



HAL
open science

A Weissman-type estimator of the conditional marginal expected shortfall

Yuri Goegebeur, Armelle Guillou, Nguyen Khanh Le Ho, Jing Qin

► **To cite this version:**

Yuri Goegebeur, Armelle Guillou, Nguyen Khanh Le Ho, Jing Qin. A Weissman-type estimator of the conditional marginal expected shortfall. 2020. hal-02613135v2

HAL Id: hal-02613135

<https://hal.science/hal-02613135v2>

Preprint submitted on 12 Dec 2020 (v2), last revised 24 Sep 2021 (v4)

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

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

A Weissman-type estimator of the conditional marginal expected shortfall

Yuri Goegebeur⁽¹⁾, Armelle Guillou⁽²⁾, Nguyen Khanh Le Ho⁽¹⁾, Jing Qin⁽¹⁾

⁽¹⁾ Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

⁽²⁾ Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

Abstract

The marginal expected shortfall is an important risk measure in finance, which has been extended recently to the case where the random variables of main interest $(Y^{(1)}, Y^{(2)})$ are observed together with a covariate $X \in \mathbb{R}^d$. This leads to the concept of conditional marginal expected shortfall. It is defined as $\theta_p(x_0) = \mathbb{E}[Y^{(1)} | Y^{(2)} \geq Q_{Y^{(2)}}(1-p|x_0); x_0]$, where p is small and $Q_{Y^{(2)}}(\cdot|x_0)$ denotes the conditional quantile function of $Y^{(2)}$, given $X = x_0$. In this paper, we propose an estimator for $\theta_p(x_0)$ allowing extrapolation outside the $Y^{(2)}$ -data range, i.e., valid for $p < 1/n$. The main asymptotic properties of this estimator have been established, using empirical processes arguments combined with the multivariate extreme value theory. The finite sample behavior of the proposed estimator is evaluated with a simulation experiment, and the practical applicability is illustrated on a medical dataset.

Keywords: Conditional marginal expected shortfall, extrapolation, Pareto-type distribution.

1 Introduction

Risk measures have been extensively studied in the literature, but mainly in the univariate framework. The extension to the multivariate context is very recent and crucial for many applications. For instance, in finance, the Marginal Expected Shortfall (MES) is an important tool when measuring the systemic risk of financial institutions. It is defined as $\theta_p := \mathbb{E}[Y^{(1)} | Y^{(2)} \geq Q_{Y^{(2)}}(1-p)]$ where $Y^{(1)}$ denotes the loss of the equity return of a financial institution, while $Y^{(2)}$ is that of the entire market and $Q_{Y^{(2)}}(1-p)$ the corresponding $(1-p)$ -quantile. This MES measure has been estimated in Cai et al. (2015) using extreme value arguments in the case $p = O(1/n)$, where n is the sample size. Recently, Goegebeur et al. (2020) have considered also the estimation of the marginal expected shortfall, but this time in case where the random variables of main interest $(Y^{(1)}, Y^{(2)})$ are recorded together with a random covariate $X \in \mathbb{R}^d$. This leads to the concept of conditional marginal expected shortfall, given $X = x_0$, and defined as

$$\theta_p(x_0) = \mathbb{E} \left[Y^{(1)} \middle| Y^{(2)} \geq Q_{Y^{(2)}}(1-p|x_0); x_0 \right].$$

In the sequel, we will denote by $F_j(\cdot|x)$ the continuous conditional distribution function of $Y^{(j)}, j = 1, 2$, given $X = x$, and use the notation $\bar{F}_j(\cdot|x)$ for the conditional survival function and $U_j(\cdot|x)$ for the associated tail quantile function defined as $U_j(\cdot|x) = \inf\{y : F_j(y|x) \geq 1 - 1/\cdot\}$. Also, we will define by f_X the density function of the covariate X and by x_0 a reference position such that $x_0 \in \text{Int}(S_X)$, the interior of the support $S_X \subset \mathbb{R}^d$ of f_X , which is assumed to be non-empty. Considering $(Y_i^{(1)}, Y_i^{(2)}, X_i), i = 1, \dots, n$, independent copies of $(Y^{(1)}, Y^{(2)}, X)$, Goegebeur et al. (2020) proposed the following nonparametric estimator for $\theta_{k/n}(x_0)$

$$\bar{\theta}_{\frac{k}{n}}(x_0) := \frac{\frac{1}{k} \sum_{i=1}^n K_{h_n}(x_0 - X_i) Y_i^{(1)} \mathbb{1}_{\{Y_i^{(2)} \geq \hat{U}_2(n/k|x_0)\}}}{\hat{f}_n(x_0)},$$

where $K_{h_n}(\cdot) := K(\cdot/h_n)/h_n^d$, with K a joint density function on \mathbb{R}^d , k an intermediate sequence such that $k \rightarrow \infty$ with $k/n \rightarrow 0$, h_n a positive non-random sequence of bandwidths with $h_n \rightarrow 0$ if $n \rightarrow \infty$, $\mathbb{1}_A$ the indicator function on the event A , and $\hat{f}_n(x_0) := (1/n) \sum_{i=1}^n K_{h_n}(x_0 - X_i)$ is a classical kernel density estimator. Here, $\hat{U}_2(\cdot|x_0)$ is an estimator for $U_2(\cdot|x_0)$, defined as $\hat{U}_2(\cdot|x_0) := \inf\{y : \hat{F}_{n,2}(y|x_0) \geq 1 - 1/\cdot\}$ where

$$\hat{F}_{n,2}(y|x_0) := \frac{\frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{Y_i^{(2)} \leq y\}}}{\hat{f}_n(x_0)}. \quad (1)$$

The asymptotic behavior of $\bar{\theta}_{k/n}(x_0)$ has been established by Goegebeur et al. (2020) and is recalled in Theorem 5.1 in Section 5. Due to the conditions $k, n \rightarrow \infty$ with $k/n \rightarrow 0$, the $Y^{(2)}$ -quantile is intermediate, and thus the estimator $\bar{\theta}_{k/n}(x_0)$ cannot be used for extrapolation outside the $Y^{(2)}$ -data range. This reduces considerably the potential of applications of the estimator in practice. The aim of this paper is to solve this issue and to define a new estimator which allows extrapolation and thus which is valid for $p < 1/n$.

The remainder of the paper is organized as follows. In Section 2, we introduce our estimator for the conditional marginal expected shortfall allowing extrapolation and we establish its main asymptotic properties. The efficiency of our estimator is examined with a small simulation study in Section 3. Finally, in Section 4 we illustrate the performance of our estimator on the PIMA Indians Diabetes dataset. Some preliminary results are given in Section 5, whereas the proofs of the main results are postponed to Section 6.

2 Estimator and asymptotic properties

We assume that $Y^{(1)}$ and $Y^{(2)}$ are positive random variables, and that they follow a conditional Pareto-type model. Let RV_ψ denote the class of regularly varying functions at infinity with index ψ , i.e., positive measurable functions f satisfying $f(tx)/f(t) \rightarrow x^\psi$, as $t \rightarrow \infty$, for all $x > 0$. If $\psi = 0$, then we call f a slowly varying function at infinity.

Assumption (D) For all $x \in S_X$, the conditional survival function of $Y^{(j)}, j = 1, 2$, given $X = x$, satisfies

$$\bar{F}_j(y|x) = A_j(x)y^{-1/\gamma_j(x)} \left(1 + \frac{1}{\gamma_j(x)} \delta_j(y|x) \right),$$

where $A_j(x) > 0$, $\gamma_j(x) > 0$, and $|\delta_j(\cdot|x)|$ is normalized regularly varying at infinity with index $-\beta_j(x)$, $\beta_j(x) > 0$, i.e.,

$$\delta_j(y|x) = B_j(x) \exp \left(\int_1^y \frac{\varepsilon_j(u|x)}{u} du \right),$$

with $B_j(x) \in \mathbb{R}$ and $\varepsilon_j(y|x) \rightarrow -\beta_j(x)$ as $y \rightarrow \infty$. Moreover, we assume $y \rightarrow \varepsilon_j(y|x)$ to be a continuous function.

Clearly, Assumption (D) implies that $U_j(\cdot|x), j = 1, 2$, satisfy

$$U_j(y|x) = [A_j(x)]^{\gamma_j(x)} y^{\gamma_j(x)} (1 + a_j(y|x)) \quad (2)$$

where $a_j(y|x) = \delta_j(U_j(y|x)|x)(1 + o(1))$, and thus $|a_j(\cdot|x)| \in RV_{-\beta_j(x)\gamma_j(x)}$.

Now, to estimate $\theta_p(x_0)$ it is required to impose an assumption on the right-hand upper tail dependence of $(Y^{(1)}, Y^{(2)})$, conditional on a value of the covariate X . Let $R_t(y_1, y_2|x) := t\mathbb{P}(\bar{F}_1(Y^{(1)}|x) \leq y_1/t, \bar{F}_2(Y^{(2)}|x) \leq y_2/t | X = x)$.

Assumption (R) For all $x \in S_X$ we have as $t \rightarrow \infty$ and $h \downarrow 0$ that

$$R_t(y_1, y_2|x) \rightarrow R(y_1, y_2|x),$$

uniformly in $y_1, y_2 \in (0, T]$, for any $T > 0$, and in $x \in B(x_0, h)$.

Assuming (D) with $\gamma_1(x_0) < 1$ and (R), one can show (see Cai et al., 2015, Proposition 1) that

$$\lim_{p \rightarrow 0} \frac{\theta_p(x_0)}{U_1(1/p|x_0)} = - \int_0^\infty R(s, 1|x_0) ds^{-\gamma_1(x_0)},$$

from which the following approximation can be deduced

$$\theta_p(x_0) \sim \frac{U_1(1/p|x_0)}{U_1(n/k|x_0)} \theta_{\frac{k}{n}}(x_0) \sim \left(\frac{k}{np} \right)^{\gamma_1(x_0)} \theta_{\frac{k}{n}}(x_0).$$

Thus, to estimate $\theta_p(x_0)$, we need first to estimate $\gamma_1(x_0)$. We propose the following estimator

$$\hat{\gamma}_{1, k_1}(x_0) := \frac{\frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \left(\ln Y_i^{(1)} - \ln \hat{U}_1(n/k_1|x_0) \right) \mathbb{1}_{\{Y_i^{(1)} \geq \hat{U}_1(n/k_1|x_0)\}}}{\hat{f}_n(x_0)}, \quad (3)$$

based on an intermediate sequence k_1 such that $k_1 \rightarrow \infty$ with $k_1/n \rightarrow 0$. This sequence may be different from k , the one used in the estimator $\bar{\theta}_{\frac{k}{n}}(x_0)$, but the two sequences are linked together

(see Theorem 2.2 below). On the contrary, for convenience, we use the same bandwidth h_n and kernel K for both the estimation of $\gamma_1(x_0)$ and $\theta_p(x_0)$. Note that $\hat{\gamma}_{1,k_1}(x_0)$ is a local version of the Hill estimator (Hill, 1975) from the univariate extreme value context.

Now, we are able to define a Weissman-type estimator for $\theta_p(x_0)$, given by

$$\hat{\theta}_p(x_0) = \left(\frac{k}{np}\right)^{\hat{\gamma}_{1,k_1}(x_0)} \bar{\theta}_{\frac{k}{n}}(x_0).$$

Our aim in this paper is to establish the asymptotic behavior of $\hat{\theta}_p(x_0)$, which requires the asymptotic behavior of $\hat{\gamma}_{1,k_1}(x_0)$ in terms of the process on which the estimator $\bar{\theta}_{\frac{k}{n}}(x_0)$ is based on (see Theorem 5.1 in Section 5). To reach this goal, some assumptions due to the regression context are required. In particular, f_X and the functions appearing in $\bar{F}_j(y|x)$, $j = 1, 2$, are assumed to satisfy the following Hölder conditions. Let $\|\cdot\|$ denote some norm on \mathbb{R}^d .

Assumption (\mathcal{H}) *There exist positive constants M_{f_X} , M_{A_j} , M_{γ_j} , M_{B_j} , M_{ε_j} , η_{f_X} , η_{A_j} , η_{γ_j} , η_{B_j} and η_{ε_j} , where $j = 1, 2$, such that for all $x, z \in S_X$:*

$$\begin{aligned} |f_X(x) - f_X(z)| &\leq M_{f_X} \|x - z\|^{\eta_{f_X}}, \\ |A_j(x) - A_j(z)| &\leq M_{A_j} \|x - z\|^{\eta_{A_j}}, \\ |\gamma_j(x) - \gamma_j(z)| &\leq M_{\gamma_j} \|x - z\|^{\eta_{\gamma_j}}, \\ |B_j(x) - B_j(z)| &\leq M_{B_j} \|x - z\|^{\eta_{B_j}}, \\ \sup_{y \geq 1} |\varepsilon_j(y|x) - \varepsilon_j(y|z)| &\leq M_{\varepsilon_j} \|x - z\|^{\eta_{\varepsilon_j}}. \end{aligned}$$

We also impose a condition on the kernel function K , which is a standard condition in local estimation.

Assumption (\mathcal{K}) *K is a bounded density function on \mathbb{R}^d , with support S_K included in the unit ball in \mathbb{R}^d , with respect to the norm $\|\cdot\|$.*

Our first aim, now, is to show the weak convergence, denoted \rightsquigarrow , of the process based on $\hat{\gamma}_{1,k_1}(x_0)$, but in terms of the same process as the one used in Theorem 2.3 from Goegebeur et al. (2020).

Theorem 2.1 *Assume (\mathcal{D}), (\mathcal{H}), (\mathcal{K}), $x_0 \in \text{Int}(S_X)$ with $f_X(x_0) > 0$, and $y \rightarrow F_1(y|x_0)$, is strictly increasing. Consider sequences $k_1 \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, in such a way that $k_1/n \rightarrow 0$, $k_1 h_n^d \rightarrow \infty$, $h_n^{\eta_{\varepsilon_1}} \ln n/k_1 \rightarrow 0$, $\sqrt{k_1 h_n^d} h_n^{\eta_{f_X} \wedge \eta_{A_1}} \rightarrow 0$, $\sqrt{k_1 h_n^d} h_n^{\eta_{\gamma_1}} \ln n/k_1 \rightarrow 0$, $\sqrt{k_1 h_n^d} |\delta_1(U_1(n/k_1|x_0)|x_0)| \rightarrow 0$. Then we have,*

$$\sqrt{k_1 h_n^d} (\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)) \rightsquigarrow \frac{\gamma_1(x_0)}{f_X(x_0)} \left[\int_0^1 W(z, \infty) \frac{1}{z} dz - W(1, \infty) \right],$$

where $W(z, \infty)$ is a zero centered Gaussian process with covariance function

$$\mathbb{E}(W(z, \infty)W(\bar{z}, \infty)) = \|K\|_2^2 f_X(x_0) (z \wedge \bar{z}),$$

with $\|K\|_2 := \sqrt{\int_{\mathbb{R}^d} K^2(u) du}$.

Note that the variance of the limiting distribution of $\hat{\gamma}_{1,k_1}(x_0)$, after normalization, is given by $\gamma_1^2(x_0)\|K\|_2^2/f_X(x_0)$, compared to an asymptotic variance of γ_1^2 for the Hill estimator in the univariate context.

Before stating the weak convergence of $\hat{\theta}_p(x_0)$, we need to introduce a second order condition, usual in the extreme value context.

Assumption (S). *There exist $\beta > \gamma_1(x_0)$ and $\tau < 0$ such that, as $t \rightarrow \infty$ and $h \downarrow 0$ we have*

$$\sup_{x \in B(x_0, h)} \sup_{0 < y_1 < \infty, \frac{1}{2} \leq y_2 \leq 2} \frac{|R_t(y_1, y_2|x) - R(y_1, y_2|x_0)|}{y_1^\beta \wedge 1} = O(t^\tau).$$

Our final result is now the following.

Theorem 2.2 *Assume (D), (H), (K), (S) with $x \rightarrow R(y_1, y_2|x)$ being a continuous function, and $y \rightarrow F_j(y|x_0)$, $j = 1, 2$, are strictly increasing. Let $x_0 \in \text{Int}(S_X)$ such that $f_X(x_0) > 0$. Consider sequences $k = \lfloor n^\alpha \ell_1(n) \rfloor$, $k_1 = \lfloor n^{\alpha_1} \ell_2(n) \rfloor$ and $h_n = n^{-\Delta} \ell_3(n)$, where ℓ_1, ℓ_2 and ℓ_3 are slowly varying functions at infinity, with $\alpha \in (0, 1)$ and*

$$\alpha \leq \alpha_1 < \min \left(\frac{\alpha}{d} [d + 2(\eta_{f_X} \wedge \eta_{A_1} \wedge \eta_{\gamma_1})], \frac{\alpha + 2\gamma_1(x_0)\beta_1(x_0)}{1 + 2\gamma_1(x_0)\beta_1(x_0)} \right),$$

and

$$\begin{aligned} & \max \left(\frac{\alpha}{d + 2\gamma_1(x_0)(\eta_{A_1} \wedge \eta_{\gamma_1})}, \frac{\alpha}{d + 2(1 - \gamma_1(x_0))(\eta_{A_2} \wedge \eta_{\gamma_2} \wedge \eta_{B_2} \wedge \eta_{\varepsilon_2} \wedge \eta_{f_X})}, \right. \\ & \frac{\alpha}{d} - \frac{2(1 - \alpha)\gamma_1^2(x_0)\beta_1(x_0)}{d + d(\beta_1(x_0) + \varepsilon)\gamma_1(x_0)}, \frac{\alpha - 2(1 - \alpha)(\gamma_1(x_0) \wedge (\beta_2(x_0)\gamma_2(x_0)) \wedge (-\tau))}{d}, \\ & \left. \frac{\alpha_1}{d + 2(\eta_{f_X} \wedge \eta_{A_1} \wedge \eta_{\gamma_1})}, \frac{\alpha_1 - 2(1 - \alpha_1)\gamma_1(x_0)\beta_1(x_0)}{d} \right) < \Delta < \frac{\alpha}{d}. \end{aligned}$$

Then, for $\gamma_1(x_0) < 1/2$ and p satisfying $p \leq \frac{k}{n}$ such that $\frac{\ln k/(np)}{\sqrt{k_1 h_n^d}} \rightarrow 0$ and $\sqrt{\frac{k}{k_1}} \ln \frac{k}{np} \rightarrow r \in [0, \infty]$, we have

$$\begin{aligned} & \min \left(\sqrt{kh_n^d}, \frac{\sqrt{k_1 h_n^d}}{\ln k/(np)} \right) \left(\frac{\hat{\theta}_p(x_0)}{\theta_p(x_0)} - 1 \right) \\ & \rightsquigarrow \min(r, 1) \frac{\gamma_1(x_0)}{f_X(x_0)} \left(\int_0^1 W(y, \infty) \frac{1}{y} dy - W(1, \infty) \right) \\ & + \min \left(1, \frac{1}{r} \right) \left\{ -(1 - \gamma_1(x_0)) \frac{W(\infty, 1)}{f_X(x_0)} + \frac{1}{f_X(x_0)} \frac{\int_0^\infty W(y, 1) dy^{-\gamma_1(x_0)}}{\int_0^\infty R(y, 1|x_0) dy^{-\gamma_1(x_0)}} \right\}, \end{aligned}$$

where $W(y_1, y_2)$ is a zero centered Gaussian process with covariance function

$$\mathbb{E}(W(y_1, y_2)W(\bar{y}_1, \bar{y}_2)) = \|K\|_2^2 f_X(x_0) R(y_1 \wedge \bar{y}_1, y_2 \wedge \bar{y}_2|x_0),$$

$W(y, \infty)$ is the limiting process of Theorem 2.1, and $W(\infty, y)$ is a zero centered Gaussian process with covariance function

$$\mathbb{E}(W(\infty, y)W(\infty, \bar{y})) = \|K\|_2^2 f_X(x_0)(y \wedge \bar{y}).$$

The variance of the limiting random variable in Theorem 2.2, denoted \mathbb{W} , is given by

$$\begin{aligned} \text{Var}(\mathbb{W}) = & \frac{\|K\|_2^2}{f_X(x_0)} \left\{ (\min(r, 1))^2 \gamma_1^2(x_0) \right. \\ & + \left(\min\left(1, \frac{1}{r}\right) \right)^2 \left[\gamma_1^2(x_0) - 1 - c^2 \int_0^\infty R(s, 1|x_0) ds^{-2\gamma_1(x_0)} \right] \\ & + 2 \min\left(r, \frac{1}{r}\right) \gamma_1(x_0) \left[(1 - \gamma_1(x_0) + c) R(1, 1|x_0) \right. \\ & \left. \left. - \int_0^1 \left(1 - \gamma_1(x_0) + cs^{-\gamma_1(x_0)}(1 - \gamma_1(x_0) - \gamma_1(x_0) \ln s)\right) \frac{R(s, 1|x_0)}{s} ds \right] \right\}, \quad (4) \end{aligned}$$

where $c := \left(\int_0^\infty R(s, 1|x_0) ds^{-\gamma_1(x_0)}\right)^{-1}$.

3 Simulation experiment

In this section, we illustrate the finite-sample performance of our conditional marginal expected shortfall estimator with a simulation experiment. To this aim, we consider the three models already used in Goegebeur et al. (2020), but this time in the case where $p < 1/n$, i.e., allowing extrapolation outside the $Y^{(2)}$ -data range. These models are the followings:

Model 1: The conditional logistic copula model, defined as

$$C(u_1, u_2|x) = e^{-[(-\ln u_1)^x + (-\ln u_2)^x]^{1/x}}, \quad u_1, u_2 \in [0, 1], \quad x \geq 2. \quad (5)$$

In this model, X is uniformly distributed in the interval $[2, 10]$, and the marginal distribution functions of $Y^{(1)}$ and $Y^{(2)}$ are Fréchet distributions:

$$F_j(y) = e^{-y^{-1/\gamma_j}}, \quad y > 0,$$

$j = 1, 2$. We set $\gamma_1 = 0.25$ and $\gamma_2 = 0.5$. It can be shown that this model satisfies Assumption (S) with $R(y_1, y_2|x) = y_1 + y_2 - (y_1^x + y_2^x)^{1/x}$, $\tau = -1$ and $\beta = 1 - \varepsilon$ for some small $\varepsilon > 0$.

Model 2: The conditional distribution of $(Y^{(1)}, Y^{(2)})$ given $X = x$ is that of

$$(|Z_1|^{\gamma_1(x)}, |Z_2|^{\gamma_2(x)}),$$

where (Z_1, Z_2) follow a bivariate standard Cauchy distribution with density function

$$f(z_1, z_2) = \frac{1}{2\pi} (1 + z_1^2 + z_2^2)^{-3/2}, \quad (z_1, z_2) \in \mathbb{R}^2.$$

Here, X is uniformly distributed in $[0, 1]$ and

$$\begin{aligned}\gamma_1(x) &= 0.4 [0.1 + \sin(\pi x)] \left[1.1 - 0.5e^{-64(x-0.5)^2} \right], \\ \gamma_2(x) &= 0.1 + 0.1x.\end{aligned}$$

Again, Assumption (\mathcal{S}) is satisfied for this model with $R(y_1, y_2|x) = y_1 + y_2 - \sqrt{y_1^2 + y_2^2}$, $\tau = -1$ and $\beta = 2$ (see, e.g., Cai et al., 2015, in the context without covariates).

Model 3: We consider again the conditional logistic copula model defined in (5) but this time with conditional Burr distributions for the marginal distribution functions of $Y^{(1)}$ and $Y^{(2)}$, i.e.,

$$F_j(y|x) = 1 - \left(\frac{\beta_j}{\beta_j + y^{\tau_j(x)}} \right)^{\lambda_j}, \quad y > 0; \quad \beta_j, \lambda_j, \tau_j(x) > 0,$$

$j = 1, 2$. We set $\beta_1 = \beta_2 = 1$, $\lambda_1 = 1$, $\lambda_2 = 0.5$, and

$$\tau_1(x) = 2e^{0.2x}, \quad \tau_2(x) = 8/\sin(0.3x).$$

Similarly to Model 1, this model satisfies (\mathcal{S}) .

Note that for all these models, Assumption (\mathcal{D}) is satisfied since all the marginal conditional distributions are standard examples from this class of heavy-tailed distributions (see, e.g., Beirlant et al., 2009, Table 1).

We simulate 500 datasets of size $n = 500$ and 1000 from each model. For each sample, we compute $\hat{\theta}_p(x_0)$ for two different values of p : $1/(2n)$ and $1/(5n)$, and three different sets of values of the covariate position x_0 : $\{3, 5, 7\}$ for Model 1 and Model 3, and $\{0.3, 0.5, 0.8\}$ for Model 2.

Concerning the kernel function K , we use the bi-quadratic function

$$K(x) = \frac{15}{16}(1 - x^2)^2 \mathbb{1}_{\{x \in [-1, 1]\}},$$

for both the estimation of $\hat{\gamma}_{1, k_1}(x_0)$ and $\bar{\theta}_{k/n}(x_0)$. To compute these estimators, we need also to select a bandwidth h_n . To this aim, we use the cross-validation procedure introduced by Yao (1999), and already used in the extreme value framework by Daouia et al. (2011, 2013) and Escobar-Bach et al. (2018a), and defined as:

$$h_{cv} := \operatorname{argmin}_{h_n \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left(\mathbb{1}_{\{Y_i^{(2)} \leq Y_j^{(2)}\}} - \hat{F}_{n, h_n, 2, -i}(Y_j^{(2)} | X_i) \right)^2,$$

where \mathcal{H} is the grid of values defined as $R_X \times \{0.05, 0.10, \dots, 0.30\}$, with R_X the range of the covariate X , and

$$\hat{F}_{n, h_n, 2, -i}(y|x) := \frac{\sum_{k=1, k \neq i}^n K_{h_n}(x - X_k) \mathbb{1}_{\{Y_k^{(2)} \leq y\}}}{\sum_{k=1, k \neq i}^n K_{h_n}(x - X_k)}.$$

Concerning now, the choice of the sequence k_1 in the estimation of $\gamma_1(x_0)$, a graphical assessment is used. The retained value of k_1 corresponds to the smallest value of k after which the median of the estimates $\hat{\gamma}_{1,k}(x_0)$ over the 500 replications exhibit a stable part. This choice is common in extreme value theory where we search at a plateau of an extreme value estimator as a function of the intermediate sequence.

Figures 1 to 6 display the boxplots of the ratios between the estimates $\hat{\theta}_p(x_0)$ and the true value $\theta_p(x_0)$ based on the 500 replications for the different values of x_0 (corresponding to the rows) and the two sample sizes (corresponding to the columns: $n = 500$ on the left, $n = 1000$ on the right). Figures 1 and 2 correspond to Model 1 for $p = 1/(2n)$ and $1/(5n)$, respectively. Similarly, Figures 3 and 4 correspond to Model 2, and Figures 5 and 6 to Model 3.

Based on these simulations, we can draw the following conclusions:

- Our estimator $\hat{\theta}_p(x_0)$ performs quite well in all the situations, although its efficiency depends obviously on the model and also on the covariate position. This is expected because as is clear from our models, the marginal distributions in Model 1 do not depend on the covariates but the dependence structure does. On the contrary, the dependence structure in Model 2 does not depend on the covariate but the marginal distributions depend on x_0 , and in Model 3 both of them depend on x_0 . Thus, Model 1 is less challenging than the two other models;
- Note that the estimation in Model 2 is difficult and depends a lot on the value of the covariate. This can be explained by the fact that the plot of $\gamma_1(x)$ as a function of x exhibits two local maxima, one of which being close to 0.3, and a local minimum at 0.5. Thus, we get an underestimation of $\theta_p(x_0)$ near the local maxima and an overestimation of $\theta_p(x_0)$ near the local minima, due to the local nature of the estimation. Outside these neighborhoods, the estimation is without bias. This is the case for $x_0 = 0.8$.
- As expected, the smaller p is, the more difficult is the estimation, due to an important extrapolation outside the $Y^{(2)}$ -data range. This results in an increase of the variability of the estimates. Note that the values $p = 1/(2n)$ and $p = 1/(5n)$ correspond to quite severe extrapolations since the estimation is done locally, and the local number of observations is much smaller than n .

4 Application to Indians diabetes data

In this section we illustrate the developed methodology on the PIMA Indians Diabetes dataset. This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases, and contains observations on nine variables, for 768 subjects aged 21 to 81 years. We estimate the conditional marginal expected shortfall for the diastolic blood pressure, $Y^{(1)}$, given that the body mass index, BMI, $Y^{(2)}$, exceeds a large quantile and given a value for

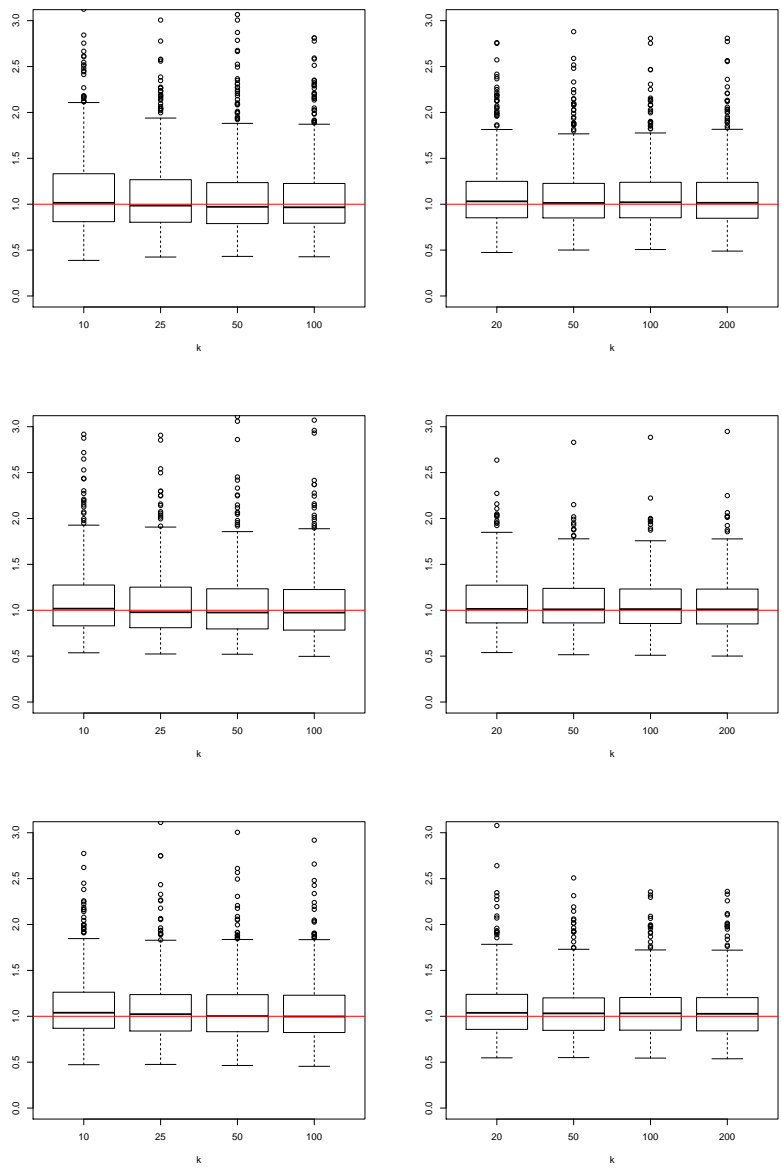


Figure 1: Boxplots of $\hat{\theta}_p(x_0)/\theta_p(x_0)$ for Model 1 based on 500 replications of sample sizes $n = 500$ (left column) and $n = 1000$ (right column) for $p = 1/(2n)$. From top to bottom we have $x_0 = 3$, $x_0 = 5$ and $x_0 = 7$. The values of k are taken as 2%, 5%, 10% and 20% of n .

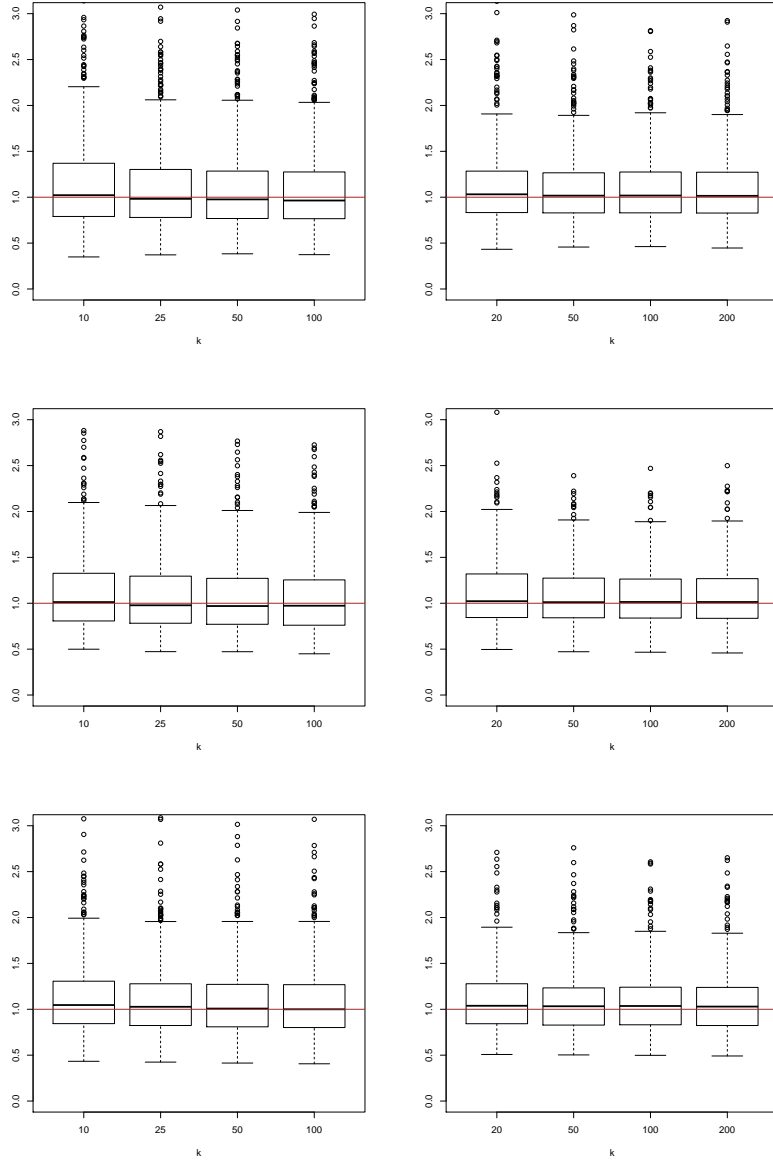


Figure 2: Boxplots of $\hat{\theta}_p(x_0)/\theta_p(x_0)$ for Model 1 based on 500 replications of sample sizes $n = 500$ (left column) and $n = 1000$ (right column) for $p = 1/(5n)$. From top to bottom we have $x_0 = 3$, $x_0 = 5$ and $x_0 = 7$. The values of k are taken as 2%, 5%, 10% and 20% of n .

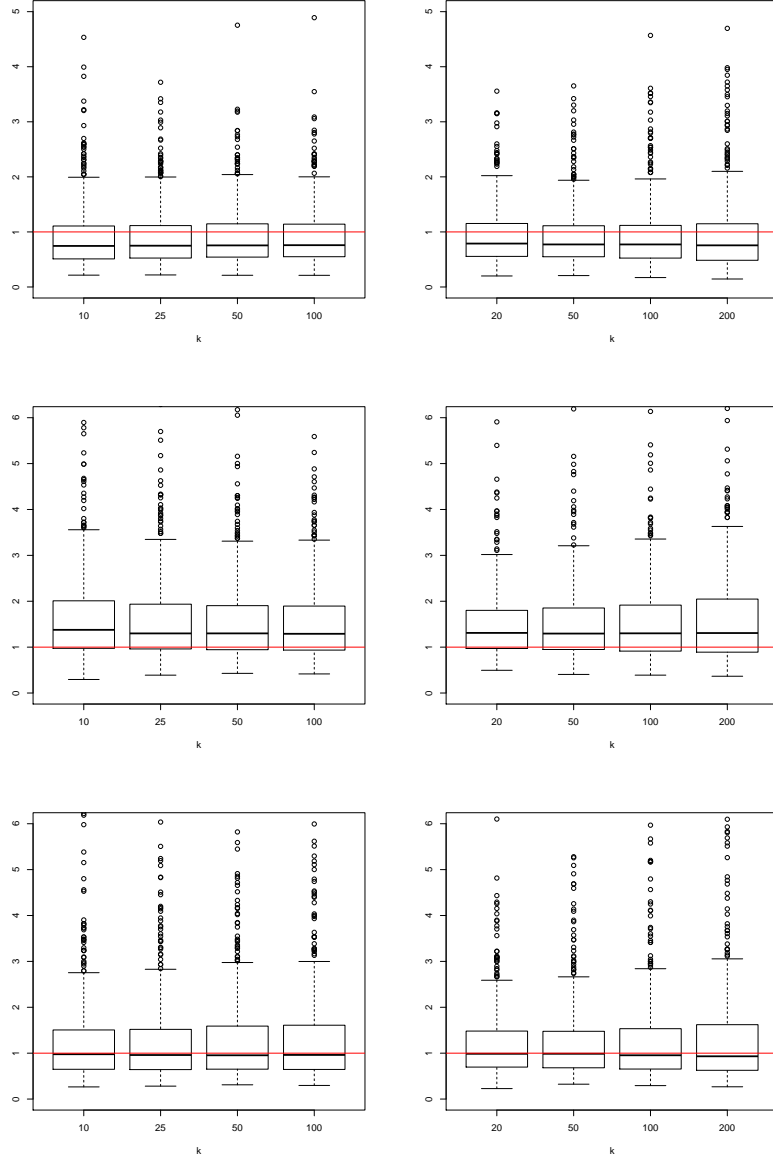


Figure 3: Boxplots of $\hat{\theta}_p(x_0)/\theta_p(x_0)$ for Model 2 based on 500 replications of sample sizes $n = 500$ (left column) and $n = 1000$ (right column) for $p = 1/(2n)$. From top to bottom we have $x_0 = 0.3$, $x_0 = 0.5$ and $x_0 = 0.8$. The values of k are taken as 2%, 5%, 10% and 20% of n .

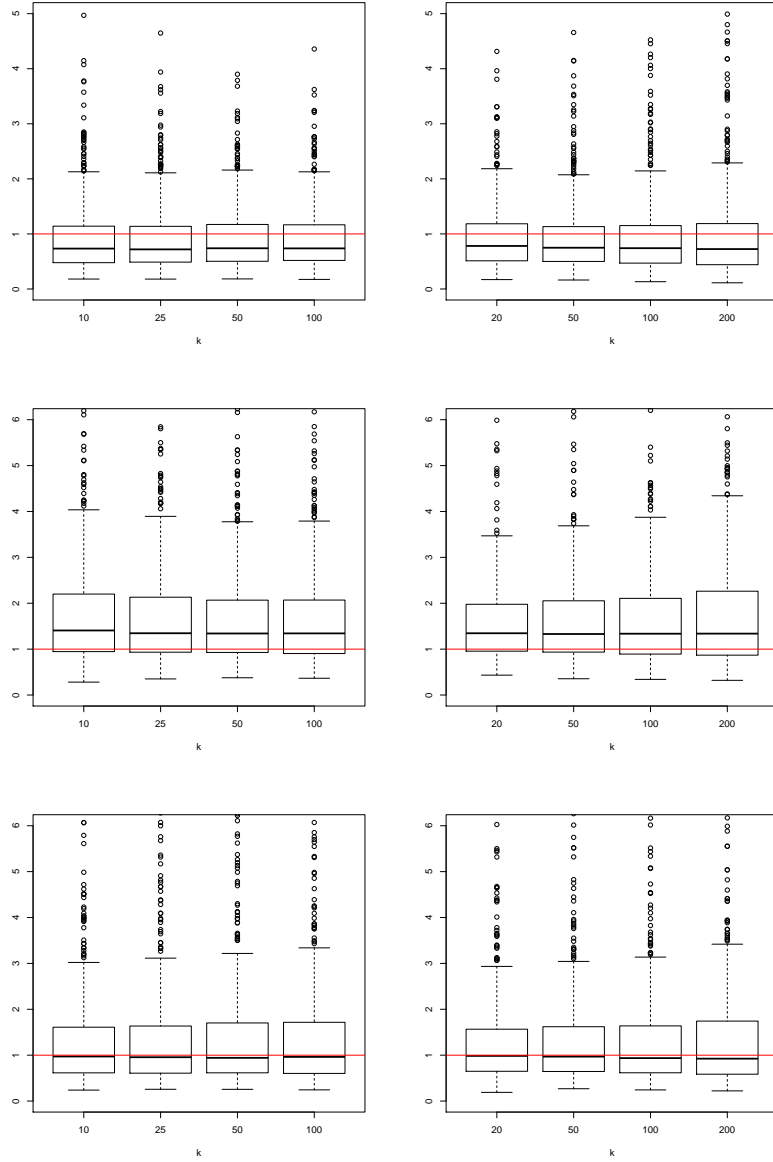


Figure 4: Boxplots of $\hat{\theta}_p(x_0)/\theta_p(x_0)$ for Model 2 based on 500 replications of sample sizes $n = 500$ (left column) and $n = 1000$ (right column) for $p = 1/(5n)$. From top to bottom we have $x_0 = 0.3$, $x_0 = 0.5$ and $x_0 = 0.8$. The values of k are taken as 2%, 5%, 10% and 20% of n .

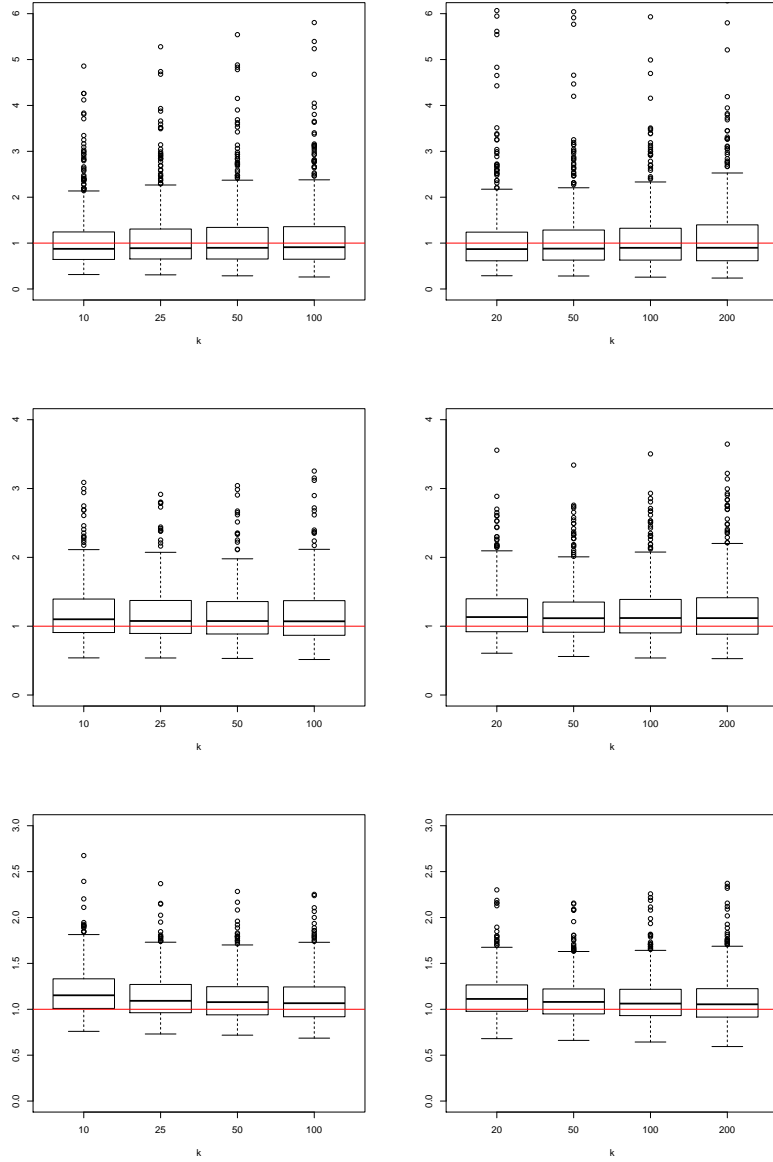


Figure 5: Boxplots of $\hat{\theta}_p(x_0)/\theta_p(x_0)$ for Model 3 based on 500 replications of sample sizes $n = 500$ (left column) and $n = 1000$ (right column) for $p = 1/(2n)$. From top to bottom we have $x_0 = 3$, $x_0 = 5$ and $x_0 = 7$. The values of k are taken as 2%, 5%, 10% and 20% of n .

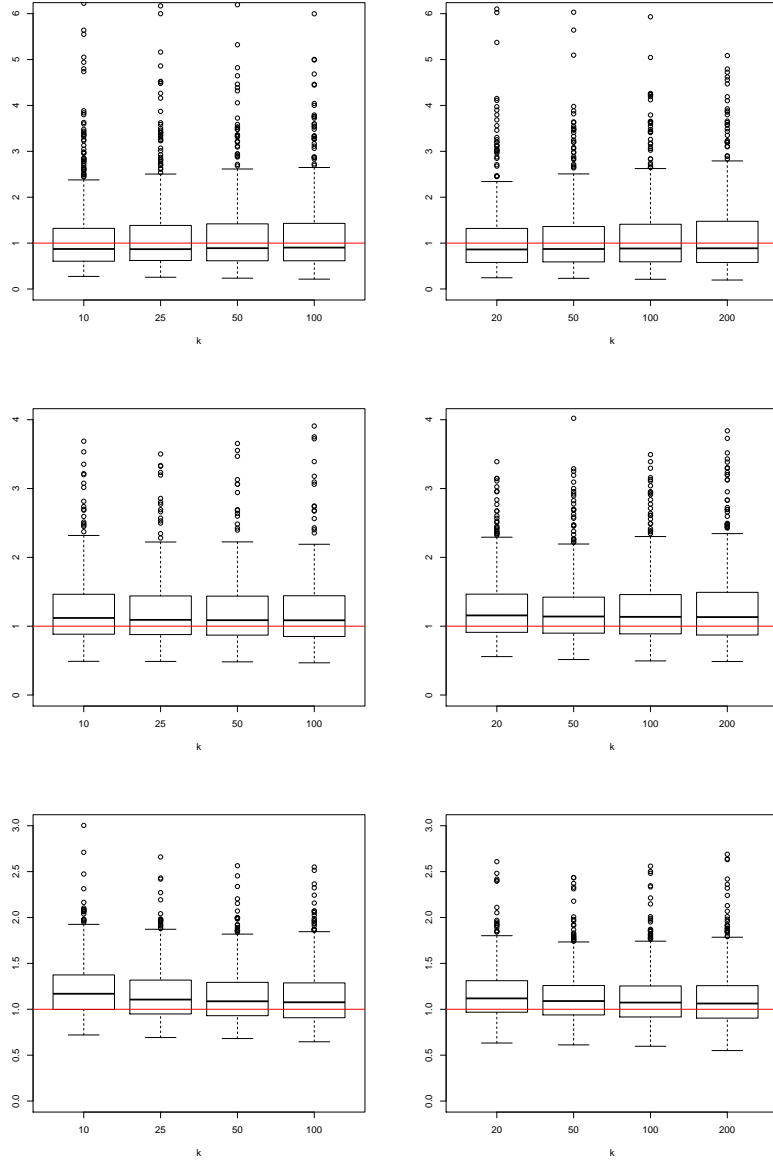


Figure 6: Boxplots of $\hat{\theta}_p(x_0)/\theta_p(x_0)$ for Model 3 based on 500 replications of sample sizes $n = 500$ (left column) and $n = 1000$ (right column) for $p = 1/(5n)$. From top to bottom we have $x_0 = 3$, $x_0 = 5$ and $x_0 = 7$. The values of k are taken as 2%, 5%, 10% and 20% of n .

the covariate age, X . We only consider the subjects for whom $Y^{(1)}$ and $Y^{(2)}$ are non-missing, giving $n = 729$. In Gardes and Girard (2015) these data were considered with the purpose of illustrating the estimation of the dependence function $R(y_1, y_2|x)$. The estimation results reported below are obtained with the bi-quadratic kernel function, and the bandwidth selection is done with the cross-validation criterion described in the simulation section. The scatterplot of blood pressure versus BMI, shown in Figure 7, indicates a positive association between these variables. In order to verify the Pareto-type behavior of $Y^{(1)}$ and $Y^{(2)}$, we construct the local Pareto quantile-quantile plots of the $Y^{(1)}$ and $Y^{(2)}$ data, respectively, for which the X coordinate is in a neighborhood of age=25 and age=55, see Figure 8. Clearly, the local Pareto quantile-quantile plots become approximately linear in the largest observations, which confirms underlying Pareto-type distributions (see Beirlant et al., 2004, for a general discussion of Pareto quantile-quantile plots). In Figure 9 we show the plot of the Hill estimates for blood pressure as a function of k , based on the data shown in the local Pareto quantile-quantile plots. When focusing on the stable horizontal parts of these plots we can clearly see that the theoretical requirement $\gamma_1(x_0) < 0.5$ is satisfied. Similar local Pareto quantile-quantile and Hill plots were obtained at other ages. Next, we investigate the asymptotic dependence assumption by plotting an estimate for $R(2, 2|x_0)$ as a function of k , at several values of x_0 . To this aim, we consider an adjustment of the estimator proposed in Escobar-Bach et al. (2018b), which is for the context of estimation of $L(y_1, y_2|x_0)$, to the context of estimation of $R(y_1, y_2|x_0)$, namely

$$\widehat{R}(y_1, y_2|x_0) = \frac{\frac{1}{k} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{\widehat{F}_{n,1}(Y_i^{(1)}|X_i) \leq \frac{k}{n} y_1, \widehat{F}_{n,2}(Y_i^{(2)}|X_i) \leq \frac{k}{n} y_2\}}}{\widehat{f}_n(x_0)}, \quad (6)$$

where $\widehat{F}_{n,1}$ is a kernel estimator for F_1 , of the same form as $\widehat{F}_{n,2}$ given in (1). For the bandwidth parameter h_n in (6) we take the bandwidth used in $\widehat{F}_{n,2}$, obtained with cross-validation, and which turns out to be the same as the cross-validated bandwidth for $\widehat{F}_{n,1}$ (by coincidence). In Figure 10 we plot $\widehat{R}(2, 2|x_0)$ as a function of k , at $x_0 = 30$ and $x_0 = 50$. Clearly, the displays show a positive estimate for $R(2, 2|x_0)$, which gives evidence of asymptotic dependence of $Y^{(1)}$ and $Y^{(2)}$ given $X = x_0$. Finally, we illustrate the estimation of the conditional marginal expected shortfall of blood pressure given a BMI that exceeds a high quantile and given a certain age. In Figure 11, we show $\widehat{\theta}_p(x_0)$ as a function of age for $p = 0.5\%$, 1% , 5% and 10% . The k -value for the estimation of the conditional marginal expected shortfall at a given x_0 was obtained by plotting $\widehat{\theta}_p(x_0)$ as a function of k , whereafter k is selected by a stability criterion as described in Goegebeur et al. (2019). As expected, smaller values of p lead to larger estimates of the conditional marginal expected shortfall. In Figure 12 we show $\widehat{\theta}_p(x_0)$ at several positions x_0 , and for $p = 1\%$ (left) and $p = 5\%$ (right), along with pointwise 95% confidence intervals for $\theta_p(x_0)$, obtained from Theorem 2.2. These confidence intervals are given by

$$\left[\frac{\widehat{\theta}_p(x_0)}{1 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sqrt{\widehat{\text{Var}}(\mathbb{W})}}{a_n}}, \frac{\widehat{\theta}_p(x_0)}{1 - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sqrt{\widehat{\text{Var}}(\mathbb{W})}}{a_n}} \right],$$

where $a_n := \min(\sqrt{k h_n^d}, \sqrt{k_1 h_n^d}/(\ln k/(np)))$, Φ^{-1} denotes the standard normal quantile function and $\widehat{\text{Var}}(\mathbb{W})$ is an estimate for the asymptotic variance given in (4), obtained by using the local

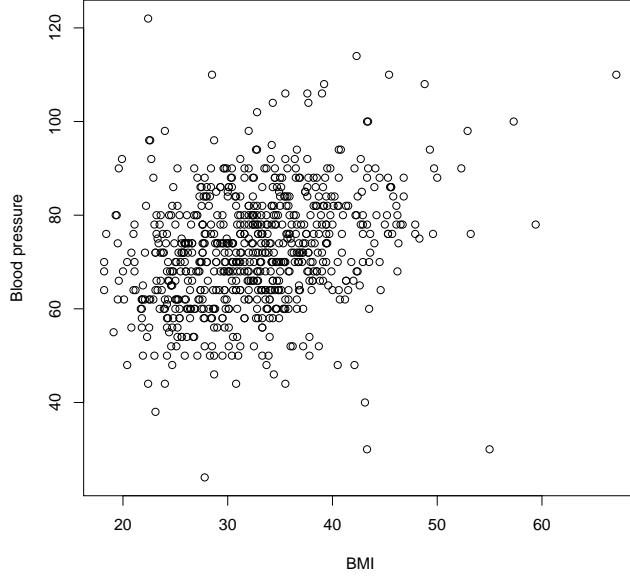


Figure 7: Indians diabetes dataset: scatterplot of blood pressure versus BMI.

Hill estimate (3) for $\gamma_1(x_0)$ and (6) as estimate for $R(y_1, y_2|x_0)$. The confidence intervals are clearly wider for $p = 1\%$ than for $p = 5\%$, reflecting the higher uncertainty of the estimate at $p = 1\%$ due to the fact that the estimation is based on fewer observations. Similarly, the confidence intervals are wider at covariate positions where the data are sparse. These findings are obviously in line with what can be expected.

5 Preliminary results

To be self contained, we recall below Theorem 2.3 from Goegebeur et al. (2020) which states the weak convergence of $\bar{\theta}_{\frac{k}{n}}(x_0)$.

Theorem 5.1 *Assume (\mathcal{D}) , (\mathcal{H}) , (\mathcal{K}) , (\mathcal{S}) with $x \rightarrow R(y_1, y_2|x)$ being a continuous function, and $y \rightarrow F_j(y|x_0)$, $j = 1, 2$, are strictly increasing. Let $x_0 \in \text{Int}(S_X)$ such that $f_X(x_0) > 0$. Consider sequences $k = \lfloor n^\alpha \ell_1(n) \rfloor$ and $h_n = n^{-\Delta} \ell_3(n)$, where ℓ_1 and ℓ_3 are slowly varying functions at infinity, with $\alpha \in (0, 1)$ and*

$$\begin{aligned} & \max \left(\frac{\alpha}{d + 2\gamma_1(x_0)(\eta_{A_1} \wedge \eta_{\gamma_1})}, \frac{\alpha}{d + 2(1 - \gamma_1(x_0))(\eta_{A_2} \wedge \eta_{\gamma_2} \wedge \eta_{B_2} \wedge \eta_{\varepsilon_2} \wedge \eta_{f_X})}, \right. \\ & \left. \frac{\alpha}{d} - \frac{2(1 - \alpha)\gamma_1^2(x_0)\beta_1(x_0)}{d + d(\beta_1(x_0) + \varepsilon)\gamma_1(x_0)}, \frac{\alpha - 2(1 - \alpha)(\gamma_1(x_0) \wedge (\beta_2(x_0)\gamma_2(x_0)) \wedge (-\tau))}{d} \right) \\ & < \Delta < \frac{\alpha}{d}. \end{aligned}$$

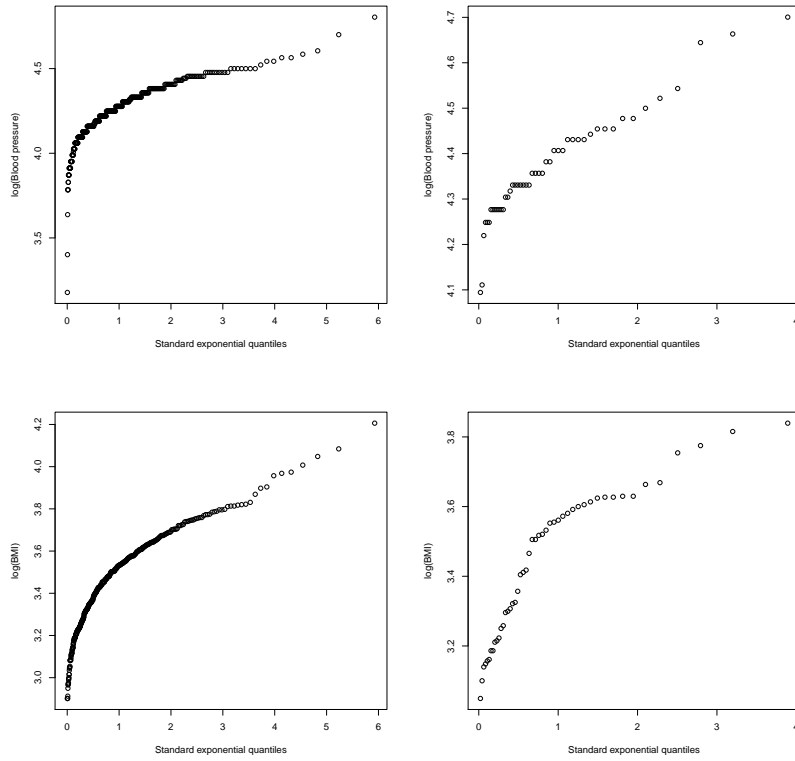


Figure 8: Indians diabetes dataset: local Pareto quantile-quantile plots of blood pressure (top row) and BMI (bottom row), at age=25 (left) and age=55 (right).

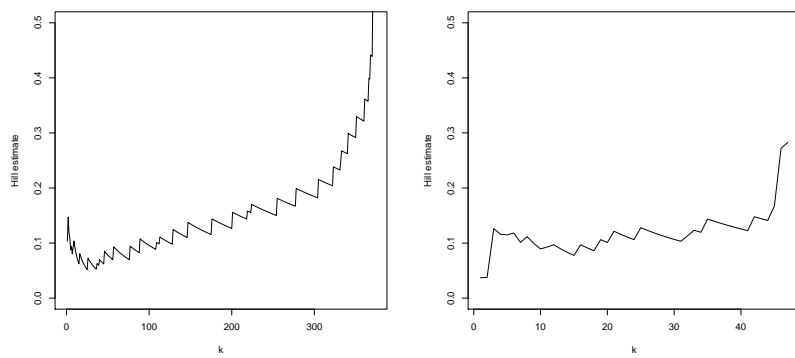


Figure 9: Indians diabetes dataset: local Hill plots of blood pressure for age=25 (left) and age=55 (right).

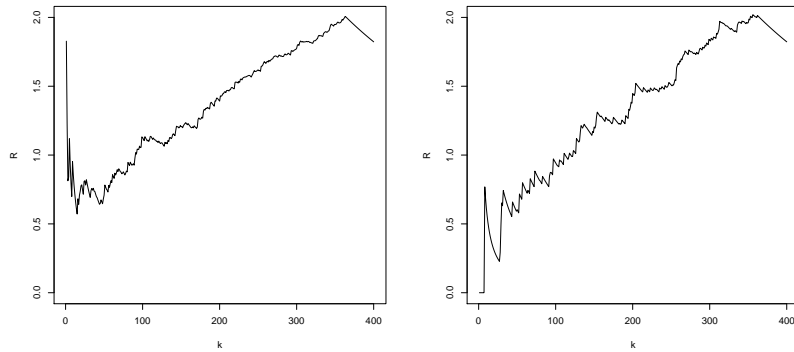


Figure 10: Indians diabetes dataset: $\hat{R}(2, 2|x_0)$ as a function of k , for $x_0 = 30$ (left) and $x_0 = 50$ (right).

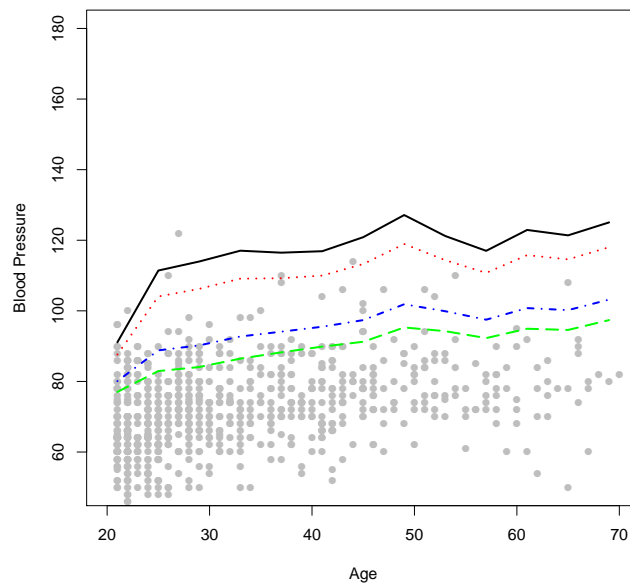


Figure 11: Indians diabetes dataset: $\hat{\theta}_p(x_0)$ as a function of x_0 for $p = 0.5\%$ (solid line) 1% (dotted line), 5% (dashed-dotted line) and 10% (dashed line).

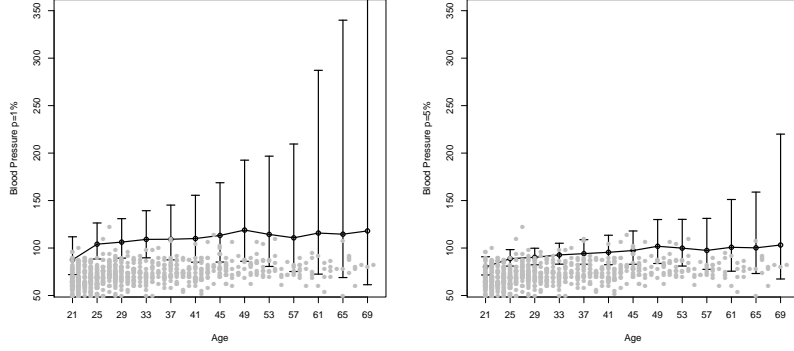


Figure 12: Indians diabetes dataset: pointwise 95% confidence interval for $\theta_p(x_0)$ at several values of x_0 for $p = 1\%$ (left) and $p = 5\%$ (right).

Then, for $\gamma_1(x_0) < 1/2$, we have

$$\sqrt{kh_n^d} \left(\frac{\bar{\theta}_{\frac{k}{n}}(x_0)}{\theta_{\frac{k}{n}}(x_0)} - 1 \right) \rightsquigarrow -(1 - \gamma_1(x_0)) \frac{W(\infty, 1)}{f_X(x_0)} + \frac{1}{f_X(x_0)} \frac{\int_0^\infty W(s, 1) ds^{-\gamma_1(x_0)}}{\int_0^\infty R(s, 1|x_0) ds^{-\gamma_1(x_0)}}.$$

Now, remark that, assuming that $F_1(y|x_0)$ is strictly increasing in y , we have

$$\begin{aligned} \hat{\gamma}_{1,k_1}(x_0) &= \frac{1}{\hat{f}_n(x_0)} \frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \int_{\hat{U}_1(n/k_1|x_0)}^{Y_i^{(1)}} \frac{1}{u} du \mathbb{1}_{\{Y_i^{(1)} \geq \hat{U}_1(n/k_1|x_0)\}} \\ &= \frac{1}{\hat{f}_n(x_0)} \int_{\hat{U}_1(n/k_1|x_0)}^\infty \frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \frac{1}{u} \mathbb{1}_{\{Y_i^{(1)} \geq u\}} du \\ &= \frac{1}{\hat{f}_n(x_0)} \int_{\hat{U}_1(n/k_1|x_0)}^\infty \frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{\bar{F}_1(Y_i^{(1)}|x_0) \leq \frac{k_1}{n} \frac{n}{k_1} \bar{F}_1(u|x_0)\}} \frac{1}{u} du \\ &= \frac{\gamma_1(x_0)}{\hat{f}_n(x_0)} \int_0^1 \frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{\bar{F}_1(Y_i^{(1)}|x_0) \leq \frac{k_1}{n} \frac{n}{k_1} \bar{F}_1(z^{-\gamma_1(x_0)} \hat{U}_1(n/k_1|x_0)|x_0)\}} \frac{1}{z} dz \\ &= \frac{\gamma_1(x_0)}{\hat{f}_n(x_0)} \int_0^1 T_n(\hat{s}_n(z|x_0)|x_0) \frac{1}{z} dz, \end{aligned} \tag{7}$$

where

$$\begin{aligned} T_n(y|x_0) &:= \frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{\bar{F}_1(Y_i^{(1)}|x_0) \leq k_1/n y\}}, \\ \text{and } \hat{s}_n(z|x_0) &:= \frac{n}{k_1} \bar{F}_1 \left(z^{-\gamma_1(x_0)} \hat{U}_1(n/k_1|x_0) \middle| x_0 \right). \end{aligned}$$

Thus we need to study the asymptotic properties of $T_n(y|x_0)$, $\hat{s}_n(z|x_0)$ and $\hat{U}_1(n/k_1|x_0)$.

We start by showing the weak convergence of the process based on $T_n(y|x_0)$, first when the process is centered around its expectation (Theorem 5.2) and then when it is centered around the dominant term of its expectation (Corollary 5.1).

Theorem 5.2 *Assume (\mathcal{D}) , (\mathcal{H}) , (\mathcal{K}) , and $x_0 \in \text{Int}(S_X)$ with $f_X(x_0) > 0$, and $y \rightarrow F_1(y|x_0)$ is strictly increasing. Consider sequences $k_1 \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, in such a way that $k_1/n \rightarrow 0$, $k_1 h_n^d \rightarrow \infty$ and $h_n^{\eta_{\gamma_1} \wedge \eta_{\varepsilon_1}} \ln n/k_1 \rightarrow 0$. Then for $\eta \in [0, 1/2)$, we have,*

$$\sqrt{k_1 h_n^d} \left(\frac{T_n(y|x_0) - \mathbb{E}(T_n(y|x_0))}{y^\eta} \right) \rightsquigarrow \frac{W(y, \infty)}{y^\eta} \quad (8)$$

in $D((0, T])$, for any $T > 0$.

Proposition 5.1 *Assume (\mathcal{D}) , (\mathcal{H}) , (\mathcal{K}) , and $x_0 \in \text{Int}(S_X)$ with $f_X(x_0) > 0$. Consider sequences $k_1 \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, in such a way that $k_1/n \rightarrow 0$, and $h_n^{\eta_{\gamma_1} \wedge \eta_{\varepsilon_1}} \ln n/k_1 \rightarrow 0$. Then for $\eta \in [0, 1)$, we have,*

$$\begin{aligned} \frac{\mathbb{E}(T_n(y|x_0))}{y^\eta} &= y^{1-\eta} f_X(x_0) + O\left(h_n^{\eta_{f_X} \wedge \eta_{A_1}}\right) + O\left(h_n^{\eta_{\gamma_1}} \ln \frac{n}{k_1}\right) + O\left(\left|\delta_1\left(U_1\left(\frac{n}{k_1}|x_0\right)|x_0\right)\right|h_n^{\eta_{B_1}}\right)\right) \\ &\quad + O\left(\left|\delta_1\left(U_1\left(\frac{n}{k_1}|x_0\right)|x_0\right)\right|h_n^{\eta_{\varepsilon_1}} \ln \frac{n}{k_1}\right), \end{aligned}$$

where the O -terms are uniform in $y \in (0, T]$, for any $T > 0$.

Corollary 5.1 *Assume (\mathcal{D}) , (\mathcal{H}) , (\mathcal{K}) , $x_0 \in \text{Int}(S_X)$ with $f_X(x_0) > 0$, and $y \rightarrow F_1(y|x_0)$ is strictly increasing. Consider sequences $k_1 \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, in such a way that $k_1/n \rightarrow 0$, $k_1 h_n^d \rightarrow \infty$, $h_n^{\eta_{\varepsilon_1}} \ln n/k_1 \rightarrow 0$, $\sqrt{k_1 h_n^d} h_n^{\eta_{f_X} \wedge \eta_{A_1}} \rightarrow 0$, $\sqrt{k_1 h_n^d} h_n^{\eta_{\gamma_1}} \ln n/k_1 \rightarrow 0$, $\sqrt{k_1 h_n^d} |\delta_1(U_1(n/k_1|x_0)|x_0)| h_n^{\eta_{B_1}} \rightarrow 0$, and $\sqrt{k_1 h_n^d} |\delta_1(U_1(n/k_1|x_0)|x_0)| h_n^{\eta_{\varepsilon_1}} \ln n/k_1 \rightarrow 0$. Then for $\eta \in [0, 1/2)$, we have,*

$$\sqrt{k_1 h_n^d} \left(\frac{T_n(y|x_0)}{y^\eta} - y^{1-\eta} f_X(x_0) \right) \rightsquigarrow \frac{W(y, \infty)}{y^\eta}$$

in $D((0, T])$, for any $T > 0$.

In the sequel, for convenient representation, the limiting process in Theorem 5.2 and Corollary 5.1 will be defined on the same probability space as the original random variables, via the Skorohod construction, but it should be kept in mind that it is only in distribution equal to the original process. The Skorohod representation theorem gives then, with keeping the same notation, that

$$\sup_{y \in (0, T]} \left| \sqrt{k_1 h_n^d} \left(\frac{T_n(y|x_0)}{y^\eta} - y^{1-\eta} f_X(x_0) \right) - \frac{W(y, \infty)}{y^\eta} \right| \rightarrow 0 \text{ a.s. ,}$$

as $n \rightarrow \infty$.

For the intermediate quantile estimate $\widehat{U}_1(n/k_1|x_0)$, we recall Lemma 5.6 from Goegebeur et al. (2020), which is used several times in our proofs, and which states the weak convergence of $\widehat{u}_n := \widehat{U}_1(n/k_1|x_0)/U_1(n/k_1|x_0)$.

Lemma 5.1 Assume (\mathcal{D}) , (\mathcal{H}) , (\mathcal{K}) , $x_0 \in \text{Int}(S_X)$ with $f_X(x_0) > 0$ and $y \rightarrow F_1(y|x_0)$ is strictly increasing. Consider sequences $k_1 \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$, in such a way that $k_1/n \rightarrow 0$, $k_1 h_n^d \rightarrow \infty$, $h_n^{\eta_{\varepsilon_1}} \ln n/k_1 \rightarrow 0$, $\sqrt{k_1 h_n^d} h_n^{\eta_{f_X} \wedge \eta_{A_1}} \rightarrow 0$, $\sqrt{k_1 h_n^d} h_n^{\eta_{\gamma_1}} \ln n/k_1 \rightarrow 0$, $\sqrt{k_1 h_n^d} |\delta_1(U_1(n/k_1|x_0)|x_0)| \rightarrow 0$. Then, as $n \rightarrow \infty$, we have

$$\sqrt{k_1 h_n^d} (\hat{u}_n - 1) \rightsquigarrow \frac{\gamma_1(x_0)W(1, \infty)}{f_X(x_0)}.$$

From Lemma 5.1, we can show now the uniform convergence in probability of $\hat{s}_n(z|x_0)$ towards z for any $z \in (0, T]$.

Lemma 5.2 Under the assumptions of Lemma 5.1, for any $T > 0$, we have

$$\sup_{z \in (0, T]} |\hat{s}_n(z|x_0) - z| = o_{\mathbb{P}}(1).$$

Proof of the preliminary results

Proof of Theorem 5.2. Recall that

$$\frac{T_n(y|x_0)}{y^\eta} = \frac{1}{k_1} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{\bar{F}_1(Y_i^{(1)}|x_0) \leq \frac{k_1}{n} y\}} \frac{1}{y^\eta}.$$

The proof of Theorem 5.2 follows the lines of proof of Theorem 2.1 in Goegebeur et al. (2020). Below, we only outline the main differences and refer to the latter paper otherwise. To start, we need some notations from empirical process theory with changing function classes, see for instance van der Vaart and Wellner (1996). Let P be the distribution measure of $(Y^{(1)}, X)$, and denote the expected value under P as $Pf := \int f dP$ for any real-valued measurable function $f : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$. For a function class \mathcal{F} , let $N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$, denote the minimal number of ε -brackets needed to cover \mathcal{F} . The bracketing integral is then defined as

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\ln N_{[]}(\varepsilon, \mathcal{F}, L_2(P))} d\varepsilon.$$

We introduce our sequence of classes \mathcal{F}_n on $\mathbb{R} \times \mathbb{R}^d$ as

$$\mathcal{F}_n := \{(u, z) \rightarrow f_{n,y}(u, z), y \in (0, T]\},$$

where

$$f_{n,y}(u, z) := \sqrt{\frac{n h_n^d}{k_1}} K_{h_n}(x_0 - z) \mathbb{1}_{\{\bar{F}_1(u|x_0) \leq \frac{k_1}{n} y\}} \frac{1}{y^\eta}.$$

Denote also by F_n an envelope function of the class \mathcal{F}_n . Now, according to Theorem 19.28 in van der Vaart (1998), the weak convergence of the stochastic process (8) follows from the

following four conditions. Let ρ_{x_0} be a semimetric, possibly depending on x_0 , making $(0, T]$ totally bounded. We have to prove that

$$\sup_{\rho_{x_0}(y, \bar{y}) \leq \delta_n} P(f_{n,y} - f_{n,\bar{y}})^2 \longrightarrow 0 \text{ for every } \delta_n \searrow 0, \quad (9)$$

$$PF_n^2 = O(1), \quad (10)$$

$$PF_n^2 \mathbb{1}_{\{F_n > \varepsilon \sqrt{n}\}} \longrightarrow 0 \text{ for every } \varepsilon > 0, \quad (11)$$

$$J_{\square}(\delta_n, \mathcal{F}_n, L_2(P)) \longrightarrow 0 \text{ for every } \delta_n \searrow 0. \quad (12)$$

We start with verifying condition (9), with $\rho_{x_0}(y, \bar{y}) := |y - \bar{y}|$. Without loss of generality, we may assume that $y \leq \bar{y}$. We have

$$P(f_{n,y} - f_{n,\bar{y}})^2 = \frac{nh_n^d}{k_1} \mathbb{E} \left[K_{h_n}^2(x_0 - X) \left(\frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}}}{y^\eta} - \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y}\}}}{\bar{y}^\eta} \right)^2 \right].$$

We consider now two cases.

Case 1: $y \leq \delta_n$. We have

$$\left(\frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}}}{y^\eta} - \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y}\}}}{\bar{y}^\eta} \right)^2 \leq 3 \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}}}{y^{2\eta}} + \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y}\}}}{\bar{y}^{2\eta}}.$$

This implies that

$$\begin{aligned} P(f_{n,y} - f_{n,\bar{y}})^2 &\leq 3 \frac{nh_n^d}{k_1} \mathbb{E} \left(K_{h_n}^2(x_0 - X) \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}}}{y^{2\eta}} \right) \\ &\quad + \frac{nh_n^d}{k_1} \mathbb{E} \left(K_{h_n}^2(x_0 - X) \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y}\}}}{\bar{y}^{2\eta}} \right) \\ &= 3 \frac{nh_n^d}{k_1} \int_{\mathbb{R}^d} \frac{1}{h_n^{2d}} K^2 \left(\frac{x_0 - v}{h_n} \right) \frac{P \left(\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y | X = v \right)}{y^{2\eta}} f_X(v) dv \\ &\quad + \frac{nh_n^d}{k_1} \int_{\mathbb{R}^d} \frac{1}{h_n^{2d}} K^2 \left(\frac{x_0 - v}{h_n} \right) \frac{P \left(\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y} | X = v \right)}{\bar{y}^{2\eta}} f_X(v) dv \\ &= 3 \frac{n}{k_1} \int_{S_K} K^2(v) \frac{P \left(\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y | X = x_0 - h_nv \right)}{y^{2\eta}} f_X(x_0 - h_nv) dv \\ &\quad + \frac{n}{k_1} \int_{S_K} K^2(v) \frac{P \left(\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y} | X = x_0 - h_nv \right)}{\bar{y}^{2\eta}} f_X(x_0 - h_nv) dv. \end{aligned}$$

Since $P\left(\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y \middle| X = x_0 - h_nv\right) = \bar{F}_1\left(U_1\left(\frac{n}{k_1y}|x_0\right) \middle| x_0 - h_nv\right)$, this yields

$$\begin{aligned} P(f_{n,y} - f_{n,\bar{y}})^2 &\leq 3y^{1-2\eta} \int_{S_K} K^2(v) f_X(x_0 - h_nv) dv \\ &\quad + 3 \int_{S_K} K^2(v) \left[\frac{1}{y^{2\eta}} \frac{n}{k_1} \bar{F}_1\left(U_1\left(\frac{n}{k_1y}|x_0\right) \middle| x_0 - h_nv\right) - y^{1-2\eta} \right] f_X(x_0 - h_nv) dv \\ &\quad + \bar{y}^{1-2\eta} \int_{S_K} K^2(v) f_X(x_0 - h_nv) dv \\ &\quad + \int_{S_K} K^2(v) \left[\frac{1}{\bar{y}^{2\eta}} \frac{n}{k_1} \bar{F}_1\left(U_1\left(\frac{n}{k_1\bar{y}}|x_0\right) \middle| x_0 - h_nv\right) - \bar{y}^{1-2\eta} \right] f_X(x_0 - h_nv) dv. \end{aligned}$$

Using Lemma 5.1 in Goegebeur et al. (2020) and the fact that $\rho_{x_0}(y, \bar{y}) \leq \delta_n$ which implies $\bar{y} \leq 2\delta_n$, we get

$$P(f_{n,y} - f_{n,\bar{y}})^2 \leq 5\delta_n^{1-2\eta} \int_{S_K} K^2(v) f_X(x_0 - h_nv) dv + o(1),$$

where the $o(1)$ -term does not depend on y and \bar{y} .

Case 2: $y > \delta_n$. In that case

$$\begin{aligned} \left(\frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}}}{y^\eta} - \frac{\mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y}\}}}{\bar{y}^\eta} \right)^2 &\leq \left(\frac{1}{y^\eta} - \frac{1}{\bar{y}^\eta} \right)^2 \mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}} \\ &\quad + \frac{1}{y^{2\eta}} \left\{ \mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y}\}} - \mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}y\}} \right\}, \end{aligned}$$

from which we deduce that

$$\begin{aligned} P(f_{n,y} - f_{n,\bar{y}})^2 &\leq \frac{(y^\eta - \bar{y}^\eta)^2}{(y\bar{y})^{2\eta}} \int_{S_K} K^2(v) \frac{n}{k_1} \bar{F}_1\left(U_1(n/(k_1y)|x_0) \middle| x_0 - h_nv\right) f_X(x_0 - h_nv) dv \\ &\quad + \frac{1}{y^{2\eta}} \frac{n}{k_1} \int_{S_K} K^2(v) P\left(\frac{k_1}{n}y \leq \bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n}\bar{y} \middle| X = x_0 - h_nv\right) f_X(x_0 - h_nv) dv. \end{aligned}$$

These two terms on the right-hand side of the above inequality can be handled similarly as those in case 3 in the proof of Theorem 2.1 in Goegebeur et al. (2020).

Now, a natural envelope function of the class \mathcal{F}_n is

$$F_n(u, z) := \sqrt{\frac{nh_n^d}{k_1}} K_{h_n}(x_0 - z) \frac{\mathbb{1}_{\{\bar{F}_1(u|x_0) \leq k_1T/n\}}}{[(n/k_1)\bar{F}_1(u|x_0)]^\eta}.$$

Thus, according again to the proof of Theorem 2.1 in Goegebeur et al. (2020), conditions (10) and (11) are satisfied.

Finally, we need to show condition (12). Without loss of generality we assume $T = 1$. Consider for $a, \theta < 1$:

$$\begin{aligned}\mathcal{F}_n^{(1)}(a) &:= \{f_{n,y} \in \mathcal{F}_n : y \leq a\}, \\ \mathcal{F}_n^{(2)}(\ell) &:= \{f_{n,y} \in \mathcal{F}_n : \theta^{\ell+1} \leq y \leq \theta^\ell\},\end{aligned}$$

where $\ell = 0, \dots, \lfloor \ln a / \ln \theta \rfloor$.

The class $\mathcal{F}_n^{(1)}(a)$ has been already studied in the proof of Theorem 2.1 in Goegebeur et al. (2020) and $\mathcal{F}_n^{(2)}(\ell)$ can be dealt with similar arguments as for $\mathcal{F}_n(\ell, m)$ from the latter paper, since we have the following bounds

$$\begin{aligned}\underline{u}_n(u, z) &:= \sqrt{\frac{nh_n^d}{k_1}} K_{h_n}(x_0 - z) \frac{\mathbb{1}_{\{\bar{F}_1(u|x_0) \leq k_1/n \theta^{\ell+1}\}}}{\theta^{\ell\eta}} \\ &\leq f_{n,y}(u, z) \leq \sqrt{\frac{nh_n^d}{k_1}} K_{h_n}(x_0 - z) \frac{\mathbb{1}_{\{\bar{F}_1(u|x_0) \leq k_1/n \theta^\ell\}}}{\theta^{(\ell+1)\eta}} =: \bar{u}_n(u, z).\end{aligned}$$

This concludes the proof of Theorem 5.2.

Proof of Proposition 5.1. We have

$$\begin{aligned}\frac{\mathbb{E}[T_n(y|x_0)]}{y^\eta} &= \frac{1}{y^\eta} \frac{n}{k_1} \mathbb{E} \left[K_{h_n}(x_0 - X) \mathbb{1}_{\{\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n} y\}} \right] \\ &= \frac{1}{y^\eta} \frac{n}{k_1} \int_{\mathbb{R}^d} \frac{1}{h_n^d} K \left(\frac{x_0 - u}{h_n} \right) P \left(\bar{F}_1(Y^{(1)}|x_0) \leq \frac{k_1}{n} y \mid X = u \right) f_X(u) du \\ &= \frac{1}{y^\eta} \frac{n}{k_1} \int_{S_K} K(u) \bar{F}_1 \left(U_1 \left(\frac{n}{k_1 y} \mid x_0 \right) \mid x_0 - h_n u \right) f_X(x_0 - h_n u) du \\ &= y^{1-\eta} f_X(x_0) + y^{1-\eta} \int_{S_K} K(u) [f_X(x_0 - h_n u) - f_X(x_0)] du \\ &\quad + f_X(x_0) \int_{S_K} K(u) \left[\frac{\frac{n}{k_1} \bar{F}_1 \left(U_1 \left(\frac{n}{k_1 y} \mid x_0 \right) \mid x_0 - h_n u \right)}{y^\eta} - y^{1-\eta} \right] du \\ &\quad + \int_{S_K} K(u) \left[\frac{\frac{n}{k_1} \bar{F}_1 \left(U_1 \left(\frac{n}{k_1 y} \mid x_0 \right) \mid x_0 - h_n u \right)}{y^\eta} - y^{1-\eta} \right] [f_X(x_0 - h_n u) - f_X(x_0)] du.\end{aligned}$$

Following the lines of proof of Lemma 5.1 in Goegebeur et al. (2020), we have

$$\begin{aligned}\left| \frac{\frac{n}{k_1} \bar{F}_1 \left(U_1 \left(\frac{n}{k_1 y} \mid x_0 \right) \mid x_0 - h_n u \right)}{y^\eta} - y^{1-\eta} \right| &= O \left(h_n^{\eta A_1} \right) + O \left(h_n^{\eta \gamma_1} \ln \frac{n}{k_1} \right) + O \left(\left| \delta_1 \left(U_1 \left(\frac{n}{k_1} \mid x_0 \right) \mid x_0 \right) \right| h_n^{\eta B_1} \right) \\ &\quad + O \left(\left| \delta_1 \left(U_1 \left(\frac{n}{k_1} \mid x_0 \right) \mid x_0 \right) \right| h_n^{\eta \varepsilon_1} \ln \frac{n}{k_1} \right)\end{aligned}$$

with O -terms which are uniform in $y \in (0, T]$, for any $T > 0$. This yields Proposition 5.1.

Proof of Corollary 5.1. Using the decomposition

$$\begin{aligned} \sqrt{k_1 h_n^d} \left(\frac{T_n(y|x_0)}{y^\eta} - y^{1-\eta} f_X(x_0) \right) &= \sqrt{k_1 h_n^d} \left(\frac{T_n(y|x_0) - \mathbb{E}(T_n(y|x_0))}{y^\eta} \right) \\ &\quad + \sqrt{k_1 h_n^d} \left(\frac{\mathbb{E}(T_n(y|x_0))}{y^\eta} - y^{1-\eta} f_X(x_0) \right), \end{aligned}$$

combined with Theorem 5.2 and Proposition 5.1 yields Corollary 5.1.

Proof of Lemma 5.2. We have, for $z \in (0, T]$ and any $\varepsilon > 0$ and $\zeta \in (0, \beta_1(x_0)]$

$$\begin{aligned} |\widehat{s}_n(z|x_0) - z| &= \left| \frac{n}{k_1} \overline{F}_1 \left(z^{-\gamma_1(x_0)} \widehat{u}_n U_1(n/k_1|x_0) \middle| x_0 \right) - z \right| \\ &= \left| \frac{\overline{F}_1 \left(z^{-\gamma_1(x_0)} \widehat{u}_n U_1(n/k_1|x_0) \middle| x_0 \right)}{\overline{F}_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right)} - z \right| \\ &\leq z \left| \widehat{u}_n^{-1/\gamma_1(x_0)} - 1 \right| \frac{1 + \frac{1}{\gamma_1(x_0)} \delta_1 \left(z^{-\gamma_1(x_0)} \widehat{u}_n U_1(n/k_1|x_0) \middle| x_0 \right)}{1 + \frac{1}{\gamma_1(x_0)} \delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right)} \\ &\quad + \frac{z}{\gamma_1(x_0)} \frac{\left| \delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right) \right|}{1 + \frac{1}{\gamma_1(x_0)} \delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right)} \left| \frac{\delta_1 \left(z^{-\gamma_1(x_0)} \widehat{u}_n U_1(n/k_1|x_0) \middle| x_0 \right)}{\delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right)} - 1 \right| \\ &\leq CT \left| \widehat{u}_n^{-1/\gamma_1(x_0)} - 1 \right| \\ &\quad + CT \left| \delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right) \right| \left\{ \left| \frac{\delta_1 \left(z^{-\gamma_1(x_0)} \widehat{u}_n U_1(n/k_1|x_0) \middle| x_0 \right)}{\delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right)} - \left(z^{-\gamma_1(x_0)} \widehat{u}_n \right)^{-\beta_1(x_0)} \right| \right. \\ &\quad \left. + \left| \left(z^{-\gamma_1(x_0)} \widehat{u}_n \right)^{-\beta_1(x_0)} - 1 \right| \right\} \\ &\leq CT \left| \widehat{u}_n^{-1/\gamma_1(x_0)} - 1 \right| \\ &\quad + C\varepsilon \left| \delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right) \right| T^{1+\gamma_1(x_0)\beta_1(x_0)\pm\gamma_1(x_0)\zeta} \widehat{u}_n^{-\beta_1(x_0)\pm\zeta} \\ &\quad + C \left| \delta_1 \left(U_1(n/k_1|x_0) \middle| x_0 \right) \right| T \left\{ T^{\gamma_1(x_0)\beta_1(x_0)} \widehat{u}_n^{-\beta_1(x_0)} + 1 \right\}, \end{aligned}$$

for n large, with arbitrary large probability, by Proposition B.1.10 in de Haan and Ferreira (2006). In the above, the notation $a^{\pm\bullet}$ means a^\bullet if $a \geq 1$ and $a^{-\bullet}$ if $a < 1$. Using Lemma 5.1, Lemma 5.2 follows.

6 Proofs of the main results

Proof of Theorem 2.1. Using (7), we have the following decomposition

$$\begin{aligned}
\sqrt{k_1 h_n^d} (\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)) &= \frac{\gamma_1(x_0)}{f_X(x_0)} \int_0^1 W(z, \infty) \frac{1}{z} dz \\
&+ \gamma_1(x_0) \sqrt{k_1 h_n^d} \int_0^1 \left[\frac{\hat{s}_n(z|x_0)}{z} - 1 \right] dz \\
&+ \frac{\gamma_1(x_0)}{f_X(x_0)} \int_0^1 [W(\hat{s}_n(z|x_0), \infty) - W(z, \infty)] \frac{1}{z} dz \\
&+ \gamma_1(x_0) \int_0^1 \left\{ \sqrt{k_1 h_n^d} \left[\frac{T_n(\hat{s}_n(z|x_0)|x_0)}{f_X(x_0) \hat{s}_n^\eta(z|x_0)} - \hat{s}_n^{1-\eta}(z|x_0) \right] - \frac{W(\hat{s}_n(z|x_0), \infty)}{f_X(x_0) \hat{s}_n^\eta(z|x_0)} \right\} \frac{\hat{s}_n^\eta(z|x_0)}{z} dz \\
&- \frac{\gamma_1(x_0)}{f_X(x_0) \hat{f}_n(x_0)} \int_0^1 T_n(\hat{s}_n(z|x_0)|x_0) \frac{1}{z} dz \sqrt{\frac{k_1}{n}} \sqrt{n h_n^d} (\hat{f}_n(x_0) - f_X(x_0)) \\
&=: \frac{\gamma_1(x_0)}{f_X(x_0)} \int_0^1 W(z, \infty) \frac{1}{z} dz + \sum_{i=1}^4 T_{i,n}.
\end{aligned}$$

We study each term separately.

Concerning $T_{1,n}$, following the lines of proof of Lemma 5.2, we have

$$T_{1,n} = \gamma_1(x_0) \sqrt{k_1 h_n^d} \left[\hat{u}_n^{-1/\gamma_1(x_0)} - 1 \right] + o_{\mathbb{P}}(1).$$

Now, combining Lemma 5.1 with a Taylor expansion, we have

$$T_{1,n} \rightsquigarrow -\frac{\gamma_1(x_0)}{f_X(x_0)} W(1, \infty).$$

Concerning $T_{2,n}$, for $\delta \in (0, 1)$, we use the decomposition

$$\begin{aligned}
T_{2,n} &= \frac{\gamma_1(x_0)}{f_X(x_0)} \left\{ \int_0^\delta [W(\hat{s}_n(z|x_0), \infty) - W(z, \infty)] \frac{1}{z} dz + \int_\delta^1 [W(\hat{s}_n(z|x_0), \infty) - W(z, \infty)] \frac{1}{z} dz \right\} \\
&=: T_{2,n}^{(1)} + T_{2,n}^{(2)}.
\end{aligned}$$

Using Lemma 5.2 combined with Potter's bounds (see Proposition B.1.9 (5) in de Haan and Ferreira, 2006), we have for any $\zeta \in (0, 1/\gamma_1(x_0))$, for n large, and with arbitrary large probability

$$\begin{aligned}
|T_{2,n}^{(1)}| &\leq \frac{\gamma_1(x_0)}{f_X(x_0)} \sup_{z \in (0, 2]} \frac{|W(z, \infty)|}{z^\eta} \left[\int_0^\delta \frac{\hat{s}_n^\eta(z|x_0)}{z} dz + \frac{\delta^\eta}{\eta} \right] \\
&\leq C \int_0^\delta \frac{1}{z} \left[z^{-\gamma_1(x_0)} \hat{u}_n \right]^{\eta(\pm\zeta - 1/\gamma_1(x_0))} dz + C \delta^\eta \\
&\leq C \hat{u}_n^{(\pm\zeta - 1/\gamma_1(x_0))\eta} \delta^{(1 - \gamma_1(x_0)\zeta)\eta} + C \delta^\eta \\
&= C \varepsilon^{1 - \gamma_1(x_0)\zeta} + C \varepsilon, \text{ by choosing } \delta^\eta = \varepsilon.
\end{aligned}$$

Now, concerning $T_{2,n}^{(2)}$, remark that following the lines of proof of Lemma 5.2, we have

$$\sup_{z \in [\delta, 1]} \left(k_1 h_n^d \right)^{1/4} |\hat{s}_n(z|x_0) - z| = o_{\mathbb{P}}(1),$$

from which we deduce that, for any $\xi > 0$

$$\begin{aligned} & \mathbb{P} \left(\sup_{z \in [\delta, 1]} |W(\hat{s}_n(z|x_0), \infty) - W(z, \infty)| > \xi \right) \\ & \leq \mathbb{P} \left(\sup_{z \in [\delta, 1]} |W(\hat{s}_n(z|x_0), \infty) - W(z, \infty)| > \xi, \sup_{z \in [\delta, 1]} |\hat{s}_n(z|x_0) - z| \leq \frac{1}{(k_1 h_n^d)^{1/4}} \right) \\ & \quad + \mathbb{P} \left(\sup_{z \in [\delta, 1]} |\hat{s}_n(z|x_0) - z| > \frac{1}{(k_1 h_n^d)^{1/4}} \right) \\ & \leq \mathbb{P} \left(\sup_{z \in [\delta, 1], |y-z| \leq \frac{1}{(k_1 h_n^d)^{1/4}}} |W(y, \infty) - W(z, \infty)| > \xi \right) + \mathbb{P} \left(\sup_{z \in [\delta, 1]} |\hat{s}_n(z|x_0) - z| > \frac{1}{(k_1 h_n^d)^{1/4}} \right) \\ & = o(1), \end{aligned}$$

by the continuity of $W(\cdot, \infty)$. This implies that

$$|T_{2,n}^{(2)}| \leq \varepsilon \ln \frac{1}{\delta} = -\frac{\varepsilon}{\eta} \ln \varepsilon.$$

Hence, $T_{2,n} = o_{\mathbb{P}}(1)$.

Concerning $T_{3,n}$, from Lemma 5.2, we have, for n large, with arbitrary large probability

$$|T_{3,n}| \leq \frac{\gamma_1(x_0)}{f_X(x_0)} \sup_{y \in (0, 2]} \left| \sqrt{k_1 h_n^d} \left(\frac{T_n(y|x_0)}{y^\eta} - y^{1-\eta} f_X(x_0) \right) - \frac{W(y, \infty)}{y^\eta} \right| \int_0^1 \frac{\hat{s}_n^\eta(z|x_0)}{z} dz.$$

Then, by Corollary 5.1 combined with the Skorohod representation theorem, we can conclude that $T_{3,n} = o_{\mathbb{P}}(1)$.

Finally, $T_{4,n} = o_{\mathbb{P}}(1)$ using the properties of the kernel density estimator.

This achieves the proof of Theorem 2.1.

Proof of Theorem 2.2. We use the decomposition

$$\frac{\hat{\theta}_p(x_0)}{\theta_p(x_0)} = \underbrace{\left\{ \left(\frac{k}{np} \right)^{\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)} \right\}}_{T_{5,n}} \underbrace{\left\{ \frac{\bar{\theta}_{\frac{k}{n}}(x_0)}{\theta_{\frac{k}{n}}(x_0)} \right\}}_{T_{6,n}} \underbrace{\left\{ \frac{\left(\frac{k}{np} \right)^{\gamma_1(x_0)} \theta_{\frac{k}{n}}(x_0)}{\theta_p(x_0)} \right\}}_{T_{7,n}},$$

from which we deduce that

$$\frac{\hat{\theta}_p(x_0)}{\theta_p(x_0)} - 1 = \left\{ \left(\frac{k}{np} \right)^{\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)} - 1 \right\} T_{6,n} T_{7,n} + \left\{ \frac{\bar{\theta}_{\frac{k}{n}}(x_0)}{\theta_{\frac{k}{n}}(x_0)} - 1 \right\} T_{7,n} + \left\{ \frac{(\frac{k}{np})^{\gamma_1(x_0)} \theta_{\frac{k}{n}}(x_0)}{\theta_p(x_0)} - 1 \right\}. \quad (13)$$

We will study the three terms $(T_{i,n} - 1)$, $i = 5, 6, 7$, separately.

Concerning the term $(T_{5,n} - 1)$, remark that, assuming $\frac{\ln k/(np)}{\sqrt{k_1 h_n^d}} \rightarrow 0$ and using Theorem 2.1, we have

$$\begin{aligned} \left(\frac{k}{np} \right)^{\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)} - 1 &= \exp \left\{ \sqrt{k_1 h_n^d} [\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)] \frac{\ln k/(np)}{\sqrt{k_1 h_n^d}} \right\} - 1 \\ &= \sqrt{k_1 h_n^d} [\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)] \frac{\ln k/(np)}{\sqrt{k_1 h_n^d}} (1 + o_{\mathbb{P}}(1)), \end{aligned}$$

from which we deduce that

$$\frac{\sqrt{k_1 h_n^d}}{\ln k/(np)} \left\{ \left(\frac{k}{np} \right)^{\hat{\gamma}_{1,k_1}(x_0) - \gamma_1(x_0)} - 1 \right\} \rightsquigarrow \frac{\gamma_1(x_0)}{f_X(x_0)} \left[\int_0^1 W(z, \infty) \frac{1}{z} dz - W(1, \infty) \right]. \quad (14)$$

The asymptotic behavior of the term $(T_{6,n} - 1)$ has been already established in Theorem 5.1.

Now, concerning the term $(T_{7,n} - 1)$, remark that

$$\begin{aligned} \frac{(\frac{k}{np})^{\gamma_1(x_0)} \theta_{k/n}(x_0)}{\theta_p(x_0)} - 1 &= \left(\frac{\theta_{k/n}(x_0)/U_1(n/k|x_0)}{\theta_p(x_0)/U_1(1/p|x_0)} - 1 \right) \frac{U_1(n/k|x_0)(\frac{k}{np})^{\gamma_1(x_0)}}{U_1(1/p|x_0)} \\ &\quad + \frac{U_1(n/k|x_0)(\frac{k}{np})^{\gamma_1(x_0)}}{U_1(1/p|x_0)} - 1. \end{aligned} \quad (15)$$

Under assumption (\mathcal{D}) , (2) yields

$$\frac{U_1(n/k|x_0)(\frac{k}{np})^{\gamma_1(x_0)}}{U_1(1/p|x_0)} - 1 = \frac{1 + a_1(n/k|x_0)}{1 + a_1(1/p|x_0)} - 1 = o\left(\frac{1}{\sqrt{k h_n^d}}\right), \quad (16)$$

since $\sqrt{k h_n^d} |\delta_1(U_1(n/k|x_0)|x_0)| \rightarrow 0$ and $\sqrt{k h_n^d} |\delta_1(U_1(1/p|x_0)|x_0)| \rightarrow 0$ under our assumptions.

Moreover

$$\begin{aligned}
\frac{\theta_{k/n}(x_0)}{U_1(n/k|x_0)} &= \int_0^\infty \frac{n}{k} \mathbb{P}\left(Y^{(1)} > y_1, Y^{(2)} \geq U_2(n/k|x_0) \mid x_0\right) \frac{dy_1}{U_1(n/k|x_0)} \\
&= \int_0^\infty \frac{n}{k} \mathbb{P}\left(1 - F_1(Y^{(1)}|x_0) < 1 - F_1(y_1|x_0), 1 - F_2(Y^{(2)}|x_0) \leq \frac{k}{n} \mid x_0\right) \frac{dy_1}{U_1(n/k|x_0)} \\
&= \int_0^\infty R_{n/k}\left(\frac{n}{k} [1 - F_1(y_1|x_0)], 1 \mid x_0\right) \frac{dy_1}{U_1(n/k|x_0)} \\
&= - \int_0^\infty R_{n/k}\left(\frac{n}{k} [1 - F_1(z_1^{-\gamma_1(x_0)} U_1(n/k|x_0)|x_0)], 1 \mid x_0\right) dz_1^{-\gamma_1(x_0)} \\
&= - \int_0^\infty R(z_1, 1|x_0) dz_1^{-\gamma_1(x_0)} \\
&\quad - \int_0^\infty \left[R_{n/k}\left(\frac{n}{k} [1 - F_1(z_1^{-\gamma_1(x_0)} U_1(n/k|x_0)|x_0)], 1 \mid x_0\right) \right. \\
&\quad \quad \left. - R\left(\frac{n}{k} [1 - F_1(z_1^{-\gamma_1(x_0)} U_1(n/k|x_0)|x_0)], 1 \mid x_0\right) \right] dz_1^{-\gamma_1(x_0)} \\
&\quad - \int_0^\infty \left[R\left(\frac{n}{k} [1 - F_1(z_1^{-\gamma_1(x_0)} U_1(n/k|x_0)|x_0)], 1 \mid x_0\right) - R(z_1, 1|x_0) \right] dz_1^{-\gamma_1(x_0)} \\
&=: - \int_0^\infty R(z_1, 1|x_0) dz_1^{-\gamma_1(x_0)} + T_{8,n} + T_{9,n}.
\end{aligned}$$

Now, by Assumption (\mathcal{S})

$$\begin{aligned}
|T_{8,n}| &\leq \sup_{x \in B(x_0, h_n)} \sup_{0 < y_1 < \infty, \frac{1}{2} \leq y_2 \leq 2} \frac{|R_{n/k}(y_1, y_2|x) - R(y_1, y_2|x_0)|}{y_1^\beta \wedge 1} \\
&\quad \times \left| \int_0^\infty \left(\left\{ \frac{n}{k} [1 - F_1(z_1^{-\gamma_1(x_0)} U_1(n/k|x_0)|x_0)] \right\}^\beta \wedge 1 \right) dz_1^{-\gamma_1(x_0)} \right| \\
&= O\left(\left(\frac{n}{k}\right)^\tau\right) = o\left(\frac{1}{\sqrt{kh_n^d}}\right),
\end{aligned}$$

by our assumptions on the sequence k , and

$$\begin{aligned}
|T_{9,n}| &\leq \sup_{\frac{1}{2} \leq y_2 \leq 2} \left| \int_0^\infty \left[R\left(\frac{n}{k} [1 - F_1(z_1^{-\gamma_1(x_0)} U_1(n/k|x_0)|x_0)], y_2 \mid x_0\right) - R(z_1, y_2|x_0) \right] dz_1^{-\gamma_1(x_0)} \right| \\
&= o\left(\frac{1}{\sqrt{kh_n^d}}\right),
\end{aligned}$$

by Lemma 5.4 in Goegebeur et al. (2020). Thus

$$\frac{\theta_{k/n}(x_0)}{U_1(n/k|x_0)} = - \int_0^\infty R(z_1, 1|x_0) dz_1^{-\gamma_1(x_0)} + o\left(\frac{1}{\sqrt{kh_n^d}}\right). \quad (17)$$

A similar type of property can be obtained for $\theta_p(x_0)$ instead of $\theta_{k/n}(x_0)$. Indeed

$$\begin{aligned}
\frac{\theta_p(x_0)}{U_1(1/p|x_0)} &= - \int_0^\infty R_{1/p} \left(\frac{1}{p} \left[1 - F_1(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0) \right], 1|x_0 \right) dz_1^{-\gamma_1(x_0)} \\
&= - \int_0^\infty R(z_1, 1|x_0) dz_1^{-\gamma_1(x_0)} \\
&\quad - \int_0^\infty \left[R_{1/p} \left(\frac{1}{p} \left[1 - F_1(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0) \right], 1|x_0 \right) \right. \\
&\quad \quad \left. - R \left(\frac{1}{p} \left[1 - F_1(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0) \right], 1|x_0 \right) \right] dz_1^{-\gamma_1(x_0)} \\
&\quad - \int_0^\infty \left[R \left(\frac{1}{p} \left[1 - F_1(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0) \right], 1|x_0 \right) - R(z_1, 1|x_0) \right] dz_1^{-\gamma_1(x_0)} \\
&=: - \int_0^\infty R(z_1, 1|x_0) dz_1^{-\gamma_1(x_0)} + T_{10,n} + T_{11,n}.
\end{aligned}$$

Clearly $T_{10,n} = O(p^{-\tau}) = o\left(\frac{1}{\sqrt{kh_n^d}}\right)$, by our assumptions on k and p . For $T_{11,n}$, we follow the lines of proof of the second part of Lemma 5.4 in Goegebeur et al. (2020), using the Lipschitz property of the function R , for $T_n \rightarrow \infty$, we have

$$\begin{aligned}
\sqrt{kh_n^d} |T_{11,n}| &\leq \sqrt{kh_n^d} \left| \int_0^{T_n} \left[R \left(\frac{1}{p} \left[1 - F_1 \left(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0 \right) \right], 1|x_0 \right) - R(z_1, 1|x_0) \right] dz_1^{-\gamma_1(x_0)} \right| \\
&\quad + \sqrt{kh_n^d} \left| \int_{T_n}^\infty \left[R \left(\frac{1}{p} \left[1 - F_1 \left(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0 \right) \right], 1|x_0 \right) - R(z_1, 1|x_0) \right] dz_1^{-\gamma_1(x_0)} \right| \\
&\leq -\sqrt{kh_n^d} \int_0^{T_n} \left| \frac{1}{p} \left[1 - F_1 \left(z_1^{-\gamma_1(x_0)} U_1(1/p|x_0)|x_0 \right) \right] - z_1 \right| dz_1^{-\gamma_1(x_0)} \\
&\quad + 2 \sup_{z_1 \geq 0} R(z_1, 1|x_0) \sqrt{kh_n^d} T_n^{-\gamma_1(x_0)} \\
&\leq -\sqrt{kh_n^d} \frac{|\delta_1 \left(U_1 \left(\frac{1}{p} |x_0 \right) |x_0 \right)|}{|\gamma_1(x_0) + \delta_1 \left(U_1 \left(\frac{1}{p} |x_0 \right) |x_0 \right)|} \int_0^{T_n} z_1 \left| \frac{\delta_1 \left(z_1^{-\gamma_1(x_0)} U_1 \left(\frac{1}{p} |x_0 \right) |x_0 \right)}{\delta_1 \left(U_1 \left(\frac{1}{p} |x_0 \right) |x_0 \right)} - 1 \right| dz_1^{-\gamma_1(x_0)} \\
&\quad + C \sqrt{kh_n^d} T_n^{-\gamma_1(x_0)} \\
&\leq C \sqrt{kh_n^d} \left| \delta_1 \left(U_1 \left(\frac{1}{p} |x_0 \right) |x_0 \right) \right| T_n^{1-\gamma_1(x_0)+(\beta_1(x_0)+\varepsilon)\gamma_1(x_0)} + C \sqrt{kh_n^d} T_n^{-\gamma_1(x_0)} \\
&\leq C \sqrt{kh_n^d} \left| \delta_1 \left(U_1 \left(\frac{n}{k} |x_0 \right) |x_0 \right) \right| T_n^{1-\gamma_1(x_0)+(\beta_1(x_0)+\varepsilon)\gamma_1(x_0)} + C \sqrt{kh_n^d} T_n^{-\gamma_1(x_0)},
\end{aligned}$$

for n large. Then, if α and Δ are chosen as stated in Theorem 2.2 and $T_n = n^\kappa$ with κ chosen such that

$$\frac{\alpha - \Delta d}{2\gamma_1(x_0)} < \kappa < \frac{2(1 - \alpha)\gamma_1(x_0)\beta_1(x_0) - (\alpha - \Delta d)}{2[1 - \gamma_1(x_0) + (\beta_1(x_0) + \varepsilon)\gamma_1(x_0)]},$$

we have

$$\frac{\theta_p(x_0)}{U_1(1/p|x_0)} = - \int_0^\infty R(z_1, 1|x_0) dz_1^{-\gamma_1(x_0)} + o\left(\frac{1}{\sqrt{kh_n^d}}\right). \quad (18)$$

Combining (15), (16), (17) and (18), we deduce that

$$\frac{\left(\frac{k}{np}\right)^{\gamma_1(x_0)} \theta_{k/n}(x_0)}{\theta_p(x_0)} - 1 = o\left(\frac{1}{\sqrt{kh_n^d}}\right). \quad (19)$$

Finally, decomposition (13) combined with Theorem 5.1, (14) and (19) achieves the proof of Theorem 2.2.

Acknowledgement

The research of Armelle Guillou was supported by the French National Research Agency under the grant ANR-19-CE40-0013-01/ExtremReg project and an International Emerging Action (IEA-00179).

References

- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley.
- Beirlant, J., Joossens, E. and Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, *Journal of Statistical Planning and Inference*, 139, 2800–2815.
- Cai, J.J., Einmahl, J.H.J., de Haan, L. and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 417–442.
- Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, 19, 2557–2589.
- Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves, *TEST*, 20, 311–333.
- Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2018a). Local robust estimation of the Pickands dependence function, *Annals of Statistics*, 46, 2806–2843.
- Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2018b). Local estimation of the conditional stable tail dependence function, *Scandinavian Journal of Statistics*, 45, 590–617.

Gardes, L. and Girard, S. (2015). Nonparametric estimation of the conditional tail copula, *Journal of Multivariate Analysis*, 137, 1–16.

Goegebeur, Y., Guillou, A., Ho, N.K.L. and Qin, J. (2020). Conditional marginal expected shortfall, <https://hal.archives-ouvertes.fr/hal-02272392>.

Goegebeur, Y., Guillou, A. and Qin, J. (2019). Bias-corrected estimation for conditional Pareto-type distributions with random right censoring, *Extremes*, 22, 459–498.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory, an introduction*, Springer.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, 5, 1163–1174.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, 3, Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes, with applications to statistics*, Springer Series in Statistics, Springer-Verlag, New York.

Yao, Q. (1999). Conditional predictive regions for stochastic processes, Technical report, University of Kent at Canterbury, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.2449&rep=rep1&type=pdf>.