Electronic Journal of Statistics
Vol. 14 (2020) 4421–4456
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1779

Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model

Aboubacrène Ag Ahmad[*], **El Hadji Deme**[*], **Aliou Diop**[*]

Université Gaston Berger, LERSTAD, UFR SAT, BP 234, Saint-Louis, Sénégal.
e-mail: aboubacren@yahoo.fr; eltinoyaw@yahoo.fr; alioudiop52@yahoo.fr

Stéphane Girard[†,‡] **and Antoine Usseglio-Carleve**[†]

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.
e-mail: Stephane.Girard@inria.fr; Antoine.Usseglio-Carleve@inria.fr

Abstract: We consider a location-dispersion regression model for heavy-tailed distributions when the multidimensional covariate is deterministic. In a first step, nonparametric estimators of the regression and dispersion functions are introduced. This permits, in a second step, to derive an estimator of the conditional extreme-value index computed on the residuals. Finally, a plug-in estimator of extreme conditional quantiles is built using these two preliminary steps. It is shown that the resulting semi-parametric estimator is asymptotically Gaussian and may benefit from the same rate of convergence as in the unconditional situation. Its finite sample properties are illustrated both on simulated and real tsunami data.

MSC2020 subject classifications: Primary 62G32; secondary 62G30, 62E20.
Keywords and phrases: Semi-parametric estimation, regression and dispersion functions, tail-index, extreme conditional quantile.

Received September 2020.

Contents

arXiv: 2010.00000
[*]This work is supported by CEA-MITIC and Inria via the SIMERGE project.
[†]This work is supported the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02) and under the grant ANR-19-CE40-0013-01/ExtremReg project.
[‡]S. Girard acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French Ecole polytechnique and its Foundation and sponsored by BNP Paribas.

1. Introduction

The modeling of extreme events arises in many fields such as finance, insurance or environmental science. A recurrent statistical problem is then the estimation of extreme quantiles associated with a random variable Y, see the reference books [1, 13, 24]. In many situations, Y is recorded simultaneously with a multidimensional covariate $x \in \mathbb{R}^d$, the goal being to describe how tail characteristics such as extreme quantiles or small exceedance probabilities of the response variable Y may depend on the explanatory variable x. Motivating examples include the study of extreme rainfall as a function of the geographical location [17], the assessment of the optimal cost of the delivery activity in postal services [7], the analysis of longevity [30], the description of the upper tail of claim size distributions [1], the modeling of extremes in environmental time series [37], etc.

Here, we focus on the challenging situation where Y given x is heavy-tailed. Without additional assumptions on the pair (Y, x), the estimation of extreme conditional quantiles is addressed using nonparametric methods, see for instance the recent works of [9, 19, 21]. These methods may however suffer from the curse of dimensionality which is compounded in distribution tails by the fact that observations are rare by definition. These difficulties can be partially overcome by considering parametric models [11, 5]. Semi-parametric methods have also been considered for trend modeling in extreme events [10, 27]: A nonparametric regression model of the trend is combined with a parametric model for extreme values.

Our approach belongs to this second line of works. We assume that the response variable and the covariate are linked by a location-dispersion regression model $Y = a(x) + b(x)Z$, see [39], where Z is a heavy-tailed random variable. This model is flexible since (i) no parametric assumptions are made on $a(\cdot)$, $b(\cdot)$ and Z, (ii) it allows for heteroscedasticity via the function $b(\cdot)$. Moreover, another feature of this model is that Y inherits its tail behavior from Z and thus does not depend on the covariate x. We propose to take profit of this important property to decouple the estimation of the nonparametric and extreme structures. As a consequence, we shall show that the resulting semi-parametric estimators of extreme conditional quantiles of Y given x are asymptotically Gaussian and may benefit from the same rate of convergence as in the unconditional situation. A similar idea is implemented in [29]: An extreme-value distribution with constant extreme-value index is fitted to standardized rainfall maxima. The theoretical study of heteroscedastic extremes has been initiated in [26] and further developed in [12, 15] through the introduction of a proportional tails model. The results were applied to trend detection in rainfalls and stock market returns.

This paper is organized as follows. The location-dispersion regression model for heavy-tailed distributions is presented in more details in Section 2. The associated inference methods are described in Section 3: Estimation of the regression and dispersion functions, estimation of the conditional tail-index and extreme conditional quantiles. Asymptotic results are provided in Section 4 while the finite sample behavior of the estimators is illustrated in Section 5 on simulated data and in Section 6 on tsunami data. Proofs are postponed to the Appendix.

2. Location-dispersion regression model for heavy-tailed distributions

We consider the class of location-dispersion regression models, where the relation between a random response variable $Y \in \mathbb{R}$ and a deterministic covariate vector $x \in \Pi \subset \mathbb{R}^d$, $d \geq 1$ is given by

$$Y = a(x) + b(x)Z. \tag{2.1}$$

The real random variable Z is assumed to be heavy-tailed. Denoting by $\bar{F}_Z$ its survival function, one has

$$\bar{F}_Z(z) = z^{-1/\gamma} L(z),\ z > 0. \tag{2.2}$$

Here, $\gamma > 0$ is called the conditional tail-index and L is a slowly-varying function at infinity *i.e.* for all $t > 0$,

$$\lim_{z\to\infty} \frac{L(tz)}{L(z)} = 1.$$

$\bar{F}_Z$ is said to be regularly varying at infinity with index $-1/\gamma$. This property is denoted for short by $\bar{F}_Z \in \mathcal{RV}_{-1/\gamma}$, see [3] for a detailed account on regular variations. Model (2.1) has been introduced by [39] in the random design setting where the location function $a : \Pi \to \mathbb{R}$ and the scaling function $b : \Pi \to \mathbb{R}^+ \setminus \{0\}$ are referred to as the regression and dispersion functions respectively. Combining (2.1) and (2.2) yields

$$\bar{F}_Y(y \mid x) := \mathbb{P}(Y > y \mid x) = \bar{F}_Z\left(\frac{y - a(x)}{b(x)}\right) = \left(\frac{y - a(x)}{b(x)}\right)^{-1/\gamma} L\left(\frac{y - a(x)}{b(x)}\right), \tag{2.3}$$

for $y \geq y_0(x) > a(x)$ where the functions $a(\cdot)$, $b(\cdot)$ and the conditional tail-index γ are unknown. We thus obtain a semi-parametric location-dispersion regression model for the (heavy) tail of Y given x. The main assumption is that the conditional tail-index γ is independent of the covariate. On the one hand, the proposed semi-parametric heteroscedastic modeling offers more flexibility than purely parametric approaches. On the other hand, the location-dispersion structure may circumvent the curse of dimensionality and assuming a constant conditional tail-index γ should yield more reliable estimates in small sample contexts than purely nonparametric approaches. Let us also note that, from (2.2) and (2.3), the regular variation property yields $\bar{F}_Y(y \mid x)/\bar{F}_Z(y) \to b(x)^{1/\gamma}$ as $y \to \infty$. The location-dispersion regression model can thus be interpreted as a particular case of the proportional tails model [12] with scedasis function $b(\cdot)^{1/\gamma}$. The practical consequences of this point are further discussed in Section 5.

Starting with an independent n-sample $\{(Y_1, x_1), \dots, (Y_n, x_n)\}$ from (2.1), it is clear that, since Z is not observed, $a(\cdot)$ and $b(\cdot)$ may only be estimated up to additive and multiplicative factors. This identifiability issue can be fixed by introducing some constraints on the distribution of Z. To this end, for all $\alpha \in (0, 1)$ consider $q_Z(\alpha) = \inf\{z \in \mathbb{R}; \bar{F}_Z(z) \leq \alpha\}$ the αth quantile of Z and let $(\mu_1, \mu_2, \mu_3) \in (0, 1)^3$ such that $\mu_3 < \mu_1$ and

$$q_Z(\mu_2) = 0 \text{ and } q_Z(\mu_3) - q_Z(\mu_1) = 1. \tag{2.4}$$

Let us note that the constraint (2.4) can always be fulfilled with *i.e.* $\mu_3 = 1/4$, $\mu_2 = 1/2$ and $\mu_1 = 3/4$ up to an affine transformation of $a(\cdot)$, $b(\cdot)$ and Z such that (2.1) holds. From (2.1), for all $\alpha \in (0, 1)$, the conditional quantile of Y given $x \in \Pi$ is

$$q_Y(\alpha \mid x) = a(x) + b(x) q_Z(\alpha), \tag{2.5}$$

and therefore the regression and dispersion functions are defined in an unique way by

$$a(x) = q_Y(\mu_2 \mid x) \text{ and } b(x) = q_Y(\mu_3 \mid x) - q_Y(\mu_1 \mid x), \tag{2.6}$$

for all $x \in \Pi$. This remark is the starting point of the inference procedure described hereafter.

3. Inference

Let us denote by λ the Lebesgue measure and $\|\cdot\|$ a norm on $\mathbb{R}^d$, $d \geq 1$. Consider $\{(Y_1, x_1), \dots, (Y_n, x_n)\}$ a n-sample from (2.1): $Y_i = a(x_i) + b(x_i)Z_i$, $i = 1, \dots, n$ where $Z_1, \dots, Z_n$ are independent and identically distributed (iid) from the heavy-tailed distribution (2.2). We assume that the design points x_i, $i = 1, \dots, n$ are all distinct from each other and included in Π, a compact subset of $\mathbb{R}^d$ whose Lebesgue measure of the boundary is zero. Let $\{\Pi_i,\ i = 1, \dots, n\}$ be a partition of Π such that $x_i \in \Pi_i$. A three-stage inference procedure is adopted: The regression and dispersion functions are estimated nonparametrically in Paragraph 3.1, and the conditional tail-index is then computed from the residuals in Paragraph 3.2. Finally, the extreme conditional quantiles are derived by combining a plug-in method with Weissman's extrapolation device [40] in Paragraph 3.3.

3.1. Estimation of the regression and dispersion functions

The proposed procedure relies on the choice of a smoothing estimator for the conditional quantiles. Here, a kernel estimator for $\bar{F}_Y(y \mid x)$ is considered (see for instance [33, 34]). For all $(x, y) \in \Pi \times \mathbb{R}$ let

$$\hat{\bar{F}}_{n,Y}(y \mid x) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i > y\}} \int_{\Pi_i} K_h(x - t)dt, \tag{3.1}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, $K_h(\cdot) := K(\cdot/h)/h^d$ with K a density function on $\mathbb{R}^d$ called a kernel. The associated smoothing parameter $h = h_n \to 0$ as $n \to \infty$ is a nonrandom sequence called the bandwidth. The corresponding estimator of $q_Y(\alpha \mid x)$ is defined for all $(x, \alpha) \in \Pi \times (0, 1)$ by

$$\hat{q}_{n,Y}(\alpha \mid x) = \hat{\bar{F}}^{\leftarrow}_{n,Y}(\alpha \mid x) := \inf\{y;\ \hat{\bar{F}}_{n,Y}(y \mid x) \leq \alpha\}. \tag{3.2}$$

Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function have been extensively investigated, see, for example [2, 35, 38], among others. In view of (2.6), the regression and dispersion functions are estimated by

$$\hat{a}_n(x) = \hat{q}_{n,Y}(\mu_2 \mid x) \text{ and } \hat{b}_n(x) = \hat{q}_{n,Y}(\mu_3 \mid x) - \hat{q}_{n,Y}(\mu_1 \mid x), \tag{3.3}$$

for all $x \in \Pi$.

3.2. Estimation of the conditional tail-index

The non-observed $Z_1, \ldots, Z_n$ are estimated by the residuals

$$\hat{Z}_i = (Y_i - \hat{a}_n(x_i))/\hat{b}_n(x_i), \tag{3.4}$$

for all $i = 1, \ldots, n$ where $\hat{a}_n(\cdot)$ and $\hat{b}_n(\cdot)$ are given in (3.3). In practice, nonparametric estimators can suffer from boundary effects [6, 31] and therefore only design points sufficiently far from the boundary of Π are considered. More specifically, consider $\tilde{\Pi}^{(n)} = \{x \in \mathbb{R}^d,$ such that $B(x, h) \subset \Pi\}$ the erosion of the set Π by the ball $B(0, h)$ centered at 0 and with radius h, see [36] for further details on mathematical morphology. Denote by I_n the set of indices associated with such design points $I_n = \{i \in \{1, \ldots, n\}$ such that $x_i \in \tilde{\Pi}^{(n)}\}$ and let $m_n = \text{card}(I_n)$. It can be shown that $m_n = n(1 + O(h))$, see Lemma A.3 in the Appendix.

Finally, let (k_n) be an intermediate sequence of integers, *i.e.* such that $1 < k_n \leq n$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. The $(k_n + 1)$ top order statistics associated with the pseudo-observations $\hat{Z}_i$, $i \in I_n$ are denoted by $\hat{Z}_{m_n-k_n,m_n} \leq \cdots \leq \hat{Z}_{m_n,m_n}$. The conditional tail-index is estimated using a Hill-type statistic [28]:

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=0}^{k_n-1} \log \hat{Z}_{m_n-i,m_n} - \log \hat{Z}_{m_n-k_n,m_n}, \tag{3.5}$$

built on non iid pseudo-observations.

3.3. Estimation of extreme conditional quantiles

Clearly, the purely nonparametric estimator (3.2) cannot estimate consistently extreme quantiles of levels α_n arbitrarily small. For instance, when $n\alpha_n \to 0$, the extreme quantile is likely to be larger than the maximum observation. In such a case, an extrapolation technique is necessary to estimate the so-called extreme conditional quantile $q_Y(\alpha_n \mid x)$. To this end, we propose to take profit of the structure of the location-dispersion regression model (2.5) to define the plugin estimator

$$\tilde{q}_{n,Y}(\alpha_n \mid x) = \hat{a}_n(x) + \hat{b}_n(x)\hat{q}_{n,Z}(\alpha_n), \tag{3.6}$$

where $\hat{a}_n(x)$ and $\hat{b}_n(x)$ are given in (3.3) and $\hat{q}_{n,Z}(\alpha_n)$ is the Weissman type estimator [40]:

$$\hat{q}_{n,Z}(\alpha_n) = \hat{Z}_{m_n-k_n,m_n} \left(\frac{\alpha_n m_n}{k_n} \right)^{-\hat{\gamma}_n}. \tag{3.7}$$

Again, it should be noted that $\hat{q}_{n,Z}(\alpha_n)$ is computed from the non iid pseudo-observations $\hat{Z}_i$, $i \in I_n$. Finally, by construction, the semi-parametric estimator (3.6) cannot suffer from quantile crossing, a phenomenon which can occur with quantile regression techniques.

4. Main results

The following general assumptions are required to establish the asymptotic behavior of the estimators. The first one gathers all the conditions to define a location-dispersion regression model for heavy-tailed distributions in a multidimensional fixed design setting.

(A.1) $(Y_1, x_1), \ldots, (Y_n, x_n)$ are independent observations from the location-dispersion regression model for heavy-tailed distributions defined by (2.1), (2.2) and (2.4) and such that

$$\max_{i=1,\ldots,n} \left| \lambda(\Pi_i) - \frac{\lambda(\Pi)}{n} \right| = o(1/n), \tag{4.1}$$

$$\max_{i=1,\ldots,n} \sup_{(s,t)\in\Pi_i^2} \|s-t\| = O(n^{-1/d}). \tag{4.2}$$

We refer to [33, 34] for this definition of the multidimensional fixed design setting.

The second assumption is a regularity condition.

(A.2) The functions $a(\cdot)$ and $b(\cdot)$ are twice continuously differentiable on Π, $b(\cdot)$ is lower bounded on Π, $b(t) \geq b_m > 0$ for all $t \in \Pi$, and the survival function $\bar{F}_Z(\cdot)$ is twice continuously differentiable on $\mathbb{R}$.

Under **(A.1)** and **(A.2)**, the quantile function $q_Z(\cdot)$ and the density $f_Z(\cdot) = -\bar{F}'_Z(\cdot)$ exist and we let $H_Z(\cdot) := 1/f_Z(q_Z(\cdot))$ the quantile density function and $U_Z(\cdot) = q_Z(1/\cdot)$ the tail quantile function of Z. Moreover, the conditional survival function of Y is twice continuously differentiable with respect to its second argument. The next assumption is standard in the nonparametric kernel estimation framework.

(A.3) K is a bounded and even density with symmetric support $S \subset B(0,1)$ the unit ball of $\mathbb{R}^d$ and verifying the Lipschitz property: There exists $c_K > 0$ such that

$$|K(u) - K(v)| \leq c_K \|u - v\|,$$

for all $(u,v) \in S^2$.

Under **(A.3)**, let $\|K\|_\infty = \sup_{t \in S} K(t)$ and $\|K\|_2 = \left(\int_S K^2(t)dt\right)^{1/2}$. Finally, a second-order condition is introduced, see for instance [24, eq (3.2.5)]:

(A.4) For all $t > 0$, as $z \to \infty$,

$$\frac{U_Z(tz)}{U_Z(z)} - t^\gamma \sim A(z) t^\gamma \frac{t^\rho - 1}{\rho},$$

where $\gamma > 0$, $\rho < 0$ and A is a positive or negative function such that $A(z) \to 0$ as $z \to \infty$.

From [3, Theorem 1.5.12], property (2.2) is equivalent to $U_Z \in \mathcal{RV}_\gamma$, that is $U_Z(tz)/U_Z(z) \to t^\gamma$ as $z \to \infty$ for all $t > 0$. The role of the second-order condition **(A.4)** is thus to control the rate of the previous convergence thanks to the function $A(\cdot)$. Moreover, it can be shown that $|A|$ is regularly varying with index ρ, see [24, Lemma 2.2.3]. It is then clear that ρ, referred to as the (conditional) second-order parameter, is a crucial quantity, tuning the rate of convergence of most extreme-value estimators, see [24, Chapter 3] for examples. A list of distributions satisfying **(A.4)** is provided in Table 1 together with the associated values of γ and ρ. Similarly to [34], the dimension $d = 4$ plays a special role and we thus introduce for all $d \geq 1$:

$$\kappa(d) = \left| \begin{array}{cl} 4 & \text{if} \quad d \leq 4 \\ 2d/(d-2) & \text{if} \quad d \geq 4. \end{array} \right.$$

Our first result states the joint asymptotic normality of the estimators (3.3) of the regression and dispersion functions.

Theorem 4.1. *Assume* **(A.1)**, **(A.2)**, **(A.3)** *hold and* $f_Z(q_Z(\mu_j)) > 0$ *for* $j \in \{1,2,3\}$. *If* $nh^d \to \infty$ *and* $nh^{d+\kappa(d)} \to 0$ *as* $n \to \infty$ *then, for all sequence*

Table 1
A list of heavy-tailed distributions satisfying **(A.4)** *with the associated values of γ and ρ. $\Gamma(\cdot)$ and $B(\cdot,\cdot)$ denote the Gamma and Beta functions respectively.*

Distribution (parameters)	Density function	γ	ρ
Generalised Pareto ($\sigma, \xi > 0$)	$\sigma^{-1}\left(1+\xi t/\sigma\right)^{-1-1/\xi}$ ($t>0$)	ξ	$-\xi$
Burr ($\alpha, \beta > 0$)	$\alpha\beta t^{\alpha-1}\left(1+t^{\alpha}\right)^{-\beta-1}$ ($t>0$)	$1/(\alpha\beta)$	$-1/\beta$
Fréchet ($\alpha > 0$)	$\alpha t^{-\alpha-1}\exp\left(-t^{-\alpha}\right)$ ($t>0$)	$1/\alpha$	-1
Fisher ($\nu_1, \nu_2 > 0$)	$\dfrac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2,\nu_2/2)} t^{\nu_1/2-1}(1+\nu_1 t/\nu_2)^{-(\nu_1+\nu_2)/2}$ ($t>0$)	$2/\nu_2$	$-2/\nu_2$
Inverse Gamma ($\alpha, \beta > 0$)	$\dfrac{\beta^{\alpha}}{\Gamma(\alpha)} t^{-\alpha-1}\exp(-\beta/t)$ ($t>0$)	$1/\alpha$	$-1/\alpha$
Student ($\nu > 0$)	$\dfrac{1}{\sqrt{\nu\pi}}\dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\dfrac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$1/\nu$	$-2/\nu$

$(t_n) \subset \tilde{\Pi}^{(n)}$,

$$\frac{\sqrt{nh^d}}{b(t_n)}\begin{pmatrix}\hat{a}_n(t_n)-a(t_n)\\ \hat{b}_n(t_n)-b(t_n)\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^2},\, \lambda(\Pi)\|K\|_2^2\,\Sigma\right),$$

where the coefficients of the (symmetric) matrix Σ are given by

$$\begin{aligned}
\Sigma_{1,1} &= \mu_2(1-\mu_2)H_Z^2(\mu_2),\\
\Sigma_{1,2} &= \mu_2(1-\mu_1)H_Z(\mu_1)H_Z(\mu_2) - \mu_3(1-\mu_2)H_Z(\mu_2)H_Z(\mu_3),\\
\Sigma_{2,2} &= \mu_1(1-\mu_1)H_Z^2(\mu_1) - 2\mu_3(1-\mu_1)H_Z(\mu_1)H_Z(\mu_3) + \mu_3(1-\mu_3)H_Z^2(\mu_3).
\end{aligned}$$

A uniform consistency result can also be established:

Theorem 4.2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. If $nh^d/\log n \to \infty$ and $nh^{d+\kappa(d)}/\log n \to 0$ as $n \to \infty$, then,*

$$\begin{aligned}
\sqrt{\frac{nh^d}{\log n}}\max_{i\in I_n}\left|\frac{\hat{a}_n(x_i)-a(x_i)}{b(x_i)}\right| &= O_{\mathbb{P}}(1) \text{ and}\\
\sqrt{\frac{nh^d}{\log n}}\max_{i\in I_n}\left|\frac{\hat{b}_n(x_i)-b(x_i)}{b(x_i)}\right| &= O_{\mathbb{P}}(1).
\end{aligned}$$

As a consequence of Theorem 4.2, one can prove that the residuals $\hat{Z}_i = (Y_i - \hat{a}_n(x_i))/\hat{b}_n(x_i)$, see (3.4), are close to the unobserved Z_i, $i = 1, \ldots, n$.

Corollary 4.1. *Under the assumptions of Theorem 4.2, for all $i \in I_n$,*

$$|\hat{Z}_i - Z_i| \le R_{n,i}(1+|Z_i|), \text{ where } \max_{i\in I_n} R_{n,i} = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right) = o_{\mathbb{P}}(1).$$

Our next main result provides the asymptotic normality of the conditional tail-index estimator (3.5) and the Weissman estimator (3.7) computed on the residuals.

Theorem 4.3. *Assume* **(A.1)**-**(A.4)** *hold. Let* (k_n) *be an intermediate sequence of integers such that* $nh^d/(k_n \log n) \to \infty$, $nh^{d+\kappa(d)}/\log n \to 0$ *and* $\sqrt{k_n}A(n/k_n) \to \beta \in \mathbb{R}$ *as* $n \to \infty$. *Then,*

(i) $\sqrt{k_n}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2)$.
(ii) For all sequence $(\alpha_n) \subset (0,1)$ *such that* $n\alpha_n/k_n \to 0$ *and* $\log(n\alpha_n)/\sqrt{k_n} \to 0$ *as* $n \to \infty$,

$$\frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)}\left(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2).$$

It appears that, in the location-dispersion regression model, the tail-index can be estimated at the same rate $1/\sqrt{k_n}$ as in iid case, see [22] for a review. As expected, this semi-parametric framework is a more favorable situation than the purely nonparametric one for the estimation of the conditional tail-index where the rate of convergence $1/\sqrt{k_n h^d}$ is impacted by the covariate, see for instance [9, Corollary 1 & 2], [8, Theorem 3] and [21, Theorem 2]. To be more specific, remark first that conditions $nh^d/(k_n \log n) \to \infty$ and $nh^{d+\kappa(d)}/\log n \to 0$ imply that $k_n = o\left((n/\log n)^{\kappa(d)/(d+\kappa(d))}\right)$. Second, following [24, Eq. (3.2.10)], if A is a power function, then condition $\sqrt{k_n}A(n/k_n) \to \beta$ as $n \to \infty$ yields $k_n = O\left(n^{-2\rho/(1-2\rho)}\right)$. As a conclusion, up to logarithmic factors, possible choices of sequences are then

$$h_n = n^{-1/(d+\kappa(d))} \text{ and } k_n = n^{1/(1+\max\{d/\kappa(d), -1/(2\rho)\})}. \tag{4.3}$$

If $\rho \geq -\kappa(d)/(2d)$, the rate of convergence of $\hat{\gamma}_n$ is thus $n^{\rho/(1-2\rho)}$ up to logarithmic factors which is the classical rate for estimators of the tail-index, see for instance [25, Remark 3]. For instance, in the situation where the dimension of the covariate is $d \leq 2$, then the $n^{\rho/(1-2\rho)}$ rate is reached as soon as $\rho \geq -1$. This corresponds to the challenging situation where a high bias is expected in the estimation which may occur for most usual distributions, depending on their shape parameters, see Table 1.

Theorem 4.4 states the asymptotic normality of the estimator (3.6) of extreme conditional quantiles of $Y \mid x$.

Theorem 4.4. *Assume* **(A.1)**-**(A.4)** *hold and* $f_Z(q_Z(\mu_j)) > 0$ *for* $j \in \{1,2,3\}$. *Let* (k_n) *be an intermediate sequence of integers. Suppose* $nh^d/(k_n \log n) \to \infty$, $nh^{d+\kappa(d)} \to 0$ *and* $\sqrt{k_n}A(n/k_n) \to \beta \in \mathbb{R}$ *as* $n \to \infty$. *Then, for all sequences* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(\alpha_n) \subset (0,1)$ *such that* $n\alpha_n/k_n \to 0$ *and* $\log(n\alpha_n)/\sqrt{k_n} \to 0$ *as* $n \to \infty$,

$$\frac{\sqrt{k_n}}{q_Z(\alpha_n)\log\left(\frac{k_n}{n\alpha_n}\right)}\left(\frac{\tilde{q}_{n,Y}(\alpha_n \mid t_n) - q_Y(\alpha_n \mid t_n)}{b(t_n)}\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2). \tag{4.4}$$

Remark that $b(t_n)q_Z(\alpha_n) \sim a(t_n) + b(t_n)q_Z(\alpha_n) = q_Y(\alpha_n \mid t_n)$ and therefore (4.4) can be rewritten as

$$\frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)}\left(\frac{\tilde{q}_{n,Y}(\alpha_n \mid t_n)}{q_Y(\alpha_n \mid t_n)} - 1\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2).$$

As a comparison, the rate of convergence of purely nonparametric methods involves an extra $h^{d/2}$ factor, see for instance [18, Theorem 3] or [8, Theorem 3]. The location-dispersion regression model allows to dampen this vexing effect of the dimensionality.

Finally, a uniform consistency result is also available:

Theorem 4.5. *Assume* **(A.1)**-**(A.4)** *hold. Let* (k_n) *be an intermediate sequence of integers. Suppose* $nh^d/(k_n \log n) \to \infty$, $nh^{d+\kappa(d)}/\log n \to 0$ *and* $\sqrt{k_n}A(n/k_n) \to \beta \in \mathbb{R}$ *as* $n \to \infty$. *Then, for all sequence* $(\alpha_n) \subset (0,1)$ *such that* $n\alpha_n/k_n \to 0$ *and* $\log(n\alpha_n)/\sqrt{k_n} \to 0$ *as* $n \to \infty$,

$$\frac{\sqrt{k_n}}{q_Z(\alpha_n)\log\left(\frac{k_n}{n\alpha_n}\right)}\max_{i\in I_n}\left|\frac{\tilde{q}_{n,Y}(\alpha_n \mid x_i) - q_Y(\alpha_n \mid x_i)}{b(x_i)}\right| = O_{\mathbb{P}}(1).$$

5. Illustration on simulations

5.1. Experimental design

We propose to illustrate the finite-sample performance of the estimators of the conditional tail-index and the extreme conditional quantiles on simulated data from the location-dispersion regression model. For that purpose, set $d = 2$, $\Pi = [0,1]^2$ and define the regression and dispersion functions respectively by $a(x) = 1 - \cos(\pi(x^{(1)} + x^{(2)}))$ and $b(x) = \exp(-(x^{(1)} - 0.5)^2 - (x^{(2)} - 0.5)^2)$, for $x = (x^{(1)}, x^{(2)}) \in \Pi$. Let $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$. Two distributions are considered for the heavy-tailed random variable Z:

- Let Z_0 be a standard Student-t_ν random variable where $\nu \in \{1,2,4\}$ denotes the degrees of freedom (df) and introduce $Z = Z_0/(2q_{Z_0}(\mu_3))$ the associated rescaled Student random variable. Symmetry arguments yield $q_Z(\mu_2) = 0$, $q_Z(\mu_1) = -q_Z(\mu_3)$ and $q_Z(\mu_3) = q_{Z_0}(\mu_3)/(2q_{Z_0}(\mu_3)) = 1/2$ by construction. Therefore (2.4) holds. This choice also ensures that Z is heavy-tailed with conditional tail-index $\gamma = 1/\nu$ and that the second-order condition **(A.4)** holds with $\rho = -2/\nu$, see Table 1.
- Let Z_0 be a Burr random variable with parameters $\alpha \in \{1,2,4\}$ and $\beta = 1$. We then introduce the translated and rescaled random variable

$$Z = \frac{Z_0 - \left(\mu_2^{-1} - 1\right)^{1/\alpha}}{\left(\mu_3^{-1} - 1\right)^{1/\alpha} - \left(\mu_1^{-1} - 1\right)^{1/\alpha}},$$

  such as (2.4) holds. The second-order condition **(A.4)** is also fulfilled with $\gamma = 1/\alpha$ and $\rho = -1$, see Table 1.

The design points x_i, $i = 1, \dots, n$ are chosen on a regular grid on the unit square Π. The kernel function K is the product of two quartic (or biweight) kernels:

$$K(u, v) = \left(\frac{15}{16}\right)^2 \left(1 - u^2\right)^2 \left(1 - v^2\right)^2 \mathbb{1}_{\{|u| \le 1\}} \mathbb{1}_{\{|v| \le 1\}},$$

where $(u, v) \in \mathbb{R}^2$. We set $\|x\| = \max(|x^{(1)}|, |x^{(2)}|)$ so that $\tilde{\Pi}^{(n)} = [h, 1-h]^2$. The bandwidth is fixed to $h_n^* = \sigma n^{-1/6}$ following [4] and in accordance with (4.3), where $\sigma = 12^{-1/2}$ is the standard deviation of the coordinates of the design points. This choice is optimal for density estimation in the Gaussian case, but is also known to provide good results in other settings.

5.2. Graphical illustrations

In all the experiments, $N = 100$ replications of a dataset of size $n = 10,000$ are considered. The estimation results for the regression and dispersion functions are depicted respectively on Figure 1 and Figure 2 in the situation where Z is Student-t_ν distributed for $\nu \in \{1, 2, 4\}$. The results are visually satisfying and seem independent from the degrees of freedom. This conclusion was expected since both estimators of $a(\cdot)$ and $b(\cdot)$ are based on non-extreme quantiles, they are thus robust with respect to heavy tails.

As already noticed in Section 2, in the context of proportional tails, both random variables Y and Z share the same conditional tail-index γ. This parameter can thus be estimated either by (3.5) (computed on the residuals $\hat{Z}_i$) or by the classical Hill estimator (computed on the response variables Y_i). The associated estimation results are displayed on Figure 3 as functions of the sample fraction k_n. It first appears that working on the residuals provides much better results in terms of bias than working on the initial response variable. Second, the tail-index estimator (3.5) has a stronger bias for larger values of ν. These empirical results are in line with the properties of the Student distribution. Indeed, the second-order parameter $\rho = -2/\nu$ being increasing with ν, the bias of the Hill-type estimator increases as well.

In practice, the estimation of the conditional tail-index and extreme conditional quantiles require the selection of the sample fraction k_n. This parameter is selected using a mean-squared error criterion. Assuming that $A(t) = ct^\rho$, the optimal value of k_n is given by

$$k_n^* = \left(\frac{\gamma^2(1-\rho)^2}{-2\rho c^2}\right)^{\frac{1}{1-2\rho}} n^{-\frac{2\rho}{1-2\rho}},$$

see [24, Section 3.2]. Since ρ may be difficult to estimate in practice, a miss-specified value $\rho = -1$ is considered in several works dealing with bias reduction of tail-index estimators, see for instance [14] or [23]. Letting moreover $c = \sqrt{2}$ and restricting ourselves to integer values, we end up with $k_n^* = \lfloor (\check{\gamma} n)^{2/3} \rfloor$ where $\check{\gamma}$ is a prior naive estimation of γ computed with $k_n = \lfloor n^{1/2} \rfloor$ and where $\lfloor \cdot \rfloor$ denotes the floor function. Such a choice of k_n^* fulfils the assumptions of

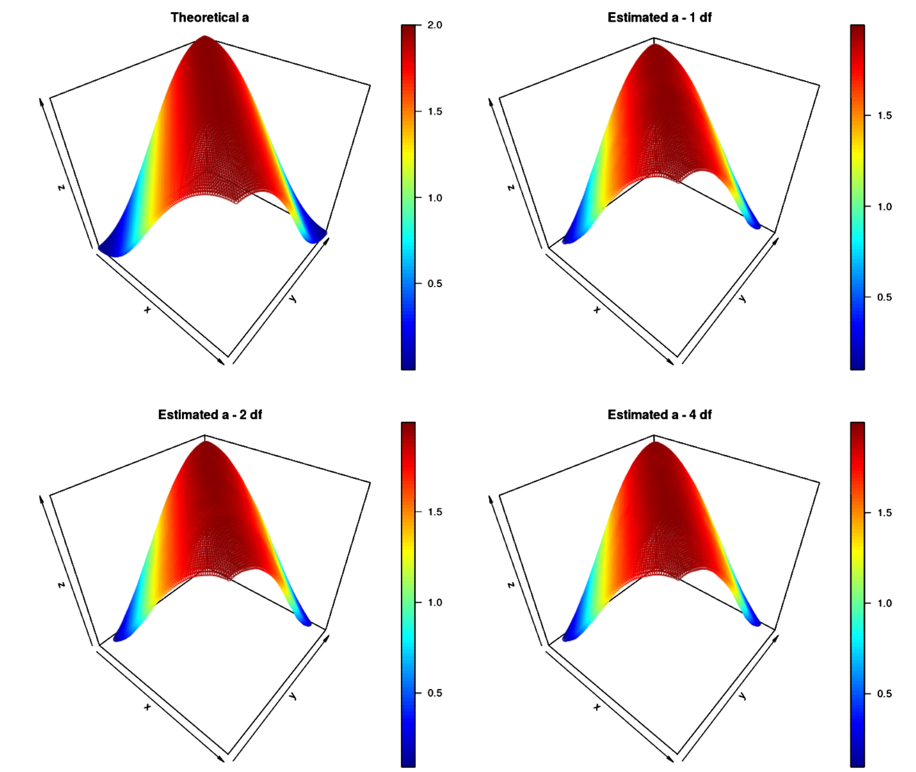

FIG 1. *Simulation results obtained on a Student-t_ν distribution. From top to bottom, left to right: Theoretical function $a(\cdot)$, and means over $N = 100$ replications of estimates $\hat{a}_n(\cdot)$ computed on $n = 10,000$ observations for $\nu \in \{1, 2, 4\}$. X-axis and y-axis range between 0 and 1, z-axis range between 0 and 2.*

Theorem 4.3–4.5 for all three considered Burr distributions and for Student-t_ν distributions with $\nu \in \{1, 2\}$. The constraints are violated in case of the Student-t_4 distribution in order to examine the robustness of the method with respect to the choice of the pair (h, k_n) which may be challenging in practice. The estimated conditional quantiles $q_Y(1/n \,|\, \cdot)$ of extreme level $\alpha_n = 1/n$ are displayed on Figure 4. As expected, the estimated extreme conditional quantiles all share the same shape despite different variation ranges.

5.3. Quantitative assessment

In this section, we propose to highlight the performances of the extreme conditional quantile estimator (3.6) thanks to a comparison with a purely nonparametric one. The nonparametric estimator is based on the ideas of the moving window approach introduced in [16]. For each $x \in \tilde{\Pi}^{(n)}$, a subsample $\left\{(Y_i^{\circledast}, x_i^{\circledast})\right\}_{i=1,\dots,n^{\circledast}} = \{(Y_i, x_i),\, 1 \le i \le n,\ \text{s.t.}\ \|x - x_i\| < h\}$ of size $n^{\circledast} =$

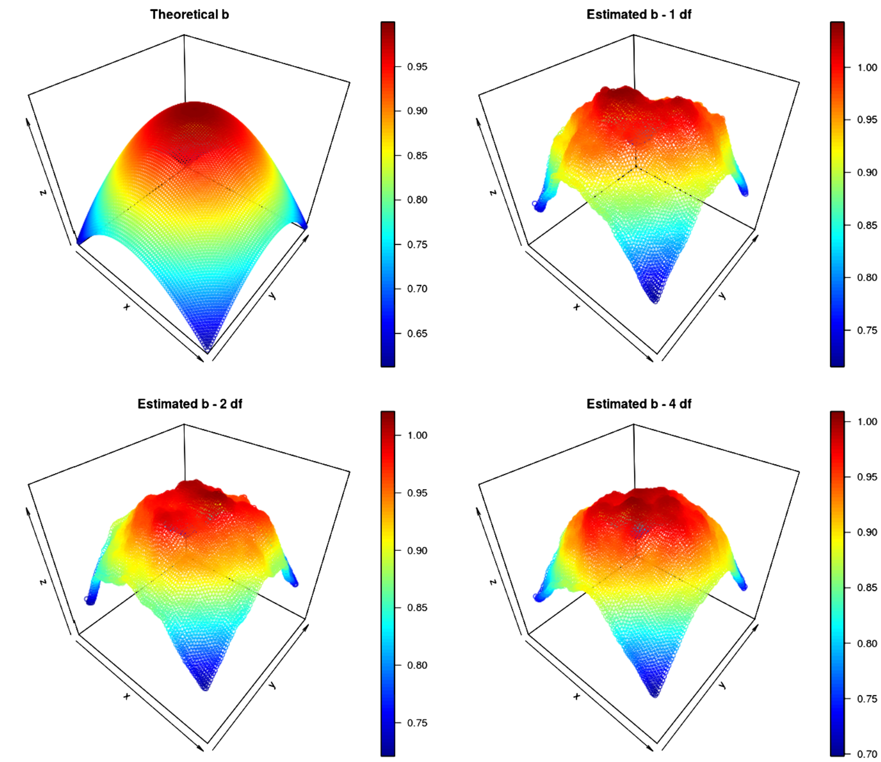

FIG 2. *Simulation results obtained on a Student-*t_ν *distribution. From top to bottom, left to right: Theoretical function* $b(\cdot)$*, and means over* $N = 100$ *replications of estimates* $\hat{b}_n(\cdot)$ *computed on* $n = 10,000$ *observations for* $\nu \in \{1, 2, 4\}$*. All three coordinates range between 0 and 1.*

$n^{\circledast}(x, h)$ is extracted from the initial sample. Letting $k_n^{\circledast} = \lfloor \sqrt{n^{\circledast}} \rfloor$, the conditional tail-index is estimated by the (local) Hill-type statistic

$$\hat{\gamma}_n^{\circledast}(x) = \frac{1}{k_n^{\circledast}} \sum_{i=0}^{k_n^{\circledast}-1} \log Y^{\circledast}_{n^{\circledast}-i, n^{\circledast}} - \log Y^{\circledast}_{n^{\circledast}-k_n^{\circledast}, n^{\circledast}},$$

and the extreme conditional quantile $q_Y(\alpha_n \,|\, x)$ is estimated by the associated Weissman-type statistic:

$$\hat{q}^{\circledast}_{n,Y}(\alpha_n \,|\, x) = Y^{\circledast}_{n^{\circledast}-k_n^{\circledast}, n^{\circledast}} \left(\frac{\alpha_n n^{\circledast}}{k_n^{\circledast}} \right)^{-\hat{\gamma}_n^{\circledast}(x)}.$$

Another option is to re-estimate γ and $q_Y(\alpha_n \,|\, x)$ taking $k_n^{\oplus} = \lfloor (\hat{\gamma}_n^{\circledast}(x) n^{\circledast})^{2/3} \rfloor$ in the above two estimators. The associated estimator of the extreme quantile is denoted by $\hat{q}^{\oplus}_{n,Y}(\alpha_n \,|\, x)$. The comparison between the true and estimated extreme conditional quantiles is based on a relative median-squared error (RMSE)

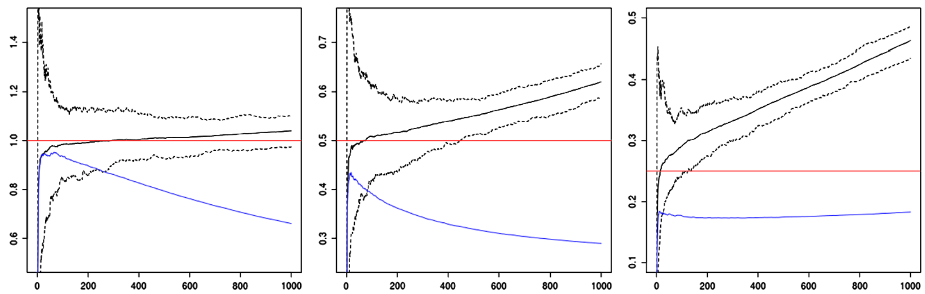

FIG 3. *Simulation results obtained on a Student-t_ν distribution for $\nu = 1$ (left), $\nu = 2$ (middle) and $\nu = 4$ (right). Mean estimate of the conditional tail-index (3.5) (continuous black line), associated 95% empirical confidence intervals (dotted lines) and mean Hill estimate computed on the response variable (continuous blue line), as functions of the sample fraction k_n. The true value $\gamma = 1/\nu$ is depicted by a red horizontal line.*

TABLE 2
Relative median squared errors associated with the estimation of the extreme conditional quantile $q_Y(1/n\,|\,\cdot)$. Results obtained with the semi-parametric estimator $\tilde{q}_{n,Y}$ and comparison with the purely nonparametric ones ($\hat{q}^{\circledast}_{n,Y}$, $\hat{q}^{\oplus}_{n,Y}$).

n	Student, $\nu = 1$	Student, $\nu = 2$	Student, $\nu = 4$
400	0.547 (0.890, 0.976)	0.129 (0.643, 0.630)	0.062 (0.442, 0.458)
1,600	0.138 (0.867, 0.893)	0.065 (0.533, 0.458)	0.020 (0.284, 0.352)
3,600	0.145 (0.855, 0.837)	0.048 (0.477, 0.431)	0.012 (0.226, 0.306)
6,400	0.061 (0.845, 0.776)	0.032 (0.456, 0.454)	0.011 (0.206, 0.253)
10,000	0.045 (0.820, 0.723)	0.026 (0.425, 0.435)	0.013 (0.184, 0.222)
n	Burr, $\alpha = 1$, $\beta = 1$	Burr, $\alpha = 2$, $\beta = 1$	Burr, $\alpha = 4$, $\beta = 1$
400	0.525 (0.746, 0.588)	0.197 (0.329, 0.285)	0.104 (0.129, 0.176)
1,600	0.182 (0.796, 0.637)	0.068 (0.348, 0.260)	0.038 (0.124, 0.168)
3,600	0.157 (0.825, 0.625)	0.056 (0.333, 0.264)	0.023 (0.118, 0.149)
6,400	0.096 (0.827, 0.591)	0.054 (0.311, 0.271)	0.020 (0.107, 0.122)
10,000	0.070 (0.845, 0.563)	0.030 (0.301, 0.262)	0.023 (0.102, 0.107)

computed on the $N = 100$ replications and the m_n design points in the square $\tilde{\Pi}^{(n)}$:

$$\operatorname{median}\left\{\operatorname{median}\left\{\left(\frac{\hat{q}^{[r]}_{n,Y}(\alpha_n\,|\,x_i)}{q_Y(\alpha_n\,|\,x_i)} - 1\right)^2,\ x_i \in \tilde{\Pi}^{(n)}\right\},\ r \in \{1,\dots,N\}\right\},$$

where $\hat{q}^{[r]}_{n,Y}(\alpha_n\,|\,\cdot)$ denotes either $\tilde{q}_{n,Y}(\alpha_n\,|\,\cdot)$, $\hat{q}^{\circledast}_{n,Y}(\alpha_n\,|\,\cdot)$ or $\hat{q}^{\oplus}_{n,Y}(\alpha_n\,|\,\cdot)$ computed on the rth replication. Here, both Student-t_ν and Burr distributions are considered with $\nu \in \{1,2,4\}$, $\alpha \in \{1,2,4\}$, $\beta = 1$, $\alpha_n = 1/n$ and $n \in \{20^2, 40^2, 60^2, 80^2, 100^2\}$. The RMSE are reported in Table 2. For all estimators, it appears that the main driver of the relative error is the tail heaviness. Unsuprisingly, the semi-parametric estimator $\tilde{q}_{n,Y}$ provides much better results than the nonparametric ones $\hat{q}^{\circledast}_{n,Y}$ and $\hat{q}^{\oplus}_{n,Y}$: Its RMSE is smaller and converges towards 0 at a faster rate when the sample size n increases.

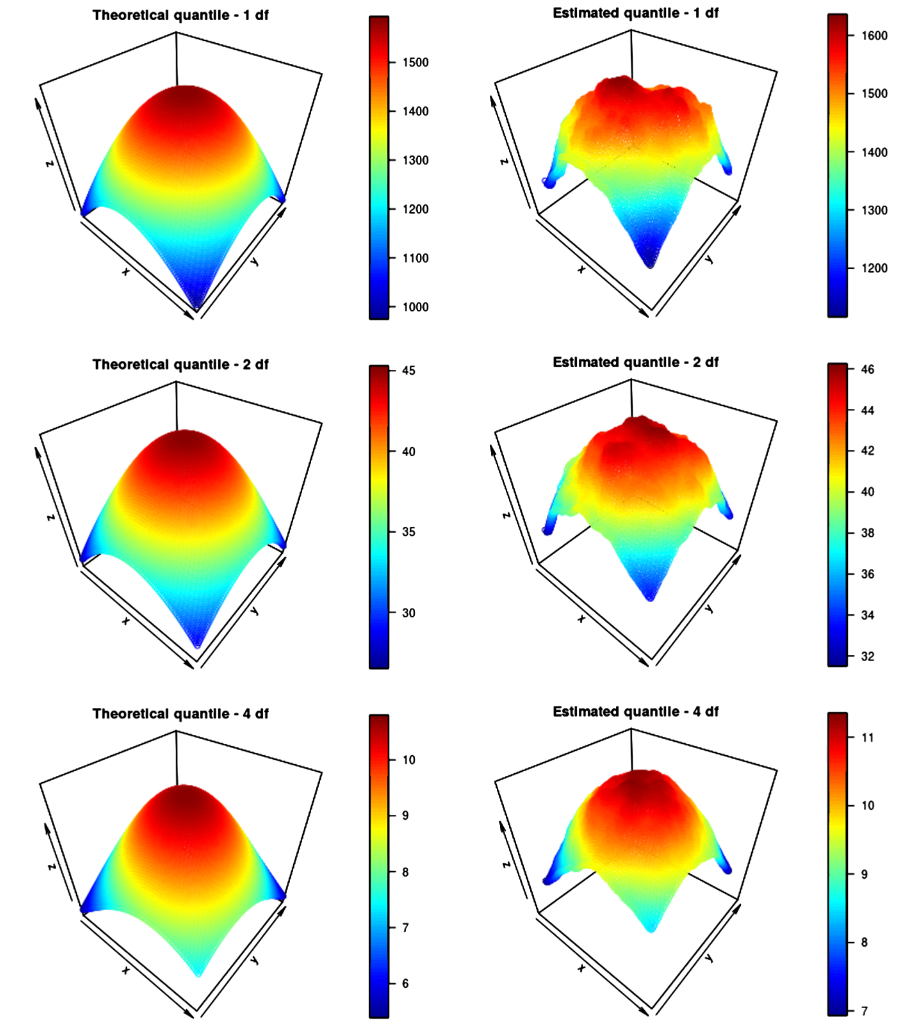

FIG 4. *Simulation results obtained on a Student-t*$_\nu$ *distribution for* $\nu = 1$ *(top),* $\nu = 2$ *(middle) and* $\nu = 4$ *(bottom). Left panels: Theoretical quantiles* $q_Y(1/n\,|\,\cdot)$. *Right panels: Means over* $N = 100$ *replications of estimates* $\tilde{q}_{n,Y}(1/n\,|\,.)$ *computed on* $n = 10,000$ *observations. X-axis and y-axis range between 0 and 1, the scale of the z-axis is the same for theoretical and estimated quantiles.*

6. Tsunami data example

The proposed illustration is based on the “Tsunami Causes and Waves” dataset, available at https://www.kaggle.com/noaa/seismic-waves. The data include the maximum wave height recorded at several stations in the world where a tsunami occured. We focus on the 2011 Tohoku tsunami, in Japan. This earthquake was the cause of the Fukushima Daiichi nuclear disaster. Indeed, a wave height greater than 15 meters (around 50 feet) flooded the nuclear plant, protected by a seawall of only 5.7 meters (19 feet). In this context, the estimation of return levels of wave heights associated with small probability is a crucial issue. Figure 5 (top-left panel) displays the maximum wave heights $Y_1, \dots, Y_n$ (in meters) recorded the 03/11/2011 at $n = 5,364$ stations with respective latitudes $x_1^{(1)}, \dots, x_n^{(1)}$ and longitudes $x_1^{(2)}, \dots, x_n^{(2)}$. Note that the values of Y are ranging from 0 to 55.88 meters (blue to red points). We propose to estimate an extreme quantile of the wave height at each station, following the methodology introduced in Section 3. The assumption of a constant conditional tail-index can be checked thanks to the test statistic $T_{4,n}$ introduced in [12]:

$$T_{4,n} = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{\hat{\gamma}_{p_i}}{\hat{\gamma}_H} - 1 \right)^2 .$$

The idea is to compare the Hill estimate $\hat{\gamma}_H$ computed on the response variables with partial ones $\hat{\gamma}_{p_i}$ computed on non-overlapping blocks indexed by $i = 1, \dots, m$. Under the hypothesis that the conditional tail-index is constant (and additional technical assumptions), it is then shown that $k_n T_{4,n} \xrightarrow{d} \chi^2_{m-1}$, see [12] for details. Following the ideas of Paragraph 5.3, we set $k_n = k_n^{\oplus} = 72$ and we choose $m = 4$ blocks as in [12], leading to $T_{4,n} \approx 2.14$ and a p-value around 0.54. The hypothesis of a constant conditional tail-index cannot be rejected, and our semi-parametric approach can thus be applied on these data.

To this end, a bandwidth has to be selected. Noticing that the standard deviations of $x^{(1)}$ and $x^{(2)}$ are respectively 1.63 and 1.16, we fixed $h_n^* = 1.63 \times n^{-1/6} \simeq 0.4$. We also set $\mu_1 = 3/4$, $\mu_2 = 1/2$ and $\mu_3 = 1/4$, these choices having no consequence in practice. The regression and dispersion functions are then estimated via (3.3) and depicted on the bi-dimensional map (Figure 5, top-right and bottom-left panels) and along the one-dimensional first principal axis (Figure 6, top panels). Note that the principal axis has been obtained by computing the eigenvector associated with the largest eigenvalue of the covariance matrix of the coordinates $(x_i^{(1)}, x_i^{(2)})$, $i = 1, \dots, n$. It appears that $\hat{a}_n(\cdot)$ and $\hat{b}_n(\cdot)$ have a similar shape with a peak in the neighbourhood of the epicenter, indicating a strong heteroscedasticity of the observed phenomenon.

The residuals $\hat{Z}_1, \dots, \hat{Z}_n$ are then computed from (3.4). The common practice is to use a graphical diagnosis to check whether these residuals have a heavy-tailed behavior. Here, a quantile-quantile plot is adopted, see the bottom-right panel of Figure 6. The log-excesses $\log(\hat{Z}_{n-i+1,n}/\hat{Z}_{n-k_n^*+1,n})$ are plotted versus the quantiles $\log(k_n^*/i)$ of the standard exponential distribution, $i = 1, \dots, k_n^*$. Note that the number of upper order statistics $k_n^* = 82$ is chosen following the approach described in Paragraph 5.2. It appears that the resulting set of points is close to the line of slope $\hat{\gamma}_n$ (computed with $k_n^* = 82$), which confirms that the heavy-tailed assumption is reasonable in this case. The proposed estimator (3.5) computed on the residuals as well as the Hill estimator computed on the output variables are both depicted as functions of k_n on the bottom-left panel of Figure 6. The first one features a nice stable behavior, confirming the heavy-tail assumption, and pointing towards a tail-index close to 0.25. As a comparison,

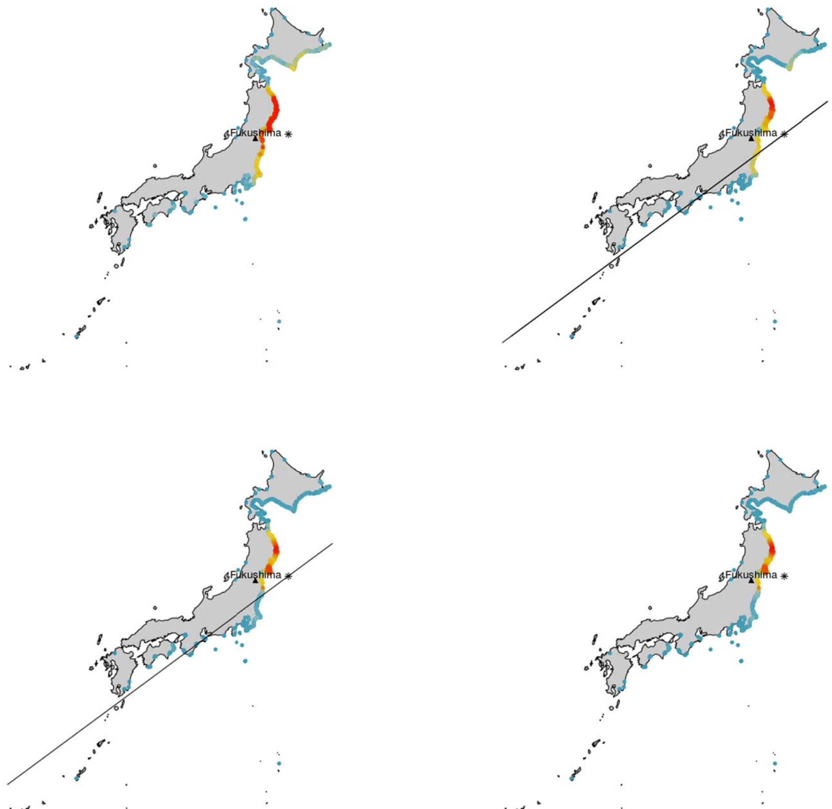

FIG 5. *Results on tsunami data. Top-left: Maximum wave height recorded at each station. Top-right: Regression function estimate* $\hat{a}_n(\cdot)$ *at each station. Bottom-left: Dispersion function estimate* $\hat{b}_n(\cdot)$ *at each station. Bottom-right: Quantile estimate* $\tilde{q}_{n,Y}(10/n \mid \cdot)$ *at each station. On all the maps, smallest and largest values are respectively depicted in blue and red. The straight line is the principal axis* $x^{(2)} = 1.64x^{(1)} + 80.35$ *computed on the coordinates of the stations, and* $*$ *represents the epicenter of the earthquake.*

the Hill estimator computed on the original output variables is less stable and yields smaller results, in accordance with the negative bias observed on simulated data (Section 5). Finally, the extreme conditional quantile estimator (3.6) is evaluated at each station with the level $\alpha_n = 10/n$. The results are reported in the bottom-right panel of Figure 5. The estimated quantiles of the maximum wave height are ranging from 0 to 60.53 meters, with largest values close to the epicenter. Note that such a quantile level means that the observed values $Y_1, \dots, Y_n$ should exceed the return levels $\tilde{q}_{n,Y}(\alpha_n \mid x_1), \dots, \tilde{q}_{n,Y}(\alpha_n \mid x_n)$ approximately 10 times in the sample. In this particular example, there are 15 waves exceeding the return levels, this empirical result does not deviate too much from the expected number of exceedances.

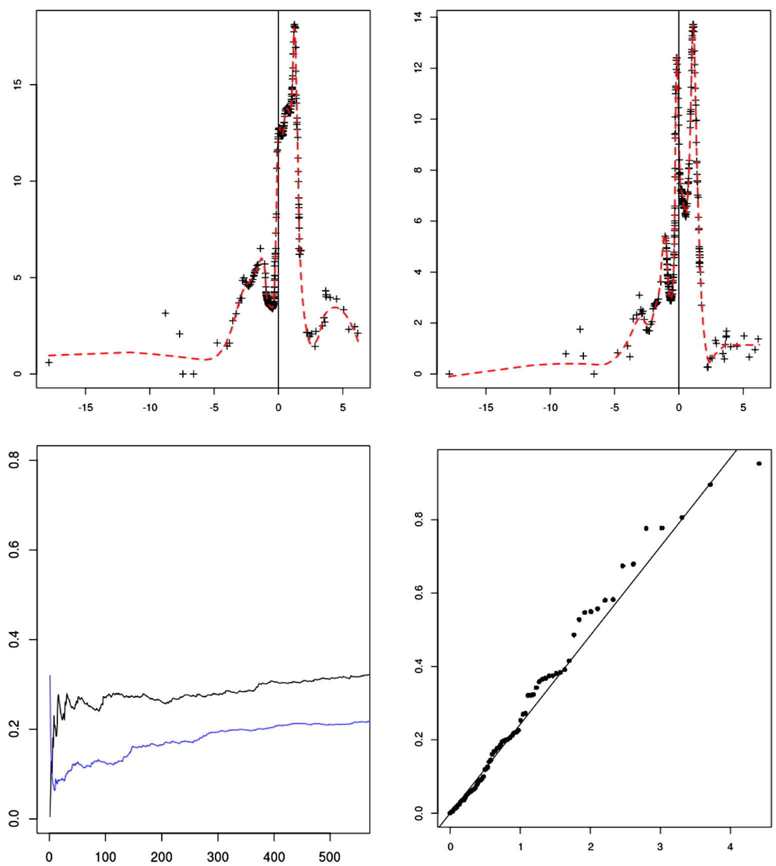

Fig 6. *Results on tsunami data. Top: Regression (left) and dispersion (right) function estimates* $\hat{a}_n(\cdot)$ *and* $\hat{b}_n(\cdot)$ *along the principal axis* $x^{(2)} = 1.64x^{(1)} + 80.35$. *The estimates at each station (black +) are smoothed (red dashed line) for the visualization sake. The vertical black line displays the projection of the epicenter on the principal axis. Bottom left: Hill estimator* (3.5) *computed on the residuals (black line) and on the original output variables (blue line) as a function of* k_n. *Bottom right: Log-excesses* $\log(\hat{Z}_{n-i+1,n}/\hat{Z}_{n-k_n^*+1,n})$ *of the residuals versus* $\log(k_n^*/i)$, $1 \le i \le k_n^* = 82$. *The straight line has slope* $\hat{\gamma}_n \simeq 0.25$.

Appendix A: Proofs

Technical lemmas are collected in Paragraph A.1 while preliminary results of general interest are provided in Paragraph A.2. Finally, the proofs of the main results are given in Paragraph A.3.

A.1. Auxiliary lemmas

The first result is an adaptation of Bochner's lemma (for twice differentiable functions) to the multidimensional fixed design setting.

Lemma A.1. *Let $\psi(\cdot \mid \cdot) : \mathbb{R}^p \times \Pi \to \mathbb{R}^+$ be a positive, twice differentiable (with respect to its second argument) function. Let us denote by $H_2[\psi](\cdot, \cdot)$ the Hessian matrix of $\psi(\cdot \mid \cdot)$ with respect to its second argument, and assume that $H_2[\psi](\cdot, \cdot)$ is continuous on $\mathbb{R}^p \times \Pi$. Let C be a compact subset of $\mathbb{R}^p$. For all sequences $(t_n) \subset \tilde{\Pi}^{(n)}$ and $(y_n) \subset C$, define*

$$\psi_n(y_n \mid t_n) := \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{\Pi_i} Q_h(t_n - s)ds,$$

where $x_i \in \Pi_i$ such that (4.1) and (4.2) hold, and $Q_h(\cdot) = Q(\cdot/h)/h^d$, where Q is an even measurable positive function with symmetric support $S \subset B(0,1)$. Then, letting $\|Q\|_1 = \int_S Q(u)du$, one has, as $n \to \infty$,

$$\psi_n(y_n \mid t_n) = \|Q\|_1 \psi(y_n \mid t_n) + O\left(n^{-1/d}\right) + O(h^2).$$

Proof. Consider the expansion

$$\begin{aligned}
& \psi_n(y_n \mid t_n) - \|Q\|_1 \psi(y_n \mid t_n) \\
= & \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{\Pi_i} Q_h(t_n - s)ds - \|Q\|_1 \psi(y_n \mid t_n) \\
= & \int_{\Pi} \psi(y_n \mid s) Q_h(t_n - s)ds - \|Q\|_1 \psi(y_n \mid t_n) \\
+ & \sum_{i=1}^{n} \psi(y_n \mid x_i) \int_{\Pi_i} Q_h(t_n - s)ds - \int_{\Pi} \psi(y_n \mid s) Q_h(t_n - s)ds \\
=: & \; T_{n,1} + T_{n,2}
\end{aligned}$$

and let us first focus on $T_{n,1}$. The change of variable $u = (t_n - s)/h$ yields

$$T_{n,1} = \int_{(t_n - \Pi)/h} \psi(y_n \mid t_n - uh) Q(u)du - \|Q\|_1 \psi(y_n \mid t_n).$$

Let us remark that $x \in B(0,1)$ implies $t_n - xh \in B(t_n, h) \subset \Pi$ since $t_n \in \tilde{\Pi}^{(n)}$ and by definition of the erosion. As a consequence, $S \subset B(0,1) \subset (t_n - \Pi)/h$ and therefore

$$T_{n,1} = \int_S [\psi(y_n \mid t_n - uh) - \psi(y_n, t_n)] Q(u)du.$$

Let $\nabla_2[\psi](\cdot, \cdot)$ denote the gradient of $\psi(\cdot \mid \cdot)$ with respect to its second argument and let $\langle \cdot, \cdot \rangle$ be the usual dot product on $\mathbb{R}^d$. A second order Taylor expansion yields, for all $y_n \in C$,

$$\psi(y_n \mid t_n - uh) - \psi(y_n \mid t_n) = h \langle \nabla_2[\psi](y_n, t_n), u \rangle + O(h^2),$$

since $H_2[\psi](\cdot, \cdot)$ is bounded on compact sets. Remarking that $\int_S uQ(u)du = 0$ shows that

$$T_{n,1} = O(h^2). \tag{A.1}$$

Let us now turn to the second term

$$T_{n,2} = \sum_{i=1}^{n} \int_{\Pi_i} [\psi(y_n \mid x_i) - \psi(y_n \mid s)] Q_h(t_n - s) ds.$$

Since $\psi(\cdot \mid \cdot)$ is continuously differentiable with respect to its second argument, there exists $c_\psi > 0$ such that

$$\begin{aligned} |T_{n,2}| &\le \sum_{i=1}^{n} \int_{\Pi_i} |\psi(y_n \mid x_i) - \psi(y_n \mid s)| \, Q_h(t_n - s) ds \\ &\le c_\psi \sum_{i=1}^{n} \int_{\Pi_i} \|x_i - s\| Q_h(t_n - s) ds. \end{aligned} \tag{A.2}$$

Moreover, under assumption (4.2),

$$\begin{aligned} |T_{n,2}| &= \sum_{i=1}^{n} \int_{\Pi_i} Q_h(t_n - s) ds \, O\left(n^{-1/d}\right) \\ &= \int_{\Pi} Q_h(t_n - s) ds \, O\left(n^{-1/d}\right) = O\left(n^{-1/d}\right). \end{aligned} \tag{A.3}$$

Finally, collecting (A.1) and (A.3), the conclusion follows. □

As a consequence of Lemma A.1, the asymptotic bias and variance of the estimator (3.1) of the conditional survival function can be derived.

Lemma A.2. *Suppose* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(y_n) \subset C$ *be two nonrandom sequences with* C *a compact subset of* $\mathbb{R}$.

(i) Then,

$$\mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right) = \bar{F}_Y(y_n \mid t_n) + O\left(n^{-1/d}\right) + O(h^2).$$

(ii) If, moreover, $nh^d \to \infty$ *as* $n \to \infty$ *and* $\liminf F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n) > 0$, *then*

$$\operatorname{var}\left(\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right) \sim \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n),$$

where F_Y *is the conditional cumulative distribution function associated with* $\bar{F}_Y$.

Proof. (i) Clearly,

$$\mathbb{E}\left[\hat{\bar{F}}_{n,Y}(y_n \mid t_n)\right] = \sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) \int_{\Pi_i} K_h(t_n - s) ds,$$

and the conclusion follows from Lemma A.1 applied with $p = 1$.

(ii) As a consequence of the independence assumption,

$$\operatorname{var}\left(\hat{\hat{F}}_{n,Y}(y_n \mid t_n)\right) = \sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) S_{n,i} - \sum_{i=1}^{n} \bar{F}_Y^2(y_n \mid x_i) S_{n,i} =: T_{n,1} - T_{n,2},$$

where

$$S_{n,i} := \left(\int_{\Pi_i} K_h(t_n - s) ds\right)^2 = \frac{1}{h^{2d}} \int_{\Pi_i} \int_{\Pi_i} K\left(\frac{t_n - s_1}{h}\right) K\left(\frac{t_n - s_2}{h}\right) ds_1 ds_2. \tag{A.4}$$

Let us write

$$K\left(\frac{t_n - s_2}{h}\right) = K\left(\frac{t_n - s_1}{h}\right) + \left[K\left(\frac{t_n - s_2}{h}\right) - K\left(\frac{t_n - s_1}{h}\right)\right],$$

with, under **(A.3)** and (4.2),

$$\left|K\left(\frac{t_n - s_2}{h}\right) - K\left(\frac{t_n - s_1}{h}\right)\right| \leq \frac{c_K \|s_2 - s_1\|}{h} = O\left(\frac{1}{n^{1/d} h}\right),$$

uniformly on $(s_1, s_2) \in \Pi_i^2$ and $i = 1, \ldots, n$. It thus follows from (4.1) that $S_{n,i}$ can be rewritten as

$$\frac{1}{h^{2d}} \int_{\Pi_i} \int_{\Pi_i} \left[K^2\left(\frac{t_n - s_1}{h}\right) + K\left(\frac{t_n - s_1}{h}\right) O\left(\frac{1}{n^{1/d} h}\right)\right] ds_1 ds_2 \tag{A.5}$$

$$= \frac{\lambda(\Pi)}{nh^{2d}} \int_{\Pi_i} K^2\left(\frac{t_n - s}{h}\right) ds \, (1 + o(1)) + O\left(\frac{1}{n^{1+1/d} h^{2d+1}}\right) \int_{\Pi_i} K\left(\frac{t_n - s}{h}\right) ds$$

$$= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \int_{\Pi_i} M_h(t_n - s) ds \, (1 + o(1)) + O\left(\frac{1}{n^{1+1/d} h^{d+1}}\right) \int_{\Pi_i} K_h(t_n - s) \, ds,$$

where we have defined $M(\cdot) = K^2(\cdot)/\|K^2\|_1 = K^2(\cdot)/\|K\|_2^2$. Replacing in $T_{n,1}$ yields

$$\begin{aligned} T_{n,1} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \left\{\sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) \int_{\Pi_i} M_h(t_n - s) ds \, (1 + o(1))\right. \\ &+ \left. O\left(\frac{1}{n^{1/d} h}\right) \sum_{i=1}^{n} \bar{F}_Y(y_n \mid x_i) \int_{\Pi_i} K_h(t_n - s) ds\right\}. \end{aligned}$$

Applying Lemma A.1 with $p = 1$ twice and recalling that $nh^d \to \infty$ as $n \to \infty$ entail

$$\begin{aligned} T_{n,1} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \left(\bar{F}_Y(y_n \mid t_n) \, (1 + o(1)) + O(h^2) + O\left(\frac{1}{n^{1/d}}\right)\right) \\ &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \bar{F}_Y(y_n \mid t_n) \, (1 + o(1)), \end{aligned}$$

under the assumption $\liminf F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n) > 0$. Similarly,

$$T_{n,2} = \frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\bar{F}_Y^2(y_n \mid t_n)\ (1 + o(1)),$$

and the conclusion follows:

$$T_{n,1} - T_{n,2} = \frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\bar{F}_Y(y_n \mid t_n)F_Y(y_n \mid t_n)\ (1 + o(1)),$$

under the assumption $\liminf F_Y(y_n \mid t_n)\bar{F}_Y(y_n \mid t_n) > 0$. □

Finally, Lemma A.3 is an adaptation of [20, Lemma A.3]. It permits to derive the error made on the estimation of the order statistics Z_{m_n-i,m_n}, $i = 0,\dots,m_n-1$ from the error made on the unsorted Z_i, $i \in I_n$.

Lemma A.3. *Recall that* $I_n = \{i \in \{1,\dots,n\}$ *such that* $x_i \in \tilde{\Pi}^{(n)}\}$ *and* $m_n = \operatorname{card}(I_n)$. *Assume* $nh^d \to \infty$ *as* $n \to \infty$.

(i) Then, $m_n = n(1 + O(h))$.

(ii) Consider (k_n) *an intermediate sequence of integers. If, for all* $i \in I_n$, $|\hat{Z}_i - Z_i| \le R_{n,i}\,(1 + |Z_i|)$, *with* $\max_{i\in I_n} R_{n,i} \xrightarrow{\mathbb{P}} 0$, *then*

$$\max_{0\le i\le k_n}\left|\log\frac{\hat{Z}_{m_n-i,m_n}}{Z_{m_n-i,m_n}}\right| = O_{\mathbb{P}}\left(\max_{i\in I_n} R_{n,i}\right).$$

Proof. (i) Let $C_n = \Pi \setminus \tilde{\Pi}^{(n)}$, $J_n = \{i \in \{1,\dots,n\}$ such that $x_i \in C_n\}$ and $N_n := \operatorname{card}(J_n)$. For all $i \in J_n$, $x_i \in C_n$ and $nh^d \to \infty$ together with (4.2) entail that $\Pi_i \subset C_n$, for n large enough. Therefore, as the sets Π_i are disjoint:

$$\sum_{i\in J_n}\lambda(\Pi_i) \le \lambda(C_n) = \lambda(\Pi) - \lambda(\tilde{\Pi}^{(n)}) = O(h),$$

in view of the absolute continuity of the erosion with respect to Lebesgue measure, see [32]. From (4.1), $\lambda(\Pi_i) \sim \lambda(\Pi)/n$ uniformly on $i = 1,\dots,n$ and thus $N_n = O(nh)$. Therefore, $m_n = n - N_n = n(1 + O(h))$ as $n \to +\infty$.

(ii) The conclusion follows by remarking that in view of (2.2) the distribution of Z has an infinite upper endpoint and by applying [20, Lemma 3]. □

A.2. Preliminary results

Let $\vee$ (resp. $\wedge$) denote the maximum (resp. the minimum). The next proposition provides a joint asymptotic normality result for the estimator (3.1) of the conditional survival function evaluated at points depending on n.

Proposition A.1. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(\alpha_j)_{j=1,\dots,J}$ *a strictly decreasing sequence in* $(0,1)$. *For all* $j \in \{1,\dots,J\}$, *define* $y_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\varepsilon_{j,n}$*, where* $\varepsilon_{j,n} \to 0$ *as* $n \to \infty$*. If* $nh^d \to \infty$ *and* $nh^{d+\kappa(d)} \to 0$ *as* $n \to \infty$*, then*

$$\left\{ \sqrt{nh^d} \left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n) \right] \right\}_{j=1,\dots,J} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^J}, \lambda(\Pi)\|K\|_2^2\, B\right),$$

where $B_{k,l} = \alpha_{k\vee\ell}(1-\alpha_{k\wedge\ell})$ *for all* $(k,\ell) \in \{1,\dots,J\}^2$*.*

Proof. Let us first remark that, for all $j \in \{1,\dots,J\}$, in view of (2.5), the sequence $y_{j,n} = a(t_n) + b(t_n)(q_Z(\alpha_j) + \varepsilon_{j,n})$ is bounded since $a(\cdot)$ and $b(\cdot)$ are continuous functions defined on compact sets and because $\varepsilon_{j,n} \to 0$ as $n \to \infty$. Besides, from (2.3), $F_Y(y_{j,n} \mid t_n) = F_Z(q_Z(\alpha_j) + \varepsilon_{j,n}) \to 1 - \alpha_j > 0$ as $n \to \infty$ and thus the assumptions of Lemma A.2(i,ii) are satisfied. Now, let $\beta \neq 0$ in $\mathbb{R}^J$, $J \geq 1$ and consider the random variable

$$\begin{aligned}
\Gamma_n &= \sum_{j=1}^J \beta_j \left\{ \hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n) \right\} \\
&= \sum_{j=1}^J \beta_j \left\{ \hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right) \right\} \\
&+ \sum_{j=1}^J \beta_j \left\{ \mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) \right) - \bar{F}_Y(y_{j,n} \mid t_n) \right\} \\
&=: \Gamma_{n,1} + \Gamma_{n,2}.
\end{aligned}$$

The random term can be expanded as

$$\Gamma_{n,1} = \sum_{i=1}^n \int_{\Pi_i} K_h(t_n - s)ds \sum_{j=1}^J \beta_j \left\{ \mathbb{1}_{\{Y_i > y_{j,n}\}} - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{j,n}\}} \right) \right\} =: \sum_{i=1}^n T_{i,n}.$$

By definition, $\mathbb{E}(\Gamma_{n,1}) = 0$, and by independence of $Y_1, \dots, Y_n$,

$$\operatorname{var}(\Gamma_{n,1}) = \sum_{i=1}^n \left(\int_{\Pi_i} K_h(t_n - s)ds \right)^2 \operatorname{var}\left(\sum_{j=1}^J \beta_j \mathbb{1}_{\{Y_i > y_{j,n}\}} \right) =: \beta^t C(n) \beta,$$

where $C^{(n)}$ is the matrix whose coefficients are defined for all $(k,\ell) \in \{1,\dots,J\}^2$ by

$$C_{k,\ell}^{(n)} = \sum_{i=1}^n S_{n,i} \operatorname{cov}\left(\mathbb{1}_{\{Y_i > y_{k,n}\}}, \mathbb{1}_{\{Y_i > y_{\ell,n}\}} \right), \tag{A.6}$$

with $S_{n,i}$ being defined in (A.4) and expanded as (A.5):

$$S_{n,i} = \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \int_{\Pi_i} M_h(t_n - s)ds(1 + o(1)) + O\left(\frac{1}{n^{1+1/d} h^{d+1}} \right) \int_{\Pi_i} K_h(t_n - s)ds,$$

see the proof of Lemma A.2. Straightforward calculations yield

$$\begin{aligned}
& \operatorname{cov}\left(\mathbb{1}_{\{Y_i>y_{k,n}\}}, \mathbb{1}_{\{Y_i>y_{\ell,n}\}}\right) \\
= \;& \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i) - \bar{F}_Y(y_{k,n} \mid x_i)\bar{F}_Y(y_{\ell,n} \mid x_i) \\
= \;& \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i) - \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i)\bar{F}_Y(y_{k,n} \wedge y_{\ell,n} \mid x_i) \\
= \;& \bar{F}_Y(y_{k,n} \vee y_{\ell,n} \mid x_i)F_Y(y_{k,n} \wedge y_{\ell,n} \mid x_i) \\
=: \;& \varphi(y_{k,n}, y_{\ell,n} \mid x_i), \qquad\qquad \text{(A.7)}
\end{aligned}$$

where $\varphi : \mathbb{R}^2 \times \Pi \to [0,1]$ is defined by $\varphi(\cdot,\cdot \mid .) = \bar{F}_Y(\cdot \vee \cdot \mid \cdot)F_Y(\cdot \wedge \cdot \mid \cdot)$. Replacing in (A.6) yields

$$\begin{aligned}
C_{k,\ell}^{(n)} &= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \sum_{i=1}^{n} \varphi(y_{k,n}, y_{\ell,n} \mid x_i) \int_{\Pi_i} M_h(t_n - s)ds \,(1 + o(1)) \\
&+ O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right) \sum_{i=1}^{n} \varphi(y_{k,n}, y_{\ell,n} \mid x_i) \int_{\Pi_i} K_h(t_n - s)ds \\
&= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} \left[\varphi(y_{k,n}, y_{\ell,n} \mid t_n) + O(h^2) + O(n^{-1/d})\right] (1 + o(1)) \\
&+ O\left(\frac{1}{n^{1+1/d}h^{d+1}}\right) \left[\varphi(y_{k,n}, y_{\ell,n} \mid t_n) + O(h^2) + O(n^{-1/d})\right] \\
&= \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} [\varphi(y_{k,n}, y_{\ell,n} \mid t_n)(1 + o(1)) + O(h^2) + O(n^{-1/d})], \qquad \text{(A.8)}
\end{aligned}$$

from Lemma A.1 applied twice with $p = 2$ and recalling that $nh^d \to \infty$. Besides, let us remark that, in view of (2.5),

$$\begin{aligned}
y_{k,n} - y_{\ell,n} &= b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_\ell) + \varepsilon_{k,n} - \varepsilon_{\ell,n}) \\
&= b(t_n)(q_Z(\alpha_k) - q_Z(\alpha_\ell))(1 + o(1)),
\end{aligned}$$

as $n \to \infty$. Therefore, assuming for instance $k < \ell$ implies $\alpha_k > \alpha_\ell$ and thus $q_Z(\alpha_k) < q_Z(\alpha_\ell)$ leading to $y_{k,n} < y_{\ell,n}$ for n large enough. More generally, $y_{k,n} \vee y_{\ell,n} = y_{k\vee\ell,n}$ and $y_{k,n} \wedge y_{\ell,n} = y_{k\wedge\ell,n}$ for n large enough and thus $\varphi(y_{k,n}, y_{\ell,n} \mid t_n) = \bar{F}_Y(y_{k\vee\ell,n} \mid t_n)F_Y(y_{k\wedge\ell,n} \mid t_n)$. From (2.3) and (2.5), we have

$$\bar{F}_Y(y_{k,n} \mid t_n) = \bar{F}_Z\left(\frac{y_{k,n} - a(t_n)}{b(t_n)}\right) = \bar{F}_Z\left(q_Z(\alpha_k) + \varepsilon_{k,n}\right) = \alpha_k + o(1),$$

in view of the continuity of $\bar{F}_Z$. As a result,

$$\varphi(y_{k,n}, y_{\ell,n} \mid t_n) \to B_{k,\ell} = \alpha_{k\vee\ell}(1 - \alpha_{k\wedge\ell}) \text{ as } n \to \infty. \qquad \text{(A.9)}$$

Collecting (A.8) and (A.9), one has

$$C_{k,\ell}^{(n)} = \frac{\lambda(\Pi)\|K\|_2^2}{nh^d} B_{k,\ell}(1 + o(1))$$

and therefore

$$\operatorname{var}(\Gamma_{n,1}) \sim \frac{\lambda(\Pi)\|K\|_2^2}{nh^d}\beta^t B\beta, \tag{A.10}$$

where B is the matrix defined by the $B_{k,\ell}$ coefficients. The proof of the asymptotic normality of $\Gamma_{n,1}$ is based on Lyapounov criteria for triangular arrays of independent random variables:

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \Big/ (\operatorname{var}(\Gamma_{n,1}))^{3/2} \to 0 \tag{A.11}$$

as $n \to \infty$. Let us highlight that the random variables $T_{i,n}$, $i = 1, \ldots, n$, are bounded:

$$\begin{aligned} |T_{i,n}| &\leq \int_{\Pi_i} K_h(t_n - s)ds \sum_{j=1}^{J} \beta_j \left|\mathbb{1}_{\{Y_i > y_{j,n}\}} - \mathbb{E}\left(\mathbb{1}_{\{Y_i > y_{j,n}\}}\right)\right| \\ &\leq \frac{\lambda(\Pi)\|K\|_\infty}{nh^d} \sum_{j=1}^{J} |\beta_j| \, (1 + o(1)) =: \zeta_n \end{aligned} \tag{A.12}$$

in view of **(A.3)** and (4.1). As a consequence, one has

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \leq \zeta_n \sum_{i=1}^{n} \mathbb{E}(T_{i,n}^2) = \zeta_n \sum_{i=1}^{n} \operatorname{var}(T_{i,n}) = \zeta_n \operatorname{var}(\Gamma_{n,1}),$$

leading to

$$\sum_{i=1}^{n} \mathbb{E}|T_{i,n}|^3 \Big/ (\operatorname{var}(\Gamma_{n,1}))^{3/2} = O\left((nh^d)^{-1/2}\right),$$

from (A.10) and (A.12). It is thus clear that (A.11) holds under the assumption $nh^d \to \infty$ and

$$\sqrt{nh^d}\Gamma_{n,1} \xrightarrow{d} \mathcal{N}\left(0, \lambda(\Pi)\|K\|_2^2 \, \beta^t B\beta\right). \tag{A.13}$$

Let us now turn to the nonrandom term. Lemma A.2(i) together with the assumptions $nh^d \to \infty$ and $nh^{d+\kappa(d)} \to 0$ as $n \to \infty$ entail

$$\begin{aligned} \sqrt{nh^d}|\Gamma_{n,2}| &\leq \sqrt{nh^d} \sum_{j=1}^{J} |\beta_j| \left|\mathbb{E}\left(\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n)\right) - \bar{F}_Y(y_{j,n} \mid t_n)\right| \\ &= O(\sqrt{nh^{d+\kappa(d)}}) = o(1). \end{aligned} \tag{A.14}$$

Finally, collecting (A.13) and (A.14), $\sqrt{nh^d}\Gamma_n$ converges to a centered Gaussian random variable with variance $\lambda(\Pi)\|K\|_2^2 \, \beta^t B\beta$, and the result follows. □

The following proposition provides the joint asymptotic normality of the estimator (3.2) of conditional quantiles. It can be read as an adaptation of classical results [2, 35, 38] to the location-dispersion regression model in the multivariate fixed design setting.

Proposition A.2. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Let* $(t_n) \subset \tilde{\Pi}^{(n)}$ *and* $(\alpha_j)_{j=1,\dots,J}$ *a strictly decreasing sequence in* $(0,1)$ *such that* $f_Z(q_Z(\alpha_j)) > 0$ *for all* $j \in \{1,\dots,J\}$. *If* $nh^d \to \infty$ *and* $nh^{d+\kappa(d)} \to 0$ *as* $n \to \infty$, *then*

$$\left\{ \frac{\sqrt{nh^d}}{b(t_n)} \left[\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n) \right] \right\}_{j=1,\dots,J} \xrightarrow{d} \mathcal{N}\left(0_{\mathbb{R}^J}, \lambda(\Pi)\|K\|_2^2\, C\right),$$

where C *is the covariance matrix defined for all* $(k,\ell) \in \{1,\dots,J\}^2$ *by* $C_{k,\ell} = \alpha_{k\vee\ell}(1-\alpha_{k\wedge\ell})H_Z(\alpha_k)H_Z(\alpha_\ell)$.

Proof. Let $s = (s_1,\dots,s_J) \in \mathbb{R}^J$, and for all $j = 1,\dots,J$,

$$\begin{aligned}
\varepsilon_{j,n} &:= s_j/\sqrt{nh^d},\\
\nu_{j,n} &:= b(t_n)\varepsilon_{j,n},\\
y_{j,n} &= q_Y(\alpha_j \mid t_n) + \nu_{j,n},\\
V_{j,n} &:= \sqrt{nh^d}\left[\hat{\bar{F}}_{n,Y}(y_{j,n} \mid t_n) - \bar{F}_Y(y_{j,n} \mid t_n)\right],\\
v_{j,n} &:= \sqrt{nh^d}\left[\alpha_j - \bar{F}_Y(y_{j,n} \mid t_n)\right].
\end{aligned}$$

These notations yield

$$\begin{aligned}
W_n(s) &:= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\frac{\sqrt{nh^d}}{b(t_n)}\left(\hat{q}_{n,Y}(\alpha_j \mid t_n) - q_Y(\alpha_j \mid t_n)\right) \le s_j\right\}\right)\\
&= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{V_{j,n} \le v_{j,n}\right\}\right).
\end{aligned}$$

From (2.3) and (2.5), the nonrandom term can be rewritten as

$$v_{j,n} = \sqrt{nh^d}\left(\alpha_j - \bar{F}_Z\left(\frac{y_{j,n} - a(t_n)}{b(t_n)}\right)\right) = \sqrt{nh^d}(\alpha_j - \bar{F}_Z\left(q_Z(\alpha_j) + \varepsilon_{j,n}\right)).$$

Since $\bar{F}_Z(\cdot)$ is differentiable, for all $j \in \{1,\dots,J\}$, there exists $\theta_{j,n} \in (0,1)$ such that

$$v_{j,n} = s_j f_Z\left(q_Z(\alpha_j) + \theta_{j,n}\varepsilon_{j,n}\right) = \frac{s_j}{H_Z(\alpha_j)}(1 + o(1)), \tag{A.15}$$

in view of the continuity of $f_Z(\cdot)$ and since $\varepsilon_{j,n} \to 0$ as $n \to \infty$. Let us now turn to the random term. Recalling that, for all $j = 1,\dots,J$, $y_{j,n} = q_Y(\alpha_j \mid t_n) + b(t_n)\varepsilon_{j,n}$, with $\varepsilon_{j,n} \to 0$ as $n \to \infty$, Proposition A.1 entails that $\{V_{j,n}\}_{j=1,\dots,J}$ converges to a centered Gaussian random vector with covariance matrix $\lambda(\Pi)\|K\|_2^2\, B$. Taking account of (A.15) yields that $W_n(s)$ converges to the cumulative distribution function of a centered Gaussian distribution with covariance matrix $\lambda(\Pi)\|K\|_2^2\, C$, evaluated at s, which is the desired result. □

The following proposition provides a uniform consistency result for the estimator (3.2) of conditional quantiles of Y given a sequence of multidimensional design points in $\tilde{\Pi}^{(n)}$, *i.e.* not too close from the boundary of Π.

Proposition A.3. *Assume* **(A.1)**, **(A.2)** *and* **(A.3)** *hold. Suppose* $nh^d/\log n \to \infty$ *and* $nh^{d+\kappa(d)}/\log n \to 0$ *as* $n \to \infty$. *Then, for all* $\alpha \in (0,1)$,

$$\sqrt{\frac{nh^d}{\log n}} \max_{i \in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1).$$

Proof. Let $v_n = (nh^d/\log n)^{1/2}$ and for all $(\varepsilon, \alpha) \in (0,1)^2$, consider

$$\begin{aligned} \kappa_1(\varepsilon,\alpha) &= 2\|K\|_2 \left(\lambda(\Pi)\alpha(1-\alpha)\left(1-\log(\varepsilon/2)\right)\right)^{1/2}, \\ \kappa_2(\alpha) &= \lambda(\Pi)\alpha(1-\alpha)\|K\|_2^2, \\ M(\varepsilon,\alpha) &= \kappa_1(\varepsilon,\alpha) H_Z(\alpha). \end{aligned}$$

Let us also introduce, for all $i \in I_n$,

$$\begin{aligned} q_{i,n}^{\pm} &= q_Y(\alpha \mid x_i) \pm M(\varepsilon,\alpha) b(x_i)/v_n, \\ \alpha_{i,n}^{\pm} &= \alpha - \mathbb{E}\left(\hat{\bar{F}}_{n,Y}\left(q_{i,n}^{\pm} \mid x_i\right)\right), \\ \xi_{i,n}^{\pm} &= \left(\hat{\bar{F}}_{n,Y} - \mathbb{E}\hat{\bar{F}}_{n,Y}\right)\left(q_{i,n}^{\pm} \mid x_i\right), \end{aligned}$$

so that the following expansion holds:

$$\begin{aligned} \delta_n &:= \mathbb{P}\left(v_n \max_{i \in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| \geq M(\varepsilon,\alpha) \right) \\ &= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{ \hat{q}_{n,Y}(\alpha \mid x_i) \geq q_{i,n}^+ \right\} \cup \left\{ \hat{q}_{n,Y}(\alpha \mid x_i) \leq q_{i,n}^- \right\} \right) \\ &= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{ \alpha \leq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^+ \mid x_i\right) \right\} \cup \left\{ \alpha \geq \hat{\bar{F}}_{n,Y}\left(q_{i,n}^- \mid x_i\right) \right\} \right) \\ &= \mathbb{P}\left(\bigcup_{i \in I_n} \left\{ \alpha_{i,n}^+ \leq \xi_{i,n}^+ \right\} \right) + \mathbb{P}\left(\bigcup_{i \in I_n} \left\{ \alpha_{i,n}^- \geq \xi_{i,n}^- \right\} \right) \\ &=: \delta_n^+ + \delta_n^-. \end{aligned}$$

Let us focus on the first term. Assumption $nh^d/\log n \to \infty$ entails that $v_n \to \infty$ as $n \to \infty$ and thus $q_{i,n}^+$ is bounded. Therefore Lemma A.2(i) shows that

$$\begin{aligned} \alpha_{i,n}^+ &= \alpha - \bar{F}_Y\left(q_{i,n}^+ \mid x_i\right) + O(h^2) + O(n^{-1/d}) \\ &= \bar{F}_Z(q_Z(\alpha)) - \bar{F}_Z\left(q_Z(\alpha) + \frac{M(\varepsilon,\alpha)}{v_n}\right) + O(h^2) + O(n^{-1/d}) \\ &= \frac{M(\varepsilon,\alpha)}{v_n} f_Z\left(q_Z(\alpha) + \frac{M(\varepsilon,\alpha)}{v_n}\theta\right) + O(h^2) + O(n^{-1/d}), \end{aligned}$$

for some $\theta \in (0,1)$, and the continuity of $f_Z(\cdot)$ then yields

$$\alpha_{i,n}^+ = \frac{M(\varepsilon,\alpha)}{v_n H_Z(\alpha)}\left(1+o(1)\right) + O(h^2) + O(n^{-1/d}) = \frac{\kappa_1(\varepsilon,\alpha)}{v_n}\left(1+o(1)\right), \quad \text{(A.16)}$$

in view of the assumption $nh^{d+\kappa(d)}/\log n \to 0$ as $n \to \infty$. As a consequence,

$$\begin{aligned} \delta_n^+ &= \mathbb{P}\left(\bigcup_{i\in I_n}\left\{\xi_{i,n}^+ \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right\}\right) \\ &\leq \sum_{i\in I_n}\mathbb{P}\left(\xi_{i,n}^+ \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right). \end{aligned} \tag{A.17}$$

Moreover,

$$\mathbb{P}\left(\xi_{i,n}^+ \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right) =: \mathbb{P}\left(\sum_{j=1}^n \tilde{X}_j \geq \frac{\kappa_1(\varepsilon,\alpha)}{v_n}(1+o(1))\right), \tag{A.18}$$

where, for all $j = 1,\dots,n$, the random variables

$$\tilde{X}_j := \left[\mathbb{1}_{\{Y_j > q_{i,n}^+\}} - \mathbb{P}\left(Y_j > q_{i,n}^+ \mid x_i\right)\right]\int_{\Pi_j} K_h(x_i - s)ds$$

are independent, centered and bounded from (4.1):

$$|\tilde{X}_j| \leq \int_{\Pi_j} K_h(x_i - s)ds \leq \frac{\lambda(\Pi)\|K\|_\infty}{nh^d}(1+o(1)).$$

Lemma A.2(ii) entails

$$\begin{aligned} \sum_{j=1}^n \mathbb{E}(\tilde{X}_j^2) &= \operatorname{var}\left(\sum_{j=1}^n \tilde{X}_j\right) = \operatorname{var}\left[\hat{\bar{F}}_{n,Y}\left(q_{i,n}^+ \mid x_i\right)\right] \\ &= \frac{\lambda(\Pi)\bar{F}_Y\left(q_{i,n}^+ \mid x_i\right)F_Y\left(q_{i,n}^+ \mid x_i\right)}{nh^d}\|K\|_2^2(1+o(1)), \\ &= \frac{\kappa_2(\alpha)}{nh^d}(1+o(1)), \end{aligned}$$

since $\alpha_{i,n}^+ \to 0$ as $n \to \infty$ from (A.16) and thus $\bar{F}_Y\left(q_{i,n}^+ \mid x_i\right) \to \alpha$ as $n \to \infty$ in view of the continuity of $\bar{F}_Y(\cdot \mid x_i)$. Bernstein's inequality for bounded random variables yields

$$\begin{aligned} (\text{A.18}) &\leq \exp\left(-\frac{\kappa_1^2(\varepsilon,\alpha)\log n}{2\kappa_2(\alpha) + \frac{2\kappa_1(\varepsilon,\alpha)(1+o(1))}{3v_n}}(1+o(1))\right) \\ &= \exp\left(-\frac{\kappa_1^2(\varepsilon,\alpha)\log n}{2\kappa_2(\alpha)}(1+o(1))\right) \\ &= \exp\left[-2\left(1-\log(\varepsilon/2)\right)\log n\left(1+o(1)\right)\right] \\ &\leq \exp\left[-\left(1-\log(\varepsilon/2)\right)\log n\right], \end{aligned} \tag{A.19}$$

for n large enough. Collecting (A.17)-(A.19) leads to

$$\delta_n^+ \leq n\exp\left[-\left(1-\log(\varepsilon/2)\right)\log n\right] = \exp\left(\log(\varepsilon/2)\log n\right) \leq \varepsilon/2$$

for n large enough. The proof that $\delta_n^- \leq \varepsilon/2$ follows the same lines. As a conclusion, we have shown that, for all $\alpha \in (0,1)$ and $\varepsilon \in (0,1)$ there exists $M(\varepsilon,\alpha) > 0$ such that

$$\mathbb{P}\left(\sqrt{\frac{nh^d}{\log n}} \max_{i\in I_n} \left| \frac{\hat{q}_{n,Y}(\alpha \mid x_i) - q_Y(\alpha \mid x_i)}{b(x_i)} \right| \geq M(\varepsilon,\alpha)\right) \leq \varepsilon,$$

which is the desired result. □

A.3. Proofs of main results

The proof of Theorem 4.1 directly relies on Proposition A.2:

Proof of Theorem 4.1. Let us remark that

$$\frac{\sqrt{nh^d}}{b(t_n)} \begin{pmatrix} \hat{a}_n(t_n) - a(t_n) \\ \hat{b}_n(t_n) - b(t_n) \end{pmatrix} = \Omega \xi_n,$$

where $\Omega = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$ and $\xi_n = \frac{\sqrt{nh^d}}{b(t_n)} \begin{pmatrix} \hat{q}_{n,Y}(\mu_3 \mid t_n) - q_Y(\mu_3 \mid t_n) \\ \hat{q}_{n,Y}(\mu_2 \mid t_n) - q_Y(\mu_2 \mid t_n) \\ \hat{q}_{n,Y}(\mu_1 \mid t_n) - q_Y(\mu_1 \mid t_n) \end{pmatrix}$.

Proposition A.2 with $J = 3$ and $\alpha_j = \mu_j$, $j = 1,\dots,J$ yields that ξ_n converges in distribution to the $\mathcal{N}\left(0_{\mathbb{R}^3},\ \lambda(\Pi)\|K\|_2^2\, C\right)$ distribution where C is given by

$$\begin{pmatrix} \mu_1\bar{\mu}_1 H_Z^2(\mu_1) & \mu_2\bar{\mu}_1 H_Z(\mu_2)(H_Z(\mu_1) & \mu_3\bar{\mu}_1 H_Z(\mu_3) H_Z(\mu_1) \\ \mu_2\bar{\mu}_1 H_Z(\mu_2) H_Z(\mu_1) & \mu_2\bar{\mu}_2 H_Z^2(\mu_2) & \mu_3\bar{\mu}_2 H_Z(\mu_2) H_Z(\mu_3) \\ \mu_3\bar{\mu}_1 H_Z(\mu_3) H_Z(\mu_1) & \mu_3\bar{\mu}_2 H_Z(\mu_2) H_Z(\mu_3) & \mu_3\bar{\mu}_3 H_Z^2(\mu_3) \end{pmatrix},$$

with $\bar{\mu}_j = 1 - \mu_j$, $j = 1,2,3$. Therefore, $\Omega\xi_n \xrightarrow{d} \mathcal{N}(0_{\mathbb{R}^2}, \lambda(\Pi)\|K\|_2^2\, \Omega C \Omega^t)$ and the conclusion follows from $\Omega C \Omega^t = \Sigma$. □

Theorem 4.2 is a straightforward consequence of Proposition A.3:

Proof of Theorem 4.2. Remarking that

$$\max_{i\in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| = \max_{i\in I_n} \left| \frac{\hat{q}_{n,Y}(\mu_2 \mid x_i) - q_Y(\mu_2 \mid x_i)}{b(x_i)} \right|,$$

the first part of the result is a consequence of Proposition A.3 applied with $\alpha = \mu_2$. Similarly,

$$\begin{aligned} \max_{i\in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| &\leq \max_{i\in I_n} \left| \frac{\hat{q}_{n,Y}(\mu_3 \mid x_i) - q_Y(\mu_3 \mid x_i)}{b(x_i)} \right| \\ &+ \max_{i\in I_n} \left| \frac{\hat{q}_{n,Y}(\mu_1 \mid x_i) - q_Y(\mu_1 \mid x_i)}{b(x_i)} \right|, \end{aligned}$$

and the conclusion follows from Proposition A.3 with $\alpha \in \{\mu_3, \mu_1\}$. □

Proof of Corollary 4.1. Remark that for all $i \in I_n$, one has

$$\begin{aligned}
|\hat{Z}_i - Z_i| &= \left|\frac{Y_i - \hat{a}_n(x_i)}{\hat{b}_n(x_i)} - Z_i\right| = \left|\frac{a(x_i) - \hat{a}_n(x_i)}{\hat{b}_n(x_i)} + \frac{\hat{b}_n(x_i) - b(x_i)}{\hat{b}_n(x_i)} Z_i\right| \\
&\leq \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \left(\left|\frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)}\right| + \left|\frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)}\right| |Z_i|\right) \\
&\leq \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \max\left\{\left|\frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)}\right|; \left|\frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)}\right|\right\} (1 + |Z_i|) \\
&=: \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \max\left\{\left|\xi_{i,n}^{(a)}\right|; \left|\xi_{i,n}^{(b)}\right|\right\} (1 + |Z_i|).
\end{aligned}$$

Let us define, for all $i \in I_n$,

$$\begin{aligned}
\xi_{i,n}^{(a)} &= \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)}, \ \xi_{i,n}^{(b)} = \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \text{ and} \\
R_{n,i} &= \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \max\left\{\left|\xi_{i,n}^{(a)}\right|; \left|\xi_{i,n}^{(b)}\right|\right\}.
\end{aligned}$$

On the one hand, Theorem 4.2 entails

$$\begin{aligned}
\max_{i \in I_n} R_{n,i} &\leq \max_{i \in I_n} \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \max\left\{\max_{i \in I_n}\left|\xi_{i,n}^{(a)}\right|; \max_{i \in I_n}\left|\xi_{i,n}^{(b)}\right|\right\} \\
&= \max_{i \in I_n} \left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{P}\left(\max_{i \in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \geq 2\right) &= \mathbb{P}\left(\max_{i \in I_n}\left|\frac{1}{1 + \xi_{i,n}^{(b)}}\right| \geq 2\right) \leq \mathbb{P}\left(\max_{i \in I_n}\left|\xi_{i,n}^{(b)}\right| \geq \frac{1}{2}\right) \\
&\leq \mathbb{P}\left(\sqrt{\frac{nh^d}{\log n}} \max_{i \in I_n}\left|\xi_{i,n}^{(b)}\right| \geq \frac{1}{2}\sqrt{\frac{nh^d}{\log n}}\right).
\end{aligned}$$

Again, Theorem 4.2 shows that the following uniform consistency holds: For all $\varepsilon > 0$, there exists $M(\varepsilon) > 0$ such that

$$\mathbb{P}\left(\sqrt{\frac{nh^d}{\log n}} \max_{i \in I_n}\left|\xi_{i,n}^{(b)}\right| \geq M(\varepsilon)\right) \leq \varepsilon.$$

Now, for n large enough, $(nh^d/\log n)^{1/2} > 2M(\varepsilon)$ so that

$$\mathbb{P}\left(\max_{i \in I_n}\left|\frac{b(x_i)}{\hat{b}_n(x_i)}\right| \geq 2\right) \leq \mathbb{P}\left(\max_{i \in I_n}\sqrt{\frac{nh^d}{\log n}}\left|\xi_{i,n}^{(b)}\right| \geq M(\varepsilon)\right) \leq \varepsilon,$$

i.e. $\max_{i\in I_n} |b(x_i)/\hat{b}_n(x_i)| = O_{\mathbb{P}}(1)$. As a result,

$$\max_{i\in I_n} R_{n,i} = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right),$$

which completes the proof of the corollary. □

Proof of Theorem 4.3. (i) Let us consider the expansion

$$\sqrt{k_n}(\hat{\gamma}_n - \gamma) = \sqrt{k_n}(\hat{\gamma}_n - \tilde{\gamma}_n) + \sqrt{k_n}(\tilde{\gamma}_n - \gamma) =: \Upsilon_{1,n} + \Upsilon_{2,n},$$

where

$$\tilde{\gamma}_n = \frac{1}{k_n}\sum_{i=0}^{k_n-1} \log Z_{m_n-i,m_n} - \log Z_{m_n-k_n,m_n}$$

is the Hill estimator computed on the unobserved random variables $Z_1, \dots, Z_n$. Recall that $m_n = \mathrm{card}(I_n)$ where $I_n = \{i \in \{1,\dots,n\}$ such that $x_i \in \tilde{\Pi}^{(n)}\}$. The first term is controlled by remarking that

$$\begin{aligned} |\Upsilon_{1,n}| &= \sqrt{k_n}|\hat{\gamma}_n - \tilde{\gamma}_n| \le \sqrt{k_n} \max_{0\le i\le k_n} \left|\log \frac{\hat{Z}_{m_n-i,m_n}}{Z_{m_n-i,m_n}}\right| \\ &= O_{\mathbb{P}}\left(\sqrt{\frac{k_n \log n}{nh^d}}\right) = o_{\mathbb{P}}(1), \end{aligned} \tag{A.20}$$

from Corollary 4.1 and Lemma A.3(ii). Let us now focus on $\Upsilon_{2,n}$. Remarking that $m_n \sim n$ as $n \to \infty$ in view of Lemma A.3(i), it is clear that $m_n/k_n \to \infty$ as $n \to \infty$. Besides, since $|A| \in \mathcal{RV}_\rho$, we thus have $A(m_n/k_n) \sim A(n/k_n)$ as $n \to \infty$. Therefore, $\sqrt{k_n}A(m_n/k_n) \to \beta$ as $n \to \infty$ and, since $Z_1, \dots, Z_n$ are iid from (2.2), classical results on Hill estimator apply, see for instance [24, Theorem 3.2.5], leading to

$$\Upsilon_{2,n} \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2). \tag{A.21}$$

The conclusion follows from (A.20) and (A.21).

(ii) Let us introduce $v_n = \sqrt{k_n}/\log(k_n/(n\alpha_n))$ and consider the Weissman estimator computed on the unobserved random variables $Z_1, \dots, Z_n$:

$$\tilde{q}_{n,Z}(\alpha_n) = Z_{m_n-k,m_n}\left(\frac{\alpha_n m_n}{k_n}\right)^{-\tilde{\gamma}_n}.$$

The following expansion holds:

$$\begin{aligned} v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)) &= v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log \tilde{q}_{n,Z}(\alpha_n)) \\ &+ v_n(\log \tilde{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n)) \\ &=: T_{1,n} + T_{2,n}, \end{aligned}$$

with

$$|T_{1,n}| \leq v_n \left|\log \frac{\hat{Z}_{m_n-k_n,m_n}}{Z_{m_n-k_n,m_n}}\right| + v_n|\hat{\gamma}_n - \tilde{\gamma}_n| \left|\log\left(\frac{\alpha_n m_n}{k_n}\right)\right| =: T_{1,1,n} + T_{1,2,n}.$$

First, $T_{1,1,n}$ is controlled by Corollary 4.1 and Lemma A.3(ii) together with the assumptions $k_n \log n/(nh^d) \to 0$ and $k_n/(n\alpha_n) \to \infty$ as $n \to \infty$,

$$T_{1,1,n} = \frac{\sqrt{k_n}}{\log\left(\frac{k_n}{n\alpha_n}\right)} O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^d}}\right) = \sqrt{\frac{k_n \log n}{nh^d}} O_{\mathbb{P}}\left(\frac{1}{\log\left(\frac{k_n}{n\alpha_n}\right)}\right) = o_{\mathbb{P}}(1). \tag{A.22}$$

Second, since $m_n \sim n$ as $n \to \infty$ (see Lemma A.3(i)),

$$T_{1,2,n} = |\Upsilon_{1,n}|(1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1), \tag{A.23}$$

in view of (A.20). Collecting (A.22) and (A.23) yields

$$T_{1,n} = v_n(\log \hat{q}_{n,Z}(\alpha_n) - \log \tilde{q}_{n,Z}(\alpha_n)) = o_{\mathbb{P}}(1). \tag{A.24}$$

Let us now focus on $T_{2,n}$. As a consequence of [24, Theorem 4.3.8], Weissman estimator inherits its asymptotic distribution from Hill estimator:

$$v_n\left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2),$$

in view of (A.21). As a result,

$$T_{2,n} \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2). \tag{A.25}$$

The conclusion follows from (A.24) and (A.25). □

Proof of Theorem 4.4. Let $v_n = \sqrt{k_n}/\log(k_n/(n\alpha_n))$ and consider the following expansion:

$$\begin{aligned}
&\frac{v_n}{b(t_n)q_Z(\alpha_n)}\left(\tilde{q}_{n,Y}(\alpha_n \mid t_n) - q_Y(\alpha_n \mid t_n)\right)\\
=&\frac{v_n}{q_Z(\alpha_n)}\left(\frac{\hat{a}_n(t_n) - a(t_n)}{b(t_n)}\right) + v_n\left(\frac{\hat{b}_n(t_n) - b(t_n)}{b(t_n)}\right) + v_n\frac{\hat{b}_n(t_n)}{b(t_n)}\left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1\right)\\
=:&\frac{\sqrt{\frac{k_n}{nh^d}}\xi_n^{(a)}}{q_Z(\alpha_n)\log\left(\frac{k_n}{n\alpha_n}\right)} + \frac{\sqrt{\frac{k_n}{nh^d}}\xi_n^{(b)}}{\log\left(\frac{k_n}{n\alpha_n}\right)} + v_n\frac{\hat{b}_n(t_n)}{b(t_n)}\left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1\right).
\end{aligned}$$

From Theorem 4.1, one has $\xi_n^{(a)} := \sqrt{nh^d}\left(\frac{\hat{a}_n(t_n)-a(t_n)}{b(t_n)}\right) = O_{\mathbb{P}}(1)$, $\xi_n^{(b)} := \sqrt{nh^d}\left(\frac{\hat{b}_n(t_n)-b(t_n)}{b(t_n)}\right) = O_{\mathbb{P}}(1)$ and thus,

$$\frac{\sqrt{\frac{k_n}{nh^d}}\xi_n^{(a)}}{q_Z(\alpha_n)\log\left(\frac{k_n}{n\alpha_n}\right)} + \frac{\sqrt{\frac{k_n}{nh^d}}\xi_n^{(b)}}{\log\left(\frac{k_n}{n\alpha_n}\right)} \xrightarrow{\mathbb{P}} 0,$$

in view of $k_n/(nh^d) \to 0$, $q_Z(\alpha_n) \to \infty$ and $n\alpha_n/k_n \to 0$ as $n \to \infty$. In addition, since $\xi_n^{(b)} = O_{\mathbb{P}}(1)$, it follows that

$$\frac{\hat{b}_n(t_n)}{b(t_n)} = 1 + \frac{\xi_n^{(b)}}{\sqrt{nh^d}} \xrightarrow{\mathbb{P}} 1. \tag{A.26}$$

Besides, from Theorem 4.3(ii),

$$\begin{aligned} v_n \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) &= v_n \left(\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n) \right) (1 + o_{\mathbb{P}}(1)) \\ &\xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2), \end{aligned} \tag{A.27}$$

and collecting (A.26) and (A.27) yields

$$v_n \frac{\hat{b}_n(t_n)}{b(t_n)} \left(\frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right) \xrightarrow{d} \mathcal{N}(\beta/(1-\rho), \gamma^2).$$

The conclusion follows. □

Proof of Theorem 4.5. Recall that $v_n = \sqrt{k_n}/\log(k_n/(n\alpha_n))$. The proof follows the same lines as the one of Theorem 4.4:

$$\begin{aligned} & \frac{v_n}{q_Z(\alpha_n)} \max_{i \in I_n} \left| \frac{\tilde{q}_{n,Y}(\alpha_n \mid x_i) - q_Y(\alpha_n \mid x_i)}{b(x_i)} \right| \\ \leq \; & \frac{v_n}{q_Z(\alpha_n)} \max_{i \in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| + v_n \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| \\ + \; & v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right|. \end{aligned}$$

From Theorem 4.2,

$$\frac{v_n}{q_Z(\alpha_n)} \max_{i \in I_n} \left| \frac{\hat{a}_n(x_i) - a(x_i)}{b(x_i)} \right| + v_n \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i) - b(x_i)}{b(x_i)} \right| \xrightarrow{\mathbb{P}} 0,$$

since $q_Z(\alpha_n) \to \infty$ and under the two assumptions that $nh^d/(k_n \log n) \to \infty$ and $n\alpha_n/k_n \to 0$ as $n \to \infty$. In addition,

$$\max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right| \leq \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} - 1 \right| + 1 = O_{\mathbb{P}}(1), \tag{A.28}$$

from Theorem 4.2, and

$$v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| = v_n \left| (\log \hat{q}_{n,Z}(\alpha_n) - \log q_Z(\alpha_n))(1 + o_{\mathbb{P}}(1)) \right| = O_{\mathbb{P}}(1), \tag{A.29}$$

in view of Theorem 4.3(ii). Collecting (A.28) and (A.29) yields

$$v_n \left| \frac{\hat{q}_{n,Z}(\alpha_n)}{q_Z(\alpha_n)} - 1 \right| \max_{i \in I_n} \left| \frac{\hat{b}_n(x_i)}{b(x_i)} \right| = O_{\mathbb{P}}(1)$$

and the conclusion follows. □

Acknowledgments

The authors thank the associate editor and two referees for their valuable comments that led to an improved presentation of the results in this article.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004). *Statistics of extremes: theory and applications*, John Wiley and Sons. MR2108013

[2] Berlinet, A., Gannoun, A. and Matzner-Løber, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles, *Statistics*, 35(2), 139–169. MR1820681

[3] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*, Encyclopedia of Mathematics and its Applications, 27, Cambridge University Press. MR1015093

[4] Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*, Oxford Statistical Science Series, 18, Oxford University Press.

[5] Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C*, 54, 207–222. MR2134607

[6] Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society: Series B*, 58, 551–563. MR1394366

[7] Daouia, A., Florens, J-P. and Simar, L. (2010). Frontier estimation and extreme value theory. *Bernoulli*, 16, 1039–1063. MR2759168

[8] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589. MR3160564

[9] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20(2), 311–333. MR2834049

[10] Davison, A. C. and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society: Series B*, 62, 191–208. MR1747404

[11] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, 52, 393–442. MR1086795

[12] Einmahl, J. H. J., de Haan, L. and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B*, 78, 31–51. MR3453645

[13] Falk, M., Hüsler, J. and Reiss, R. D. (2004). *Laws of small numbers: Extremes and rare events*, 2nd edition, Birkhäuser. MR2104478

[14] Feuerverger, A., and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *The Annals of Statistics*, 27(2), 760–781. MR1714709

[15] Gardes, L. (2015). A general estimator for the extreme value index: ap-

plications to conditional and heteroscedastic extremes. *Extremes*, 18(3), 479–510. MR3377817

[16] Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99(10), 2368–2388. MR2463396

[17] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes*, 13(2), 177–204. MR2643556

[18] Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101(2), 419–433. MR2564351

[19] Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1), 45–75. MR3179970

[20] Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2020). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Submitted*, `http://hal.inria.fr/hal-02531027`.

[21] Goegebeur, Y., Guillou, A. and Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'IHP Probabilités et Statistiques*, 51(3), 1190–1213. MR3365978

[22] Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review*, 83(2), 263–292. MR3377081

[23] Gomes, M. I., and Martins, M. J. (2004). Bias reduction and explicit semiparametric estimation of the tail index. *Journal of Statistical Planning and Inference*, 124(2), 361–378. MR2080370

[24] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, New York, Springer. MR2234156

[25] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1), 60–70. MR1615558

[26] de Haan, L., Tank, A. K. and Neves, C. (2015). On tail trend detection: modeling relative risk. *Extremes*, 18(2), 141–178. MR3351812

[27] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15, 153–167. MR1788730

[28] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, 1163–1174. MR0378204

[29] Jalbert, J., Favre, A. C., Bélisle, C. and Angers, J. F. (2017). A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society: Series C*, 66(5), 941–962. MR3715590

[30] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of American Statistical Association*, 96, 458–468. MR1939348

[31] Kyung-Joon, C. and Schucany, W. (1998). Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference*, 66, 289–304. MR1614481

[32] Matheron, G. (1981). Sur la négligeabilité du squelette et la continuité absolue des érosions. *Technical Report N-724*, Centre de Géostatistique et de Morphologie Mathématique, 1–26.

[33] Müller, H. G. and Prewitt, K. (1991). Applications of multiparameter weak convergence for adaptive nonparametric curve estimation. In *Nonparametric Functional Estimation and Related Topics*, pp. 141–166, Springer, Dordrecht. MR1154326

[34] Müller, H. G. and Prewitt, K. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *Journal of Multivariate Analysis*, 47, 1–21. MR1239102

[35] Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics and Probability Letters*, 7(5), 407–412. MR1001144

[36] Serra, J. (1983). *Image analysis and mathematical morphology*. Academic Press, Inc. MR0753649

[37] Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion). *Statistical Science*, 4, 367–393. MR1041763

[38] Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5(4), 595–645. MR0443204

[39] Van Keilegom, I. and Wang, L. (2010). Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data. *Electronic Journal of Statistics*, 4, 133–160. MR2645480

[40] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association*, 73(364), 812–815. MR0521329