\omega)$ is non-decreasing, the following inequalities hold

$$\begin{aligned} \hat{s}_{j,k,n,y}(t)(\omega) - t &\leq \hat{s}_{j,k,n,y} \left(\frac{\ell_2}{N}T \right) (\omega) - \frac{\ell_2}{N}T \leq s_{j,k,n,y} \left(\frac{\ell_2}{N}T \right) (\omega) - \frac{\ell_2}{N}T + \varepsilon < 2\varepsilon, \\ \hat{s}_{j,k,n,y}(t)(\omega) - t &\geq \hat{s}_{j,k,n,y} \left(\frac{\ell_1}{N}T \right) (\omega) - \frac{\ell_2}{N}T \geq \hat{s}_{j,k,n,y} \left(\frac{\ell_1}{N}T \right) (\omega) - \frac{\ell_1}{N}T - \varepsilon > -2\varepsilon. \end{aligned}$$

Thus, $\sup_{t \in [0, T]} \left| \hat{s}_{j,k,n,y}(t)(\omega) - t \right| < 2\varepsilon$. As a consequence,

$$\mathbb{P} \left(\left| \hat{s}_{j,k,n,y} \left(\frac{\ell}{N}T \right) - \frac{\ell}{N}T \right| < \varepsilon, \forall \ell = 0, \dots, N \right) \leq \mathbb{P} \left(\sup_{t \in [0, T]} \left| \hat{s}_{j,k,n,y}(t) - t \right| < 2\varepsilon \right).$$

Taking the complementary, using union bound and the previous pointwise convergence in probability of $\hat{s}_{j,k,n,y}$, as $N, T, \varepsilon < +\infty$ are independent of n ,

$$\mathbb{P} \left(\sup_{t \in [0, T]} \left| \hat{s}_{j,k,n,y}(t) - t \right| > \varepsilon \right) \leq \sum_{\ell=0}^N \mathbb{P} \left(\left| \hat{s}_{j,k,n,y} \left(\frac{\ell}{N}T \right) - \frac{\ell}{N}T \right| > \frac{\varepsilon}{2} \right) \xrightarrow[n \rightarrow +\infty]{} 0. \quad \square$$

In the following lemma we provide the speed rate of convergence in Lemma 3.

Lemma 4. Let $\delta_{k,n,y}$ be the scaling sequence for each margins $1 \leq k \leq d$ defined in (43). Let $y \in E$ and $0 < T < +\infty$. Under Assumptions (H1), (H2) and (I.1), (I.2), (I.3), (III.1), for $j \neq k$,

$$\delta_{k,n,y} \sup_{t \in [0, T]} \left| \hat{s}_{j,k,n,y}(t) - t \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Proof of Lemma 4. Recall $\hat{\sigma}_{n,y} := \frac{\hat{c}_{j,n,y}}{\hat{c}_{k,n,y}} \left(\frac{c_{k,y}}{c_{j,y}} \right)^{\gamma_y / \hat{\gamma}_{n,y}}$. Define $g_{j,k,n,y}(t) := \hat{s}_{j,k,n,y}(t) - t$ for $t \geq 0$ with $\hat{s}_{j,k,n,y}$ as in (51). Consider

$t_{n,y} := \left(\frac{\gamma_y}{\hat{\gamma}_{n,y}} \hat{\sigma}_{n,y} \right)^{\frac{\hat{\gamma}_{n,y}}{\hat{\gamma}_{n,y} - \gamma_y}}$ the solution of $g'_{j,k,n,y}(t) = \hat{\sigma}_{n,y} \frac{\gamma_y - t \gamma_y / \hat{\gamma}_{n,y} - 1}{\hat{\gamma}_{n,y}} = 0, \quad t \geq 0$. This corresponds to the maximum value of $g_{j,k,n,y}$ when $\gamma_y < \hat{\gamma}_{n,y}$ or its minimum value when $\gamma_y > \hat{\gamma}_{n,y}$. This is a consequence of the 2nd order optimality condition and

$$g''_{j,k,n,y}(t) = \hat{\sigma}_{n,y} \frac{\gamma_y}{\hat{\gamma}_{n,y}} (\gamma_y / \hat{\gamma}_{n,y} - 1) t^{\gamma_y / \hat{\gamma}_{n,y} - 2}, \quad t \geq 0.$$

Using (31) and (44), we easily see that

$$t_{n,y} = \left(\frac{\gamma_y}{\hat{\gamma}_{n,y}} \hat{\sigma}_{n,y} \right)^{\frac{\hat{\gamma}_{n,y}}{\hat{\gamma}_{n,y} - \gamma_y}} = \exp \left(\frac{\hat{\gamma}_{n,y}}{\hat{\gamma}_{n,y} - \gamma_y} \ln \left(\frac{\gamma_y}{\hat{\gamma}_{n,y}} \hat{\sigma}_{n,y} \right) \right) = \exp \left(\frac{\hat{\gamma}_{n,y}}{\hat{\gamma}_{n,y} - \gamma_y} o_{\mathbb{P}} \left(\delta_{k,n,y}^{-1} \right) \right) = 1 + o_{\mathbb{P}}(1).$$

For fixed $\omega \in \Omega, 0 < T < +\infty$ and n large enough, we may write $\sup_{t \in [0, T]} \left| g_{j,k,n,y}(t) \right| (\omega) \leq \max \left(\left| g_{j,k,n,y}(t_{n,y}) \right| (\omega), \left| g_{j,k,n,y}(T) \right| (\omega) \right)$. By Taylor expansion and multivariate Δ -method with (31) and (44), one readily shows that for $0 < T < +\infty$ fixed, $g_{j,k,n,y}(T) = o_{\mathbb{P}}(1/\delta_{k,n,y})$. Now, we compute

$$g_{j,k,n,y}(t_{n,y}) = t_{n,y} \left(\frac{\hat{\gamma}_{n,y}}{\gamma_y} - 1 \right).$$

Then, using (31) and (44) once more yields $g_{j,k,n,y}(t_{n,y}) = o_{\mathbb{P}}(1/\delta_{k,n,y})$. \square

Lemma 5. Let $M \in (0, +\infty)$. Let $1 \leq j \neq k \leq d$, $y \in E$ and $1 \ll k_n \ll n$ be an intermediate sequence. Under Assumptions (3), (6), (I.4), (II.2), (II.3), (III.1), we define $1 \ll \delta_{0,n,y} \ll \tilde{\delta}_{0,n,y} = \sqrt{n \bar{\alpha}_n \psi_y(h)}$ and $1 \ll \delta_{-1,n,y} \ll \tilde{\delta}_{-1,n,y} = \left(\frac{n}{k_n}\right)^{\mu_y}$ as in Eq. (21) with $\mu_y > 0$ as in Condition (6). Then, it holds that

$$\min(\delta_{0,n,y}, \delta_{-1,n,y}) \sup_{\mathbf{x} \in [0, M]^2} \left| \lambda_{j,k,y}(\mathbf{x}) - \hat{\lambda}_{j,k,n,y}(\mathbf{x}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Proof of Lemma 5. By using Eq. (5), we write

$$\sup_{\mathbf{x} \in [0, M]^2} \left| \lambda_{j,k,y}(\mathbf{x}) - \hat{\lambda}_{j,k,n,y}(\mathbf{x}) \right| = \sup_{\mathbf{x} \in [0, M]^2} \left| L_{j,k,y}(\mathbf{x}) - \hat{L}_{j,k,n,y}(\mathbf{x}) \right|.$$

For n large enough, one can get $\frac{k_n}{n} M \leq 1$. Then, the triangular inequality allows us to bound the latter quantity by

$$\sup_{\mathbf{x} \in [0, M]^2} \left| L_{j,k,y}(\mathbf{x}) - \tilde{L}_{j,k,n,y}(\mathbf{x}) \right| + \sup_{\mathbf{x} \in [0, M]^2} \left| \hat{L}_{j,k,n,y}(\mathbf{x}) - \tilde{L}_{j,k,n,y}(\mathbf{x}) \right|,$$

where $\tilde{L}_{j,k,n,y}(\mathbf{x}) := \frac{n}{k_n} \left(1 - C_{j,k,y}(\mathbf{1} - \frac{k_n}{n} \mathbf{x}) \right)$. Concerning the second term, since $\bar{\alpha}_n = o(1)$, it readily follows from (32) that for any deterministic $1 \ll \delta_{0,n,y} \ll \tilde{\delta}_{0,n,y} \ll \sqrt{n \psi_y(h_n)}$,

$$\delta_{0,n,y} \sup_{\mathbf{x} \in [0, M]^2} \left| \hat{L}_{j,k,n,y}(\mathbf{x}) - \tilde{L}_{j,k,n,y}(\mathbf{x}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Lastly, using the rate condition in (6), we have $\delta_{-1,n,y} \sup_{\mathbf{x} \in [0, M]^2} \left| L_{j,k,y}(\mathbf{x}) - \tilde{L}_{j,k,n,y}(\mathbf{x}) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. \square

Conclusion

In this paper, we present a semi-parametric method for estimating functional multivariate L^1 -expectiles at extreme risk levels (i.e., functional L^1 -MEEs). Going beyond Beck et al. [3], we establish, in this functional setting, the consistency with rate of the approximated loss function by using empirical kernel-based estimators for the tail index, the tail ratio and the upper tail dependence function. Then, we couple the loss function estimation with a gradient descent algorithm (e.g., BFGS-family). As a result, we give the consistency with rate of the approximated optimum problem for solving functional L^1 -MEEs.

A possible line of research would be to consider the L^p -norm in Eq. (7), i.e., to consider the L^p -expectiles defined in Maume-Deschamps et al. [33] in this functional setting. The question of their estimation may also be approached by the same semi-parametric methodology although it would display more intricate algebra. Along the same lines, one may consider a possible extension of Maume-Deschamps et al. [34] to functional extreme Σ -expectiles for $\Sigma \neq \mathbf{1}$. While the specific case of one dominant marginal is studied in Maume-Deschamps et al. [34], an interesting problem is to remove the multivariate regular variation hypothesis on the random vector \mathbf{X} with equivalent marginals tails. Moreover, an ambitious future work would be to connect our inference method with the online optimization theory. Indeed, we propose in our work to simultaneously estimate and optimize the loss function associated to the considered functional L^1 -MEEs, which may essentially be seen as updating with new data-points a multidimensional optimization problem. Finally, an essential future work will be to apply our estimated extreme multivariate L^1 -expectiles on real data. Most of the applications in the expectiles literature with functional covariate are motivated by spectrometry and finance. In Girard et al. [23], the authors estimate extreme univariate expectile levels of the maximum hourly log-return of Bitcoin price given the hourly log-returns from the day before. An interesting multivariate extension would be to jointly consider the Bitcoin and other currencies during the same time length. Another natural application field of the proposed extreme risk measure could be the capital allocation problem, to balance the risk level between the portfolio’s components.

Acknowledgments

The authors sincerely express their gratitude to two anonymous referees for their valuable comments and remarks which improve the quality of the present work. This work has been supported by the project ANR McLaren, France (ANR-20-CE23-0011). This work has been partially supported by the French government, through the 3IA Côte d’Azur Investments in the Future project managed by the National Research Agency (ANR), France with the reference number ANR-19-P3IA-0002.

Appendix. Regular variation framework

Definition A.1 (Multivariate Regular Variation). A random vector $\mathbf{X} \in \mathbb{R}^d$ is regularly varying with index $\alpha \geq 0$ if there exists a random vector θ on the unit sphere \mathbb{S}^{d-1} such that for any $x \in (0, +\infty)$, the following vague convergence holds:

$$\frac{\mathbb{P} \left(\|\mathbf{X}\| > xz, \frac{\mathbf{X}}{\|\mathbf{X}\|} \in \cdot \right)}{\mathbb{P}(\|\mathbf{X}\| > z)} \xrightarrow[z \rightarrow +\infty]{} x^{-\alpha} \mathbb{P}(\theta \in \cdot).$$

Definition A.2 (Second Order Regular Variation De Haan and Ferreira [11]). Let $\gamma > 0$ and $\rho \leq 0$. We say that $U \in 2RV_{-\gamma, \rho}(+\infty)$ if there exists an auxiliary function A with ultimately constant sign and $\lim_{t \rightarrow +\infty} A(t) = 0$ such that

$$\lim_{t \rightarrow +\infty} \frac{1}{A(t)} \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = x^\gamma H_\rho(x) := x^\gamma \int_1^x u^{\rho-1} du, \quad x > 0. \tag{A.1}$$

Note that for $x > 0$, $H_\rho(x) = \ln(x)\mathbf{1}_{\{\rho=0\}} + \frac{x^\rho - 1}{\rho}\mathbf{1}_{\{\rho < 0\}}$. By Theorem 2.3.9 in De Haan and Ferreira [11], this is equivalent to

$$\lim_{t \rightarrow +\infty} \frac{1}{A\left(\frac{1}{\overline{F}(t)}\right)} \left(\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-\frac{1}{\gamma}} \right) = H_{\rho, \gamma}(x) := x^{-\frac{1}{\gamma}} \frac{x^\rho - 1}{\gamma \rho}, \quad x > 0. \tag{A.2}$$

Namely, $U \in 2RV_{\gamma, \rho}(+\infty) \iff \overline{F} \in 2RV_{-\frac{1}{\gamma}, \frac{\rho}{\gamma}}(+\infty)$. Besides, on the auxiliary functions level, Theorem 2.3.3 in De Haan and Ferreira [11] provides $A \in RV_\rho(+\infty)$, $A \circ \frac{1}{\overline{F}} \in RV_{\frac{\rho}{\gamma}}(+\infty)$.

Remark 4. Condition (H1) entails a second order regular variation behavior. Note that $c_{1,y} \equiv 1$ for any $y \in E$. For any $1 \leq j \leq d$, since $\overline{F}_{j,y} \in RV_{-1/\gamma_y}(+\infty)$, there exists $\ell_{\overline{F}_{j,y}} \in RV_0(+\infty)$, such that

$$\overline{F}_{j,y}(x) = x^{-1/\gamma_y} \ell_{\overline{F}_{j,y}}(x), \quad x > 0. \tag{A.3}$$

Also, we have $U_{j,y} \in RV_{\gamma_y}(+\infty)$ and thus, there also exists a slowly varying function $\ell_{U_{j,y}} := \ell_{j,y}$ such that $U_{j,y}(x) = x^{\gamma_y} \ell_{j,y}(x)$, for $x > 0$. Condition (H2) is inherited from Beck et al. [3], Maume-Deschamps et al. [34] and states that each marginal behaves the same way in the extreme regime. Moreover, since $\overline{F}_{j,y} \in 2RV_{-1/\gamma_y, \rho_{j,y}/\gamma_y}(+\infty)$, it follows that $U_{j,y} \in 2RV_{\gamma_y, \rho_{j,y}}(+\infty)$. According to Definition A.2, there exists auxiliary functions $A_{j,y} := A_{U_{j,y}} \in RV_{\rho_{j,y}}(+\infty)$ satisfying Eq. (A.1).

Remark 5. Examples of distributions satisfying the second order regular variation condition are:

- Log-gamma distribution defined as the exponential of the sum of two independent standard exponential random variables: $\overline{F}(x) = \frac{1}{x} (1 + \log(x)) \in 2RV_{-1,0}(+\infty)$ with $A(1/\overline{F}(t)) = 1/\log(t)$ (see Geluk et al. [20]).
- For $x > e$, $\overline{F}(x) = \frac{\log x}{x} e \in 2RV_{-1,0}(+\infty)$, with $A(1/\overline{F}(t)) = 1/\log(t)$.
- Hall-Weiss class: $\overline{F}(x) = \frac{1}{2} x^{-\alpha} (1 + x^\rho) \in 2RV_{-\alpha, \rho}(+\infty)$ for $\alpha > 0$ and $\rho < 0$ with $A(1/\overline{F}(t)) = t^\rho$ (see Geluk et al. [20]).
- Cauchy distribution: $\overline{F}(x) = \frac{1}{\pi} \tan^{-1}(\frac{1}{x}) + \frac{1}{2}(1 - \text{sign}(x)) \in 2RV_{-1,-2}(+\infty)$ (see Gardes and Girard [19]).
- Fréchet distribution: $\overline{F}(x) = 1 - e^{-x^{-\frac{1}{\gamma}}} \in 2RV_{-\gamma, -\gamma}(+\infty)$, $\gamma > 0$ (see Gardes and Girard [19]).
- Burr distribution: $\overline{F}(x) = (1 + x^\tau)^{-\lambda} \in 2RV_{-\frac{1}{\tau\lambda}, -\tau}(+\infty)$, $\tau, \lambda > 0$ (see Gardes and Girard [19]).
- Student t_ν , $\nu > 1$, distribution with density $f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$. Then, $\overline{F} \in 2RV_{-\nu, -2}(+\infty)$.

The interested reader can find more 2RV type distributions in Section 5 of Mao and Hu [30].

References

[1] C. Acerbi, Spectral measures of risk: A coherent representation of subjective risk aversion, *J. Bank. Financ.* 26 (7) (2002) 1505–1518.
 [2] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, *Math. Finance* 9 (3) (1999) 203–228.
 [3] N. Beck, E. Di Bernardino, M. Mailhot, Semi-parametric estimation of multivariate extreme expectiles, *J. Multivariate Anal.* 184 (2021) 104758.
 [4] F. Bellini, E.D. Bernardino, Risk management with expectiles, *Eur. J. Finance* 23 (6) (2017) 487–506.
 [5] F. Bellini, B. Klar, A. Müller, E. Rosazza Gianin, Generalized quantiles as risk measures, *Insurance Math. Econom.* 54 (C) (2014) 41–48.
 [6] N.H. Bingham, C.M. Goldie, J.L. Teugels, Regular Variation, in: *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, 1987.
 [7] A. Daouia, L. Gardes, S. Girard, On kernel smoothing for extremal quantile regression, *Bernoulli* 19 (5B) (2013) 2557–2589.
 [8] A. Daouia, L. Gardes, S. Girard, A. Lekina, Kernel estimators of extreme level curves, *Test* 20 (2) (2011) 311–333.
 [9] A. Daouia, S. Girard, G. Stupfler, Estimation of tail risk based on extreme expectiles, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 80 (2) (2018) 263–292.
 [10] A. Daouia, S. Girard, G. Stupfler, Tail expectile process and risk assessment, *Bernoulli* 26 (1) (2020) 531–556.
 [11] L. De Haan, A. Ferreira, *Extreme Value Theory: An Introduction*, vol. 21, Springer, 2006.
 [12] E. Di Bernardino, V. Maume-Deschamps, C. Prieur, Estimating a bivariate tail: A copula based approach, *J. Multivariate Anal.* 119 (2013) 81–100.
 [13] J.H. Einmahl, A. Krajina, J. Segers, A method of moments estimator of tail dependence, *Bernoulli* 14 (4) (2008) 1003–1026.
 [14] J.H.J. Einmahl, A. Krajina, J. Segers, An M-estimator for tail dependence in arbitrary dimensions, *Ann. Statist.* 40 (3) (2012) 1764–1793.
 [15] H.-B. Fang, K.-T. Fang, S. Kotz, The meta-elliptical distributions with given marginals, *J. Multivariate Anal.* 82 (1) (2002) 1–16.
 [16] F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Rate of uniform consistency for nonparametric estimates with functional variables, *J. Statist. Plann. Inference* 140 (2) (2010) 335–352.
 [17] F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: Inference and practical aspects, *Aust. N. Z. J. Stat.* 49 (3) (2007) 267–286.
 [18] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice* (Springer Series in Statistics), Springer-Verlag, Berlin, Heidelberg, 2006.
 [19] L. Gardes, S. Girard, Functional kernel estimators of large conditional quantiles, *Electron. J. Stat.* 6 (none) (2012) 1715–1744.
 [20] J. Geluk, L. de Haan, S. Resnick, C. Stărică, Second-order regular variation, convolution and the central limit theorem, *Stochastic Process. Appl.* 69 (2) (1997) 139–159.
 [21] I. Gijbels, M. Omelka, N. Veraverbeke, Multivariate and functional covariates and conditional copulas, *Electron. J. Stat.* 6 (none) (2012) 1273–1306.
 [22] S. Girard, G.C. Stupfler, A. Usseglio-Carleve, Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models, *Ann. Statist.* 49 (6) (2021) 3358–3382.

- [23] S. Girard, G. Stupfler, A. Usseglio-Carleve, Functional estimation of extreme conditional expectiles, *Econom. Stat.* 21 (2022) 131–158.
- [24] S. Girard, G. Stupfler, A. Usseglio-Carleve, Nonparametric extreme conditional expectile estimation, *Scand. J. Stat.* 49 (1) (2022) 78–115.
- [25] T. Gneiting, Making and evaluating point forecasts, *J. Amer. Statist. Assoc.* 106 (494) (2011) 746–762.
- [26] L. de Haan, C. Zhou, Trends in extreme value indices, *J. Amer. Statist. Assoc.* 116 (535) (2021) 1265–1279.
- [27] K. Herrmann, M. Hofert, M. Mailhot, Multivariate geometric expectiles, *Scand. Actuar. J.* 2018 (7) (2018) 629–659.
- [28] L. Hua, H. Joe, Second order regular variation and conditional tail expectation of multiple risks, *Insurance Math. Econom.* 49 (3) (2011) 537–546.
- [29] O. Kallenberg, Foundations of Modern Probability, second ed., in: Probability and its Applications (New York), Springer-Verlag, New York, 2002, p. xx+638.
- [30] T. Mao, T. Hu, Second-order properties of the Haezendonck–Goovaerts risk measure for extreme risks, *Insurance Math. Econom.* 51 (2) (2012) 333–343.
- [31] V. Maume-Deschamps, E. Masiello, A. Cuberos, High level quantile estimation of sums of risks, *Dependence Model.* 3 (2015) 141.
- [32] V. Maume-Deschamps, D. Rullière, K. Said, Impact of dependence on some multivariate risk indicators, *Methodol. Comput. Appl. Probab.* 19 (2017) 395–427.
- [33] V. Maume-Deschamps, D. Rullière, K. Said, Multivariate extensions of expectiles risk measures, *Dependence Model.* (2017).
- [34] V. Maume-Deschamps, D. Rullière, K. Said, Extremes for multivariate expectiles, *Stat. Risk Model.* 35 (3–4) (2018) 111–140.
- [35] W.K. Newey, J.L. Powell, Asymmetric least squares estimation and testing, *Econometrica* 55 (4) (1987) 819–847.
- [36] J. Nocedal, S.J. Wright, Numerical Optimization, 2e ed., Springer, New York, NY, USA, 2006.
- [37] S. Resnick, T.V. Mikosch, S.M. Robinson, Heavy-Tail Phenomena: Probabilistic and Statistical Modeling, in: Heavy-tail phenomena: probabilistic and statistical modeling, (vol. 10) Springer, 2007.
- [38] R. Schmidt, U. Stadtmüller, Non-parametric estimation of tail dependence, *Scand. J. Stat.* 33 (2) (2006) 307–335.
- [39] J. Segers, Asymptotics of empirical copula processes under non-restrictive smoothness assumptions, *Bernoulli* 18 (3) (2012) 764–782.
- [40] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Stat. Univ. Paris* 8 (1959) 229–231.
- [41] A.v.d. Vaart, Asymptotic Statistics, in: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.
- [42] C. Zhu, R.H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, *ACM Trans. Math. Software* 23 (4) (1997) 550–560.
- [43] J.F. Ziegel, Coherence and elicibility, *Math. Finance* 26 (4) (2016) 901–918.